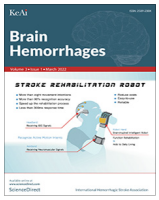
Brain Hemorrhages 5 (2024) 188–196

Contents lists available at ScienceDirect

Brain Hemorrhages

journal homepage: www.keaipublishing.com/en/journals/brain-hemorrhages/



Review article

The use of bioinformatic analysis to study intracerebral hemorrhage

Ilgiz Gareev [a], Ozal Beylerli [d], Tatiana Ilyasova [b], Andrey Mashkin [d], Huaizhang Shi [c,*]

[a] Central Research Laboratory, Bashkir State Medical University, Ufa, Republic of Bashkortostan, 3 Lenin Street 450008, Russia
[b] Department of Internal Diseases, Bashkir State Medical University, Ufa, Republic of Bashkortostan 450008, Russia
[c] Department of Neurosurgery, First Affiliated Hospital of Harbin Medical University, Harbin, Heilongjiang, China
[d] Educational and Scientific Institute of Neurosurgery, Peoples' Friendship University of Russia (RUDN University), Moscow, Russia

ARTICLE INFO

Article history:
Received 9 April 2024
Received in revised form 11 May 2024
Accepted 11 May 2024
Available online 14 May 2024

Keywords:
Intracerebral hemorrhage
Bioinformatics Analysis
Pathophysiology
Clinical Biomarkers
Therapeutic Targets
Personalized Medicine

ABSTRACT

The integration of bioinformatics analysis into intracerebral hemorrhage (ICH)research represents a paradigm shift in our approach to understanding, diagnosing, and treating this complex neurological disorder. By leveraging the power of bioinformatics, the scientific community is poised to make significant strides in combating this devastating condition, ultimately improving patient outcomes and quality of life. This study provides a comprehensive overview of the application of bioinformatics tools and techniques in elucidating the genetic, molecular, and environmental underpinnings of ICH. Through a detailed examination of genomic sequencing, transcriptomics, proteomics, and machine learning, we explore how these bioinformatics approaches have contributed to identifying genetic variants, understanding molecular pathways, and discovering biomarkers related to ICH. Challenges such as data complexity, integration of multi-omics data, and the translation of bioinformatics findings into clinical practice are discussed, alongside ethical considerations surrounding data privacy and patient consent. This study underscores the critical role of bioinformatics in advancing our understanding of ICH, offering insights into its pathophysiology, and paving the way for personalized medicine and targeted therapeutic interventions.

1. Introduction

Intracerebral hemorrhage (ICH), a type of stroke that occurs when a weakened blood vessel ruptures and bleeds into the surrounding brain, accounts for about 10–15 % of stroke cases. The complexity of its pathophysiology and the variability in its clinical presentation pose significant challenges in diagnosis, treatment, and prevention. Advances in bioinformatics have emerged as a promising avenue to address these challenges by enabling the analysis of large-scale biological data sets, thereby uncovering novel insights into the genetic, molecular, and environmental factors contributing to ICH (Fig. 1). By integrating data from genomics, proteomics, and other omics technologies, researchers are beginning to elucidate the complex molecular networks and pathways involved in the onset and progression of ICH. This review aims to explore the current landscape of bioinformatics analysis in the study of ICH, examining the tools, techniques, and methodologies that have propelled our understanding forward. We will provide a comprehensive overview of how bioinformatics analysis has been applied to study ICH, highlighting key findings, challenges, and future directions in the field.

2. Bioinformatics tools and techniques

Bioinformatics encompasses a vast field of study that combines biology, computer science, mathematics, and statistics to analyze and interpret biological data. In the context of ICH research, bioinformatics tools and techniques play a pivotal role in deciphering the complex genetic and molecular landscapes that underpin the disease. This section provides an overview of these tools and techniques, illustrating their application in neurological disorder studies.

2.1. Genomic sequencing and analysis

Genomic sequencing technologies, such as Whole Genome Sequencing (WGS) and Whole Exome Sequencing (WES), have revolutionized our understanding of the genetic underpinnings of ICH.[1] By comparing the genomes of individuals affected by ICH with those of healthy controls, researchers can identify genetic

* Corresponding author at: Department of Neurosurgery, The First Affiliated Hospital of Harbin Medical University, Harbin, Heilongjiang Province 150001, China.
E-mail address: shihuaizhang@hrbmu.edu.cn (H. Shi).

https://doi.org/10.1016/j.hest.2024.05.002

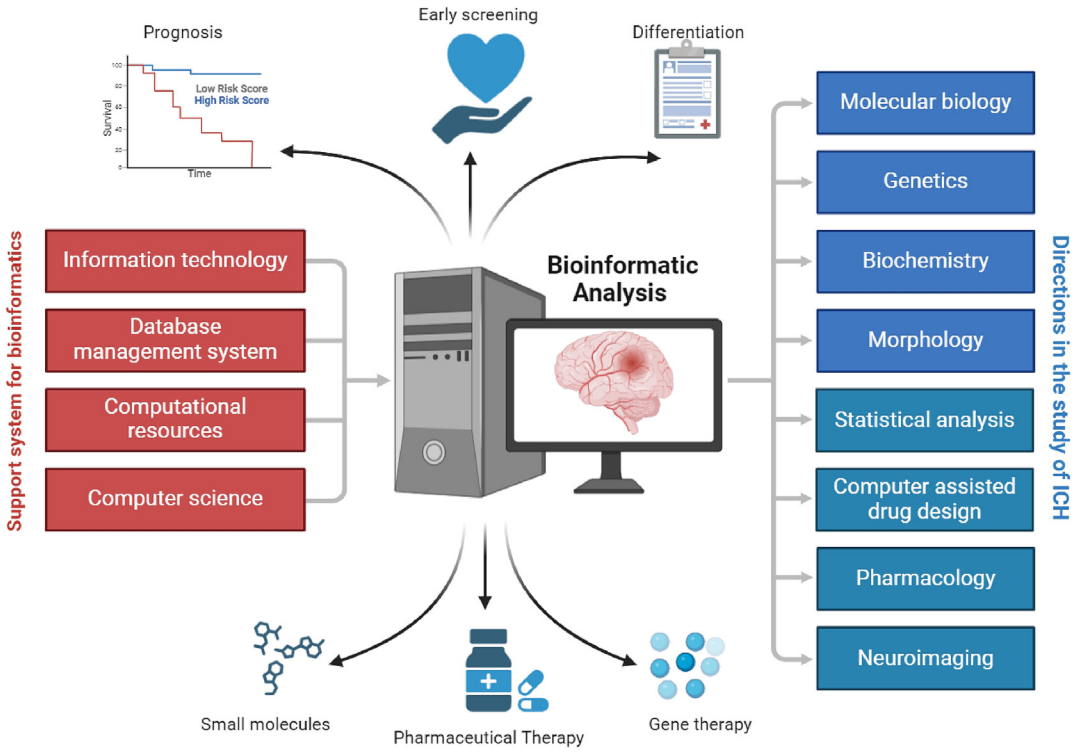


Fig. 1. Schematic illustration of the capabilities of bioinformatics analysis in intracerebral hemorrhage (ICH). To identify crucial biomarkers and therapeutic targets and establish the pathogenesis of ICH at the molecular level, bioinformatic analysis is one of key tools that may play principal roles in ICH.

variants associated with increased disease risk. Bioinformatics algorithms are crucial in this process, enabling the identification, alignment, and analysis of millions of genetic variants across the genome.

The advancement of high-throughput sequencing technologies has significantly transformed the field of human genetics and genomics, marking a pivotal shift in the approach to genetic analysis. With the widespread adoption of WGS, researchers now have the unprecedented capability to identify a comprehensive array of both common and rare genetic variants across nearly the entirety of the genome.[2] This technological leap not only enhances research into rare diseases and their clinical applications but also bolsters the identification and understanding of common diseases, along with the precise annotation of variants responsible for these conditions. As the global tally of sequenced genomes climbs into the hundreds of thousands, we stand on the threshold of a new era in which WGS emerges as the principal tool for genetic research. This represents a significant departure from the methodologies of the past decades, which predominantly relied on indirect genetic markers as surrogates for variants in adjacent regions, or limited sequencing efforts to the exonic portions of the genome. The functional analysis of variants uncovered through WGS is a critical aspect of studies in human genetics, vital for elucidating how these variants influence traits.[3] The advent of genome-wide functional genomics assays has greatly improved our ability to detect, characterize, and predict the molecular consequences of genetic variations. Nonetheless, given the intricate nature of genome functionality—a field in which our knowledge remains far from complete—there is a substantial amount of information still to be uncovered regarding the molecular impact of these variants. Such insights are crucial for understanding the potential these variants must affect more complex phenotypes at the organismal level, highlighting the ongoing journey of discovery in the realm of human genetics.

2.2. Transcriptomics and gene expression analysis

Transcriptomics, the study of the complete set of RNA transcripts produced by the genome, provides insights into the gene expression patterns associated with ICH. Techniques such as RNA sequencing (RNA-seq) allow for the quantification of gene expression levels in brain tissues or blood samples from stroke patients. Bioinformatics tools are then used to analyze these data, identifying genes that are differentially expressed in stroke patients, which may serve as potential biomarkers or therapeutic targets.

Over the last decade, single-cell RNA sequencing (scRNA-seq) has emerged as a groundbreaking technology in the life sciences, dramatically altering our approach to understanding cellular complexity and function. The advent of high-throughput scRNA-seq technologies, coupled with the development of sophisticated computational tools, has made this technology widely accessible and applicable across a broad spectrum of life science research areas. A key application of scRNA-seq technology is the development of single-cell atlases, which provide detailed maps of cell types and states within tissues, organs, and whole organisms. This endeavor has seen significant progress with the creation of comprehensive cell atlases for a variety of model organisms and humans, including Caenorhabditis elegans, planarians, Drosophila melanogaster, zebrafish, mice, Macaca fascicularis, and humans.[4–7] A notable achievement in this field is the study by Han et al., who for the first time mapped the cellular landscape of all major human organs, cataloging over 599 000 cells across 60 different tissue types. Their study established a foundational framework for understanding human cellular diversity, revealing previously uncharacterized cell hierarchies, and providing insights into cell identity and genetic networks.

The ongoing development and application of scRNA-seq are expected to generate vast amounts of data, which will be integrated into publicly accessible databases. This integration will

enhance our understanding of gene and cell functions in both health and disease. Furthermore, the combination of scRNA-seq with other large-scale genetic screening tools, such as CRISPR-based genomic screens, exemplified by technologies like Perturb-seq and LinTIMaT, is expanding the utility of scRNA-seq.[8–10] These combined approaches enable the exploration of the transcriptional consequences of gene knockouts, gene activation, or interference on a large scale, offering profound insights into how genetic variations influence cellular functions.

Despite its considerable promise, single-cell RNA sequencing faces challenges, such as the loss of spatial context due to the preparation of single-cell or single-nuclei suspensions. Although trajectory analysis can infer relationships between cell types, issues related to tissue processing and preservation may impact gene expression and cell representation. Addressing this, spatial transcriptomics techniques have been developed, marking a significant advancement in capturing the molecular architecture of tissues and organs within their spatial contexts. In 2020, spatially resolved transcriptomics was heralded as the «Method of the Year» by Nature Methods journal, highlighting its potential to revolutionize our understanding of cellular composition, complexity, and interactions within biological systems.

Looking ahead, integrating scRNA-seq into clinical diagnostics and personalized medicine presents a promising frontier. Although current scRNA-seq-based clinical studies are in early stages, focusing on understanding disease mechanisms and identifying diagnostic and therapeutic markers, the technology's potential is vast. Challenges remain, including the cost of sequencing and the need for user-friendly data analysis pipelines that can be operated by individuals without extensive bioinformatics expertise. Initiatives like the Galaxy Community's single-cell omics workbench represent steps toward making scRNA-seq more accessible for clinical applications. As technology advances and costs decrease, scRNA-seq is poised to play a pivotal role in advancing our understanding of biology and transforming clinical diagnostics and personalized medicine.

2.3. Proteomics and metabolomics

Proteomics and metabolomics represent frontier fields in the study of ICH, offering profound insights into the protein and metabolic alterations that accompany this devastating condition. While genomics and transcriptomics have laid the foundation for understanding the genetic and transcript levels, proteomics and metabolomics delve deeper into the post-genomic landscape, revealing the dynamic changes in protein expressions and metabolite concentrations critical for disease progression and resolution.[11] Proteomics, the comprehensive study of the proteome or the entire set of proteins expressed by a genome, tissue, or organism, stands at the forefront of ICH research.

Utilizing advanced mass spectrometry (MS) techniques, researchers can now quantify and profile the vast array of proteins present in the brain and circulating blood before and after an ICH.[12] This powerful approach allows for the identification of specific proteins that undergo significant changes during the stroke, shedding light on the molecular mechanisms underlying brain damage and recovery. Such insights are invaluable for developing targeted therapeutic interventions. Similarly, metabolomics—the study of small-molecule metabolites within cells, tissues, or organisms—provides a snapshot of the metabolic state affected by ICH.[13] By applying MS, scientists can detect and quantify thousands of metabolites, uncovering the metabolic pathways perturbed by the event. Understanding these metabolic changes is crucial for identifying the biochemical effects of stroke, highlighting potential metabolic biomarkers for early detection, and discovering novel therapeutic targets aimed at restoring metabolic balance. The wealth of data generated by proteomics and metabolomics studies necessitates sophisticated bioinformatic tools and methodologies for processing, analysis, and interpretation. Bioinformatics bridges the gap between raw data and meaningful biological insights, employing algorithms and statistical models to identify patterns and relationships among proteins and metabolites. Through such analysis, key molecules implicated in ICH pathophysiology are identified, offering promising avenues for biomarker discovery and the development of targeted therapies. The integration of proteomics and metabolomics into ICH research heralds a new era of molecular diagnostics and therapeutics. By unraveling the complex protein and metabolic alterations that occur, researchers are poised to identify novel biomarkers and therapeutic targets with unprecedented precision. As bioinformatic methodologies continue to evolve, the potential for personalized medicine and improved patient outcomes in ICH becomes ever more tangible.

2.4. Data mining and machine learning

Data mining and machine learning techniques are increasingly applied in bioinformatics research to uncover patterns and associations within large datasets. In ICH research, machine learning models can predict disease outcomes, identify novel disease subtypes, or suggest personalized treatment strategies based on patient data. These models rely on bioinformatics for data preprocessing, feature selection, and algorithm training and testing.

2.5. Network and pathway analysis

Network and pathway analysis are pivotal in unraveling the complex biological interactions underlying ICH. These analyses facilitate a comprehensive understanding of how genes, proteins, and metabolites interact within intricate networks, revealing the systemic impact of the disease. By dissecting these interactions, researchers can pinpoint the molecular mechanisms at play, advancing our knowledge of stroke pathology. Biological networks are intricate maps of the interactions between different molecular entities within cells. In the context of ICH, constructing these networks enables scientists to identify critical regulatory nodes—genes, proteins, or metabolites—that are significantly altered. These nodes often serve as hubs controlling key pathways that, when disrupted, contribute to the disease process. Identifying these hubs helps in understanding the disease's biology and in designing targeted therapeutic strategies.

Cytoscape emerges as a powerful bioinformatics tool for network and pathway analysis, offering researchers the ability to visualize molecular interactions and identify patterns within complex data sets. Its flexibility and comprehensive features make it invaluable for analyzing the molecular intricacies of ICH. By integrating data from various sources, Cytoscape facilitates the identification of potential biomarkers and therapeutic targets, enhancing our understanding of the disease's molecular landscape. Kyoto Encyclopedia of Genes and Genomes (KEGG) stands as a cornerstone among pathway databases, providing a vast repository of knowledge on biological pathways.[14] For ICH research, KEGG offers detailed maps of the pathways involved, highlighting how alterations in genes and metabolites can lead to pathophysiological changes. This information is crucial for uncovering the molecular drivers of stroke, offering a foundation for exploring novel therapeutic avenues. The application of network and pathway analysis in ICH research holds the promise of uncovering novel insights into the disease's molecular underpinnings. By leveraging bioinformatics platforms like Cytoscape and databases such as KEGG, researchers can delve deeper into the biology of stroke, identifying new therapeutic targets and strategies. As bioinformatics tools continue to evolve, the potential for breakthroughs in understanding and

treating ICH grows, paving the way for innovative approaches to combat this complex condition.

3. Pathophysiology of ICH

ICH, characterized by the rupture of cerebral blood vessels leading to intracranial bleeding, results from a complex interplay of genetic, molecular, and environmental factors (Fig. 2). Understanding these underlying mechanisms is crucial for the development of targeted therapies and preventive measures, highlighting the importance of comprehensive research into the disease's pathophysiology.[15] Recent advancements in genomic studies have shed light on several genetic variants that increase the risk of ICH.[16] Notably, mutations in the Collagen alpha-1 (IV) chain (COL4A1) gene, essential for the vascular basement membrane's integrity, have been linked to cerebral small vessel disease and ICH.[17] These discoveries are largely attributed to bioinformatics analyses of genome-wide association studies (GWAS), which utilize statistical software to sift through vast amounts of genetic data across populations, pinpointing risk alleles associated with the disease. Furthermore, transcriptomic analyses have unveiled significant alterations in gene expression related to critical biological processes such as blood–brain barrier (BBB) maintenance, inflammation, and apoptosis in the aftermath of hemorrhagic events. Employing RNA-seq technology coupled with bioinformatics processing, researchers have been able to map these dynamic gene expression changes, identifying potential molecular targets for intervention.

The role of microRNAs (miRNAs) and other non-coding RNAs (ncRNAs) in modulating gene expression also emerges as a significant facet of ICH pathogenesis[18–20] (Fig. 3). Bioinformatics analysis of miRNA expression profiles has revealed specific miRNAs that exhibit differential expression in ICH patients, implicating them in the regulation of pathways linked to vascular stability and inflammatory responses.[21] Integrating environmental and lifestyle factors into bioinformatics analyses enriches our understanding of their interplay with genetic predispositions in influencing ICH risk. Machine learning models have been particularly useful in evaluating the impact of hypertension, smoking, and alcohol consumption against individual genetic backgrounds, offering a comprehensive perspective on disease susceptibility. The quest for biomarkers capable of early ICH detection and prognosis has been significantly propelled forward by bioinformatics. Proteomics and metabolomics studies, leveraging advanced data analysis platforms, have identified circulating proteins and metabolites indicative of ICH onset. These biomarkers present new opportunities for non-invasive diagnostics and the monitoring of disease progression, opening avenues for timely intervention and personalized patient care. The pathophysiology of ICH encompasses a wide spectrum of genetic, molecular, and environmental components, each contributing to the disease's complexity. Through the lens of bioinformatics, researchers have begun to unravel the intricate web of interactions underlying ICH, paving the way for novel therapeutic strategies, and enhancing our ability to predict and prevent this life-threatening condition. As bioinformatics tools and methodologies continue to evolve, so too will our capacity to understand and combat ICH at its most fundamental levels.

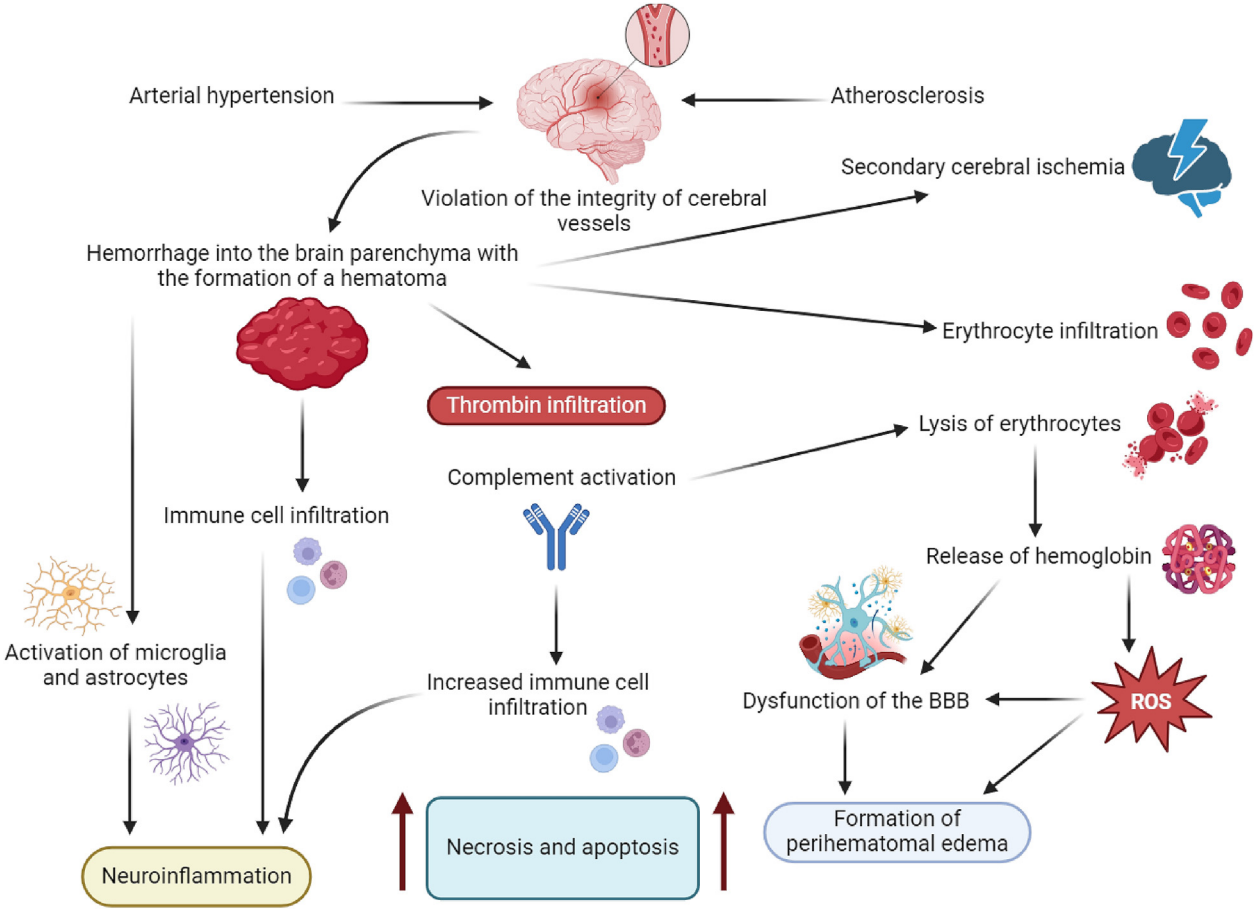


Fig. 2. The main links in the pathogenesis of intracerebral hemorrhage (ICH). The main etiological factors of ICH are hypertension and atherosclerosis of the cerebral vessels, which lead to damage to the wall and ultimately to rupture. The pathogenesis of ICH includes a complex set of pathobiochemical processes. ICH is accompanied by both primary and secondary damage to brain tissue.

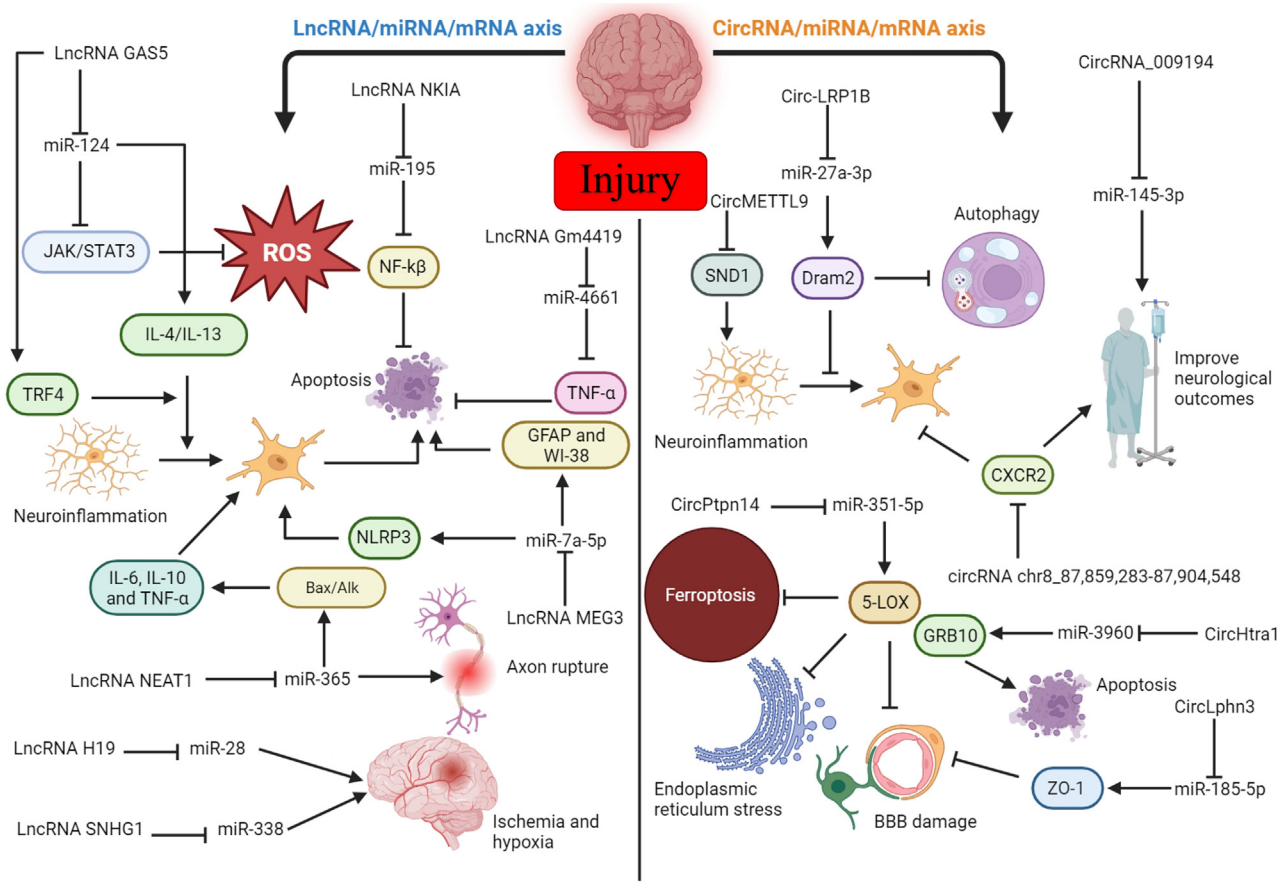


Fig. 3. Illustration demonstrating the molecular processes through epigenetic changes (non-coding RNAs (ncRNAs) (microRNAs (miRNAs), long non-coding RNAs (lncRNAs), and circular RNAs (circRNAs)) and genes) of brain damage in cerebrovascular diseases, including intracerebral hemorrhage (ICH). Due to specific changes in expression, ncRNAs can potentially be biomarkers for the diagnosis, treatment, and prognosis of ICH.

4. Case studies and applications

The application of bioinformatics in ICH research represents a paradigm shift, offering novel insights into the genetic underpinnings, molecular pathways, and potential therapeutic avenues for this devastating condition. Through the meticulous analysis of vast datasets generated by genomics, transcriptomics, proteomics, and other high-throughput technologies, bioinformatics has become a cornerstone in the quest to decipher the complex biological mechanisms of ICH. The examples highlighted below showcase the critical role of bioinformatics across various facets of stroke research, from identifying genetic risk factors to predicting patient outcomes and discovering new therapeutic targets. One of the most significant contributions of bioinformatics in this field has been in the realm of genetic research. A pivotal study leveraging WGS unearthed genetic variants associated with cerebral amyloid angiopathy, an important precursor to ICH. The bioinformatics-driven analysis of sequencing data pinpointed novel mutations in the APP gene, which is also implicated in Alzheimer's disease, thereby drawing a genetic link between these two neurological conditions. This discovery not only underscores the power of bioinformatics in uncovering hidden genetic associations but also opens new avenues for the early diagnosis and prevention of ICH.[22].

In another illustrative example, researchers employed RNA-seq to delineate gene expression profiles in brain tissues from ICH patients. The subsequent bioinformatics analysis revealed upregulation in pathways related to inflammation and immune response, shedding light on the molecular aftermath of ICHs. These insights have the potential to guide the development of targeted anti-inflammatory treatments, demonstrating the utility of transcriptomics and bioinformatics in unraveling the molecular responses to brain injury.[23] Proteomics studies have also benefited from bioinformatics, as illustrated by research analyzing cerebrospinal fluid from stroke patients to uncover protein biomarkers. Through bioinformatics processing of mass spectrometry data, researchers identified proteins with altered expression post-stroke. These proteins, linked to BBB integrity and neuronal injury, hold promise as diagnostic biomarkers, exemplifying the crucial role of bioinformatics in biomarker discovery.[24] The predictive potential of bioinformatics has been highlighted in studies employing machine learning to forecast outcomes in ICH patients based on a blend of clinical and genetic data. By training models on datasets encompassing patient characteristics, treatment interventions, and genetic markers, researchers achieved high accuracy in predicting patient outcomes. This case underscores the transformative impact of bioinformatics and machine learning in paving the way for personalized medicine, where treatment strategies can be tailored to individual risk profiles.[25–28].

Finally, network analysis studies have utilized bioinformatics to construct gene interaction networks pertinent to ICH. This comprehensive approach has unveiled key regulatory genes and pathways, offering a window into the disease's complex molecular mechanisms. By enabling a systems-level understanding of ICH, bioinformatics facilitates the exploration of novel therapeutic interventions, marking a significant stride towards unraveling the intricacies of stroke pathology.[29] Together, these examples illuminate the diverse and profound impact of bioinformatics on ICH research. By harnessing the power of computational analyses to

Table 1
Results of some studies in the field of bioinformatics analysis from Gene Expression Synthesis (GEO) datasets at intracerebral hemorrhage (ICH).

GEO datasets	Experiment type	Model of study	Results	Applications	References
GSE24265	Microarray	Human brain samples	DEGs (e.g.MAPK1/8, TNFAIP3, ATF4, and SLC2A1) are closely associated with ferroptosis after ICH	New therapeutic targets for ferroptosis	[30]
GSE24265, GSE87610, GSE92538, GSE54562, GSE54572, GSE54565, GSE54564, and GSE24095	Microarray	Human brain samples	HLA-A, HMOX1, and JUN related to endocytosis, cell adhesion, and phagosomes.	Potential pathogenic mechanisms of post-ICH depression	[31]
GSE24265	Microarray	Human brain samples	DEGs including chemokine-related, antigen immune-related, pathogen infection, cell reaction, and positive regulation of tyrosine phosphorylation and MAPK cascade were identified. CMap database identified Prestwick-1083, xamoterol, ifosfamide, methyldopate, nifurtimox, propranolol, and methoxamine as potential therapeutic agents for ICH while doxorubicin, menadione and azacitidine may increase its pathogenicity.	DEGs as potential biomarkers and small molecule compounds as new drugs	[32]
GSE24265	Microarray	Human brain samples	DEGs (IL6, TLR2, CXCL1, TIMP1, PLAUR, SERPINE1, SELE, CCL4, CCL20, and CD163) play an important role in the pathology of ICH. CMap database identified Hecogenin, Lidocaine, and NU-1025 as potential therapeutic agents for ICH	DEGs as new targets in the treatment of ICH and as prognostic markers. Small molecule compounds as new drugs	[21]
GSE24265, GSE149317, and GPL24688	Microarray and NGS	Human brain samples and rats brain samples	DEGs (IL-1β, STAT3, NLRP3 and NOD2) related to cytokine and receptor interactions and immune responses. DEGs have clinical predictive and diagnostic value	DEGs as biomarkers and targets in the treatment of ICH	[33]
GSE24265	Microarray	Human brain samples	YTHDF2, an m6A key gene, may regulate ICH progression by promoting infiltration of M1 macrophages or through the ceRNA network	Potential pathogenic mechanisms of ICH	

Abbreviations: DEGs, Differentially expressed genes; MAPK1/8, Mitogen-activated protein kinase 1/8; TNFAIP3, Tumor necrosis factor alpha-induced protein 3; ATF4, Activating transcription factor 4; SLC2A1, Solute carrier family 2, facilitated glucose transporter member 1; HLA-A, Major histocompatibility complex, class I, A; HMOX1, Heme oxygenase 1; CMap, Connectivity Map; IL-6, Interleukin-6; TLR2, Toll-like receptor 2; CXCL1, C-X-C motif chemokine ligand 1; TIMP1, Tissue inhibitor of metalloproteinase 1; PLAUR, Plasminogen activator, urokinase receptor; SERPINE1, Plasminogen activator inhibitor 1; SELE, Selectin E; CCL4, Chemokine (C–C motif) ligands 4; CCL20, C–C motif chemokine ligand 20; CD163, Cluster of differentiation 163; IL-1β, Interleukin-1β; STAT3, Signal transducer and activator of transcription 3; YTHDF2, YTH N6-methyladenosine RNA binding protein 2; ceRNA, Competing endogenous RNA.

interpret complex biological data, bioinformatics is not just enhancing our understanding of ICH but also steering the field towards more effective diagnostic, preventive, and therapeutic strategies. The most important results and discoveries in the field of bioinformatics analysis and ICH research are summarized in Table 1.[21,30–34].

5. Challenges in bioinformatics analysis

The integration of bioinformatics into ICH research represents a pivotal shift towards precision medicine, offering unprecedented opportunities to decipher the complex biological underpinnings of this condition. Yet, this integration is fraught with challenges that span the spectrum from technical and computational hurdles to ethical and privacy concerns, each adding layers of complexity to research efforts. The volume and intricacy of data generated by genomic, transcriptomic, and proteomic analyses are staggering. High-throughput methodologies, capable of generating terabytes of data from a single experiment, necessitate not only advanced bioinformatics tools but also significant computational power to ensure efficient data processing and analysis. This deluge of data places immense pressure on existing storage and management systems, demanding innovative solutions for data handling and interpretation.[35] Moreover, the challenge extends beyond mere data handling to the integration of disparate data types. ICH, like many complex diseases, is influenced by a wide array of genetic, molecular, and environmental factors. The integration of data from genomics, transcriptomics, proteomics, and other omics technologies is essential for a holistic understanding of the disease's pathology. This necessitates the development and application of sophisticated bioinformatics algorithms capable of navigating and making sense of this multidimensional data landscape, thereby uncovering patterns and connections across different biological levels.[36].

Another critical issue is the reproducibility and standardization of bioinformatics analyses. The field suffers from a lack of standardized protocols, which, coupled with variations in data processing methodologies, often leads to inconsistent and non-comparable results across studies. This variability hampers the ability to validate findings and build upon previous work, thereby impeding progress.[37–38] Ethical considerations and data privacy concerns are paramount when dealing with patient-derived data, which is often the case in ICH research. The use of genetic and health-related information necessitates rigorous ethical oversight and robust data protection measures to safeguard the confidentiality and security of personal data. This aspect of research is critical not only for maintaining public trust but also for ensuring compliance with increasingly stringent data privacy laws.[39–40].

Finally, translating the insights gained from bioinformatics research into tangible clinical applications remains a significant hurdle. The path from discovery to clinical implementation is fraught with challenges, including the need for interdisciplinary collaboration among bioinformaticians, biologists, clinicians, and other stakeholders. Developing clear and efficient pathways for the clinical validation of bioinformatics-derived hypotheses is essential for harnessing the full potential of this research to benefit patients.[41] Addressing these challenges requires a concerted effort from the scientific community, involving collaboration across disciplines, the development of innovative computational tools, and

Table 2
Advantages and disadvantages of using bioinformatics analysis for the prognosis and diagnosis of intracerebral hemorrhage (ICH).

Advantages	Disadvantages
Comprehensive analysis: Bioinformatics allows for the integration and analysis of large-scale biological data sets, including genomics, proteomics, and transcriptomics, providing a holistic view of the disease mechanisms	**Complexity and expertise:** The complexity of bioinformatics tools and data requires specialized knowledge and expertise, potentially limiting accessibility for clinicians and researchers without a background in computational biology
Early diagnosis: Through the identification of genetic markers and molecular signatures, bioinformatics can facilitate the early detection of hemorrhagic stroke, potentially before clinical symptoms manifest	**Data volume and management:** The sheer volume of data generated by genomic and proteomic analyses can be overwhelming, necessitating substantial computational resources and sophisticated data management strategies
Personalized medicine: Bioinformatics analysis can help identify individual risk factors and predict responses to treatment, enabling personalized therapeutic strategies and improving patient outcomes	**Reproducibility issues:** Variability in data collection, processing, and analysis methods can lead to issues with reproducibility and consistency across studies, complicating the validation of findings
Novel biomarker discovery: The use of bioinformatics tools can lead to the discovery of new biomarkers for ICH, enhancing diagnostic accuracy and prognostic assessment	**Integration of multiscale data:** Integrating and interpreting data from diverse sources (genomic, clinical, environmental) remains challenging, requiring advanced models that can effectively combine and analyze these data types
Predictive modeling: Advanced computational models and machine learning algorithms can predict disease progression and outcomes, offering valuable insights for clinical decision-making	**Ethical and privacy concerns:** The use of patient-derived genomic data raises ethical questions and privacy concerns, necessitating robust data protection measures and ethical oversight
Cost-Effectiveness: Over time, bioinformatics tools can become cost-effective alternatives to traditional diagnostic and prognostic methods, reducing the need for invasive procedures and expensive laboratory tests	**Clinical translation:** There is often a gap between bioinformatics discoveries and their application in clinical settings, with barriers to translating research findings into practice

the establishment of ethical frameworks for data use. As the field advances, overcoming these hurdles will be crucial for realizing the promise of bioinformatics in transforming our understanding and treatment of ICH (Table 2).

6. Future directions in bioinformatics for ICH research

The horizon of bioinformatics in the realm of ICH research holds immense promise, with advancements poised to redefine our understanding and management of this devastating condition. The evolution of computational models, particularly those leveraging the power of deep learning and artificial intelligence (AI), stands at the forefront of this transformative shift. These sophisticated models are designed to navigate the complexity and diversity of stroke-related data, encompassing everything from genetic sequences to clinical imaging. Their ability to provide nuanced predictions about disease risk, progression, and treatment responsiveness marks a significant leap forward in personalized medicine.[42–43] The integration of multi-omics data—spanning genomics, proteomics, metabolomics—and clinical information is set to offer a comprehensive lens through which to view ICH. Such an approach not only aims to elucidate the disease's multifaceted nature but also to pave the way for tailored therapeutic interventions. By capturing the intricate interplay between various biological and environmental factors, this integrated methodology seeks to herald a new era of precision medicine in stroke care.[36] Single-cell sequencing technologies are another frontier, rapidly expanding our ability to probe the cellular and molecular underpinnings of ICH with remarkable detail. Bioinformatics will be instrumental in deciphering the vast datasets generated by single-cell analyses, enabling the identification of cellular contributors to stroke pathology and the discovery of novel targets for therapeutic intervention.[44] Moreover, the future of ICH research will greatly benefit from enhanced data sharing and collaborative frameworks. The push towards

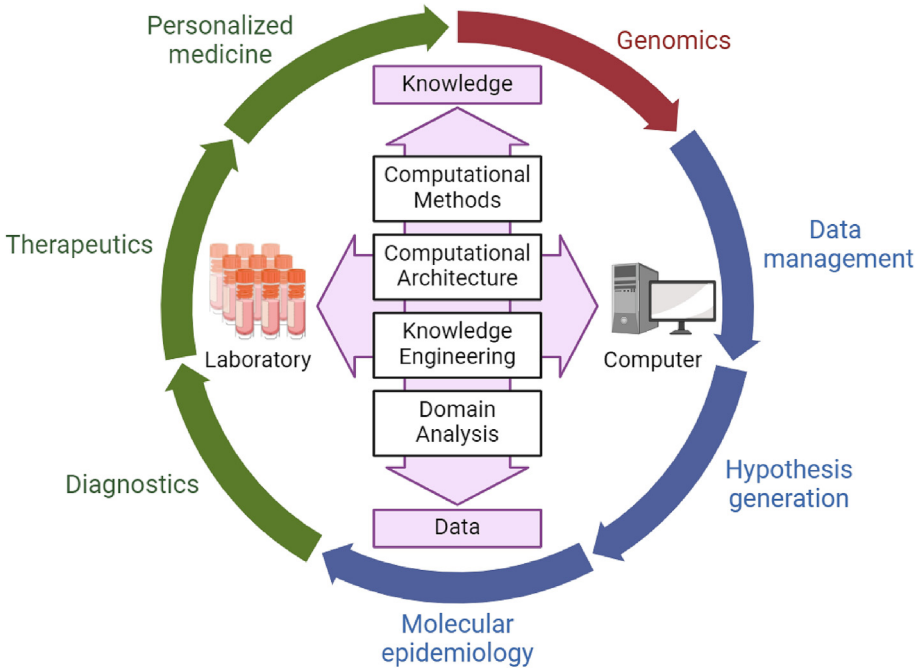


Fig. 4. Potential for bioinformatics analysis in clinical practice. The frontier of diseases research involves collaborations between medical doctors and computer scientists.

standardizing data formats and establishing open-access databases is expected to streamline the sharing of bioinformatics tools and findings. Such collaborative efforts will not only accelerate the pace of discovery but also democratize access to cutting-edge resources, fostering innovation and facilitating global research endeavors.[45] As the field progresses, the ethical and regulatory landscape governing bioinformatics research will also need to adapt. Advancements in safeguarding data privacy, alongside the ethical management of genetic information and the refinement of patient consent models, are imperative. These measures will ensure that the expansion of bioinformatics in ICH research remains aligned with the principles of responsible conduct and patient-centric care.[46].

In summary, the trajectory of bioinformatics within ICH research is marked by both exciting potential and significant challenges. The development of AI and deep learning models, the integration of multi-omics data, advancements in single-cell sequencing, and the promotion of global collaboration and ethical standards represent key areas of opportunity. Together, these developments promise to usher in a new chapter in the battle against ICH, transforming research paradigms and ultimately improving patient outcomes (Fig. 4).

7. Conclusions

Bioinformatics has emerged as a cornerstone in the study of ICH, offering profound insights into its genetic, molecular, and environmental underpinnings. Through the application of sophisticated computational tools and techniques, researchers have begun to unravel the complex biological networks that contribute to the development and progression of this devastating condition. From identifying genetic variants associated with increased disease risk to uncovering novel biomarkers for early detection, bioinformatics analysis has paved the way for advancements in diagnosis, treatment, and prevention strategies.[47] However, the journey is not without its obstacles. The sheer volume and complexity of biological data, the integration of multi-omics information, and the need for standardized methodologies pose significant challenges. Moreover, ethical considerations and the translation of bioinformatics discoveries into clinical practice require careful navigation and interdisciplinary collaboration. Looking ahead, the future of bioinformatics in ICH research holds immense promise. Advances in computational models, single-cell sequencing technologies, and data sharing initiatives, coupled with ongoing efforts to address ethical and regulatory issues, are set to further enhance our understanding and management of the disease. By fostering innovation and collaboration, the bioinformatics community is well-positioned to contribute to significant breakthroughs in ICH research. As we continue to explore the vast potential of bioinformatics, this field will play a pivotal role in shaping the future of neurological disorder research. By harnessing the power of computational analysis to interpret the complex tapestry of biological data, bioinformatics stands at the forefront of efforts to combat ICH, promising a future where personalized medicine and targeted therapies become a reality for patients around the world.

By performing bioinformatics analysis, we provide a better understanding of the role of certain elements, such as ncRNAs, in the pathogenesis of ICH, which can serve as a theoretical basis for the further development of therapeutic agents and biomarkers for early intervention. Moreover, we suggest that future studies involve larger and more heterogeneous populations and focus on comparing the diagnostic and therapeutic abilities of genes or signaling pathways in ICH among different races, ethnicities, age groups, and the presence of patient comorbidities, mainly metabolic (e.g., diabetes mellitus (DM), obesity and hypertension) and other vascular diseases (e.g., coronary heart disease (CHD) and peripheral artery disease (PAD)). Further research in bioinformatics and ICH may also create a dedicated panel consisting of more than one type of molecule.

CRediT authorship contribution statement

Ilgiz Gareev: Writing – review & editing, Writing – original draft, Project administration. **Ozal Beylerli:** Visualization, Validation, Project administration, Formal analysis, Conceptualization. **Tatiana Ilyasova:** Visualization, Validation. **Andrey Mashkin:** Data curation. **Huaizhang Shi:** Supervision, Resources, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethical Statement

This review article does not contain any studies with human participants or animals performed by any of the authors.

Acknowledgments

The authors express their gratitude to the editors of the journal for their professional and responsible approach to work!

Funding

This work was supported by the Bashkir State Medical University Strategic Academic Leadership Program (PRIORITY-2030).

References

1. Castel SE, Cervera A, Mohammadi P, et al. Modified penetrance of coding variants by cis-regulatory variation contributes to disease risk. *Nat Genet*. 2018 Sep;50(9):1327–1334. https://doi.org/10.1038/s41588-018-0192-y.
2. Lappalainen T, Scott AJ, Brandt M, Hall IM. Genomic analysis in the age of human genome sequencing. *Cell*. 2019 Mar 21;177(1):70–84. https://doi.org/10.1016/j.cell.2019.02.032.
3. McCombie WR, McPherson JD, Mardis ER. Next-generation sequencing technologies. *Cold Spring Harb Perspect Med*. 2019 Nov 1;9(11). https://doi.org/10.1101/cshperspect.a036798 a036798.
4. Cao J, Packer JS, Ramani V, et al. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science*. 2017 Aug 18;357(6352):661–667. https://doi.org/10.1126/science.aam8940.
5. Cusanovich DA, Hill AJ, Aghamirzaie D, et al. A single-cell atlas of in vivo mammalian chromatin accessibility. *Cell*. 2018 Aug 23;174(5):1309–1324.e18. https://doi.org/10.1016/j.cell.2018.06.052.
6. Cusanovich DA, Reddington JP, Garfield DA, et al. The cis-regulatory dynamics of embryonic development at single-cell resolution. *Nature*. 2018 Mar 22;555(7697):538–542. https://doi.org/10.1038/nature25981.
7. Fincher CT, Wurtzel O, de Hoog T, Kravarik KM, Reddien PW. Cell type transcriptome atlas for the planarian Schmidtea mediterranea. Science. 2018 May 25;360(6391):eaaq1736. doi: 10.1126/science.aaq1736.
8. Replogle JM, Norman TM, Xu A, et al. Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. *Nat Biotechnol*. 2020 Aug;38(8):954–961. https://doi.org/10.1038/s41587-020-0470-y.
9. Alda-Catalinas C, Bredikhin D, Hernando-Herraez I, et al. A single-cell transcriptomics CRISPR-activation screen identifies epigenetic regulators of the zygotic genome activation program. *Cell Syst*. 2020 Jul 22;11(1):25–41.e9. https://doi.org/10.1016/j.cels.2020.06.004.
10. Anzalone AV, Randolph PB, Davis JR, et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*. 2019 Dec;576(7785):149–157. https://doi.org/10.1038/s41586-019-1711-4.
11. Taylor MJ, Lukowski JK, Anderton CR. Spatially resolved mass spectrometry at the single cell: Recent innovations in proteomics and metabolomics. *J Am Soc Mass Spectrom*. 2021 Apr 7;32(4):872–894. https://doi.org/10.1021/jasms.0c00439.

12. Barallobre-Barreiro J, Chung YL, Mayr M. Proteomics and metabolomics for mechanistic insights and biomarker discovery in cardiovascular disease. *Rev Esp Cardiol (Engl Ed)*. 2013 Aug;66(8):657–661. https://doi.org/10.1016/j.rec.2013.04.009.
13. Basak T, Varshney S, Akhtar S, Sengupta S. Understanding different facets of cardiovascular diseases based on model systems to human studies: a proteomic and metabolomic perspective. *J Proteomics*. 2015 Sep 8;127(Pt A):50–60. https://doi.org/10.1016/j.jprot.2015.04.027.
14. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res*. 2017 Jan 4;45(D1):D353–D361. https://doi.org/10.1093/nar/gkw1092.
15. Magid-Bernstein J, Girard R, Polster S, et al. Cerebral hemorrhage: Pathophysiology, treatment, and future directions. *Circ Res*. 2022 Apr 15;130(8):1204–1229. https://doi.org/10.1161/CIRCRESAHA.121.319949.
16. Chauhan G, Debette S. Genetic risk factors for ischemic and ICH. *Curr Cardiol Rep*. 2016 Dec;18(12):124. https://doi.org/10.1007/s11886-016-0804-z.
17. Zhang L, Tao W, Feng H, Chen Y. Transcriptional and genomic targets of neural stem cells for functional recovery after ICH. *Stem Cells Int*. 2017;2017:2412890. https://doi.org/10.1155/2017/2412890.
18. Gareev I, Beylerli O, Zhao B. MiRNAs as potential therapeutic targets and biomarkers for non-traumatic intracerebral hemorrhage. *Biomark Res*. 2024 Feb 2;12(1):17. https://doi.org/10.1186/s40364-024-00568-y.
19. Gareev I, Beylerli O, Liang Y, et al. The role of mitochondria-targeting miRNAs in intracerebral hemorrhage. *Curr Neuropharmacol*. 2023;21(5):1065–1080. https://doi.org/10.2174/1570159X20666220507021445.
20. Guan X, Zhang Y, Gareev I, Beylerli O, Li X, Lu G, Lv L, Hai X. MiR-499a prevents astrocytes mediated inflammation in ischemic stroke by targeting PTEN. Noncoding RNA Res. 2021 Sep 28;6(3):146-152. doi: 10.1016/j.ncrna.2021.09.002. Erratum in: Noncoding RNA Res. 2022 Dec 13;8(4):693.
21. Liu Z, Zhang R, Chen X, et al. Identification of hub genes and small-molecule compounds related to intracerebral hemorrhage with bioinformatics analysis. *PeerJ*. 2019 Oct;25(7):e7782. https://doi.org/10.7717/peerj.7782.
22. Tan RYY, Traylor M, Megy K, Duarte D, Deevi SVV, Shamardina O, Mapeta RP; NIHR BioResource: Rare Diseases Consortium; Ouwehand WH, Gräf S, Downes K, Markus HS. How common are single gene mutations as a cause for lacunar stroke? A targeted gene panel study. Neurology. 2019 Nov 26;93(22):e2007-e2020. doi: 10.1212/WNL.0000000000008544.
23. Diagbouga MR, Lemeille S, Morel S, Kwak BR. Impact of disrupted cyclic stretch in intracranial aneurysms: Insights from endothelial cell transcriptomic dataset. *Data Brief*. 2023 Dec;28(52). https://doi.org/10.1016/j.dib.2023.110014 110014.
24. O'Connell GC, Stafford P, Walsh KB, Adeoye O, Barr TL. High-throughput profiling of circulating antibody signatures for stroke diagnosis using small volumes of whole blood. *Neurotherapeutics*. 2019 Jul;16(3):868–877. https://doi.org/10.1007/s13311-019-00720-9.
25. Choi JM, Seo SY, Kim PJ, et al. Prediction of hemorrhagic transformation after ischemic stroke using machine learning. *J Pers Med*. 2021 Aug 30;11(9):863. https://doi.org/10.3390/jpm11090863.
26. Huang J, Chen H, Deng J, et al. Interpretable machine learning for predicting 28-day all-cause in-hospital mortality for hypertensive ischemic or ICH patients in the ICU: A multi-center retrospective cohort study with internal and external cross-validation. *Front Neurol*. 2023 Aug;8(14):1185447. https://doi.org/10.3389/fneur.2023.1185447.
27. Lin CH, Hsu KC, Johnson KR, et al. Taiwan stroke registry investigators. Evaluation of machine learning methods to stroke outcome prediction using a nationwide disease registry. *Comput Methods Programs Biomed*. 2020 Jul;190. https://doi.org/10.1016/j.cmpb.2020.105381 105381.
28. Kurtz P, Peres IT, Soares M, Salluh JIF, Bozza FA. Hospital length of stay and 30-day mortality prediction in stroke: A machine learning analysis of 17,000 ICU Admissions in Brazil. *Neurocrit Care*. 2022 Aug;37(Suppl 2):313–321. https://doi.org/10.1007/s12028-022-01486-3.
29. Li PF, Guo SC, Liu T, et al. Integrative analysis of transcriptomes highlights potential functions of transfer-RNA-derived small RNAs in experimental intracerebral hemorrhage. *Aging (Albany NY)*. 2020 Nov 16;12(22):22794–22813. https://doi.org/10.18632/aging.103938.
30. Liu T, Li X, Cui Y, et al. Bioinformatics analysis identifies potential ferroptosis key genes in the pathogenesis of intracerebral hemorrhage. *Front Neurosci*. 2021 Jun;7(15). https://doi.org/10.3389/fnins.2021.661663 661663.
31. Zou C, Huang X, Lan X, et al. Potential genes and mechanisms linking intracerebral hemorrhage and depression: A bioinformatics-based study. *Int J Gen Med*. 2021 Apr;7(14):1213–1226. https://doi.org/10.2147/IJGM.S302916.
32. Yang Z, Wang R, Chen X, Zhao D. Identification of potential biomarkers and small-molecule compounds related to intracerebral hemorrhage with bioinformatics analysis. *Acta Neurobiol Exp (Wars)*. 2022;82(2):187–196. https://doi.org/10.55782/ane-2022-017.
33. Xiao Y, Zhang Y, Wang C, Ge Y, Gao J, Huang T. The use of multiple datasets to identify autophagy-related molecular mechanisms in intracerebral hemorrhage. *Front Genet*. 2023 Apr;3(14):1032639. https://doi.org/10.3389/fgene.2023.1032639.
34. Zhou B. Bioinformatics analysis identifies potential m6A hub genes in the pathogenesis of intracerebral hemorrhage. *J Neuroimmunol*. 2023 Dec;15(385). https://doi.org/10.1016/j.jneuroim.2023.578224 578224.
35. Navarro-González C, Huerga-Gómez A, Fazzari P. Nrg1 intracellular signaling is neuroprotective upon stroke. *Oxid Med Cell Longev*. 2019 Sep;8(2019):3930186. https://doi.org/10.1155/2019/3930186.
36. Meneri M, Bonato S, Gagliardi D, Comi GP, Corti S. New insights into cerebral vessel disease landscapes at single-cell resolution: Pathogenetic and therapeutic perspectives. *Biomedicines*. 2022 Jul 13;10(7):1693. https://doi.org/10.3390/biomedicines10071693.
37. Cokelaer T, Cohen-Boulakia S, Lemoine F. Reprohackathons: Promoting reproducibility in bioinformatics through training. *Bioinformatics*. 2023 Jun 30;39(39 Suppl 1):i11–i20. https://doi.org/10.1093/bioinformatics/btad227.
38. Miłkowski M, Hensel WM, Hohol M. Replicability or reproducibility? On the replication crisis in computational neuroscience and sharing only relevant detail. *J Comput Neurosci*. 2018 Dec;45(3):163–172. https://doi.org/10.1007/s10827-018-0702-z.
39. Heeney C, Kerr SM. Balancing the local and the universal in maintaining ethical access to a genomics biobank. *BMC Med Ethics*. 2017 Dec 28;18(1):80. https://doi.org/10.1186/s12910-017-0240-7.
40. Black L, Avard D, Zawati MH, Knoppers BM, Hébert J, Sauvageau G. Leucegene Project. Funding considerations for the disclosure of genetic incidental findings in biobank research. *Clin Genet*. 2013 Nov;84(5):397–406. https://doi.org/10.1111/cge.12190.
41. Ning M, Lopez M, Cao J, Buonanno FS, Lo EH. Application of proteomics to cerebrovascular disease. *Electrophoresis*. 2012 Dec;33(24):3582–3597. https://doi.org/10.1002/elps.201200481.
42. Morris-Blanco KC, Chokkalla AK, Arruri V, Jeong S, Probelsky SM, Vemuganti R. Epigenetic mechanisms and potential therapeutic targets in stroke. *J Cereb Blood Flow Metab*. 2022 Nov;42(11):2000–2016. https://doi.org/10.1177/0271678X221116192.
43. Shue E, Liu L, Li B, Feng Z, Li X, Hu G. Empowering beginners in bioinformatics with ChatGPT. *Quant Biol*. 2023 Jun;11(2):105–108. https://doi.org/10.15302/j-qb-023-0327.
44. Shi X, Luo L, Wang J, Shen H, Li Y, Mamtilahun M, Liu C, Shi R, Lee JH, Tian H, Zhang Z, Wang Y, Chung WS, Tang Y, Yang GY. Stroke subtype-dependent synapse elimination by reactive gliosis in mice. Nat Commun. 2021 Nov 26;12(1):6943. doi: 10.1038/s41467-021-27248-x. Erratum in: Nat Commun. 2022 Feb 28;13(1):1183.
45. Callahan A, Anderson KD, Beattie MS, Bixby JL, Ferguson AR, Fouad K, Jakeman LB, Nielson JL, Popovich PG, Schwab JM, Lemmon VP; FAIR Share Workshop Participants. Developing a data sharing community for spinal cord injury research. Exp Neurol. 2017 Sep;295:135-143. doi: 10.1016/j.expneurol.2017.05.012.
46. Vähäkangas K. Research ethics in the post-genomic era. *Environ Mol Mutagen*. 2013 Aug;54(7):599–610. https://doi.org/10.1002/em.21804.
47. Rösler W, Altenbuchinger M, Baeßler B, et al. An overview and a roadmap for artificial intelligence in hematology and oncology. *J Cancer Res Clin Oncol*. 2023 Aug;149(10):7997–8006. https://doi.org/10.1007/s00432-023-04667-5.