





# Biomedical Big Data and Artificial Intelligence in Blood

Fuhong He (和夫红) <sup>1,2,3,\*</sup>, Zhaojun Zhang (张昭军) <sup>1,2,3,\*</sup>, Xiangdong Fang (方向东) <sup>1,2,3,\*</sup>, Qian-Fei Wang (王前飞) <sup>1,2,3,\*</sup>

<sup>1</sup>China National Center for Bioinformation, Beijing 100101, China

<sup>2</sup>Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China

<sup>3</sup>University of Chinese Academy of Sciences, Beijing 100049, China

\*Corresponding authors: hefh@big.ac.cn (He F), zhangzhaojun@big.ac.cn (Zhang Z), fangxd@big.ac.cn (Fang X), wangqf@big.ac.cn (Wang QF).

The hematopoietic system has long served as an excellent model for biological and medical research, owing to its highly organized hierarchical structure, accessibility for sampling, and rapid cellular turnover. These features have enabled pivotal discoveries in stem cell biology, oncogenic transformation, and targeted therapies, exemplified by milestones such as the identification of the *BCR-ABL* fusion gene and the successful development of molecular-targeted treatments. With its intrinsic advantages, recently hematology continues to provide critical insights attributed to the rapid expansion of high-throughput omics technologies and bioinformatics, and is entering a new era that emphasizes data-driven discovery and intelligent clinical decision-making.

This special issue of *Genomics, Proteomics & Bioinformatics*, entitled “Biomedical Big Data in Blood”, includes 19 studies that collectively highlight the power of how omics integration and artificial intelligence (AI) technologies are reshaping hematological research (Table 1). These contributions cover resource construction, mechanistic exploration, translational applications, and computational modeling. Due to space limitations, we described 10 representative studies below.

## Data integration and bioinformatic tools: building the digital foundation

Robust data resources and user-friendly analytical platforms are essential for advancing omics-driven hematology research. This section features four articles that focus on the development of specialized tools and databases addressing diverse cellular contexts, model organisms, and clinical scenarios.

Wang et al. [1] developed HemaScope, an open-source bioinformatics toolkit tailored for analyzing single-cell and spatial transcriptomic data in hematopoietic systems. HemaScope integrates modules for atlas construction, lineage tracing, dynamic transcriptional analysis, and microenvironmental profiling. The toolkit has demonstrated robust performance across multiple scenarios, including bone marrow aging, acute myeloid leukemia (AML), and T-cell lymphoma. Kang et al. [2] introduced HemAtlas, a cross-species, multi-omics database dedicated to hematopoiesis. It incorporates transcriptomic, epigenomic, and spatial transcriptomic data from humans, mice, zebrafish, and *in vitro* hematopoietic stem and progenitor cell models. HemAtlas enables comparative

analyses across developmental stages and tissues, enhancing understanding of hematopoietic development and regeneration.

Zheng et al. [3] established EryDB, a comprehensive erythroid transcriptomic database integrating bulk and single-cell RNA-seq datasets across species and disease conditions. EryDB facilitates comparative exploration of erythropoiesis and identification of dysregulated pathways in erythroid-related disorders. Zhou et al. [4] developed NeoTCR, an immunoinformatic database of experimentally validated neoantigen-specific T-cell receptors (TCRs) from 18 cancer types. NeoTCR offers unified annotations of publicly available neoantigen-specific TCR sequences along with relevant neoantigen information. It also provides a one-stop platform for clonotype discovery, neoantigen annotation, and immunotherapy prediction, bridging sequencing data with precision immuno-oncology.

Collectively, these resources help close existing gaps in data accessibility, analytical capabilities, and clinical integration, laying a digital foundation for mechanistic discoveries and personalized therapeutic strategies.

## Omics-guided mechanistic studies: from molecular insights to therapeutic strategies

The convergence of multi-omics technologies and clinical expertise is transforming our understanding of hematologic diseases, bridging molecular mechanisms with therapeutic advances. As mechanistic research enters the big data era, validation and translational foresight become crucial. This section highlights four studies that leverage genomics, transcriptomics, epigenomics, and integrative omics to uncover disease pathways and therapeutic strategies.

Epigenetic reprogramming is recognized as one of the major drivers in hematological malignancies. Zhang et al. [5] identified histone methyltransferase G9a inhibitors as candidate drugs for SETD2-deficient leukemia using the Connectivity Map combined with a drug screening platform. Their integrative analysis of transcriptomic and ChIP-seq data showed that G9a inhibition upregulates *let-7a-2*, potentially via H3K9me2 reduction, suppressing oncogenic MYC signaling. This epigenetic regulation–noncoding RNA–oncogenic axis positions G9a as a promising therapeutic target. Liu et al. [6] revealed that RNA N<sup>6</sup>-methyladenosine (m<sup>6</sup>A)

Received: 30 April 2025; Accepted: 30 April 2025.

© The Author(s) 2025. Published by Oxford University Press and Science Press on behalf of the Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation and Genetics Society of China.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Table 1 Original studies, resources, and tools reported in the special issue on biomedical big data in blood

Category	Name	Brief description	Web link	Corresponding author(s)	Ref.
Resource	HemaAtlas	A multi-omics hematopoiesis database	<a href="https://ngdc.cncb.ac.cn/hematlas">https://ngdc.cncb.ac.cn/hematlas</a>	Feng Liu, Zhang Zhang	[2]
	EryDB	A comprehensive database for erythropoiesis and erythroid-related diseases	<a href="https://ngdc.cncb.ac.cn/EryDB/home">https://ngdc.cncb.ac.cn/EryDB/home</a>	Xiangdong Fang, Hongzhu Qu, Yiming Bao	[3]
	NeoTCR	An immunoinformatic database for neoantigen-specific TCRs across various cancer subtypes	<a href="http://neotcrdb.bioxai.cn">http://neotcrdb.bioxai.cn</a>	Jian Liu, Kankan Wang	[4]
	HLA typing	An HLA database to search for HLA allele frequencies across populations	<a href="http://bigdata.ibp.ac.cn/HLAtyping">http://bigdata.ibp.ac.cn/HLAtyping</a>	Shumin He, Tao Xu	
	HemaCisDB	A database for analyzing <i>cis</i> -regulatory elements across hematopoietic malignancies	<a href="https://hemacisdb.chinablood.com.cn">https://hemacisdb.chinablood.com.cn</a>	Yuxuan Liu	
	PlateletBase	A knowledgebase for platelet research and disease insights	<a href="http://plateletbase.clinlabomics.org.cn">http://plateletbase.clinlabomics.org.cn</a>	Dongsheng Wang, Jian Huang, Yang Zhang	
Tool	HemaScope	A tool for analyzing single-cell and spatial RNA-seq data of hematopoietic cells	<a href="https://doi.org/10.1093/gpbjnl/qzaf021">https://doi.org/10.1093/gpbjnl/qzaf021</a>	Dong Wang, Yan Huang	
	–	A review of biological data resources and machine learning applications in hematology research	<a href="https://zhenyiwangthu.github.io/HemaScope_Tutorial">https://zhenyiwangthu.github.io/HemaScope_Tutorial</a>	Sajuan Chen, Zhu Chen, Hai Fang, Jianfeng Li	[1]
	UNISOM	A workflow for CHIP detection from WGS and WES data	<a href="https://ngdc.cncb.ac.cn/biocode/tool/7725">https://ngdc.cncb.ac.cn/biocode/tool/7725</a>	Eric W. Klee, Shulan Tian	
	DyNDG	A dynamic network-based model, integrating DEGs to identify leukemia-related genes	<a href="https://github.com/ZhenyiWangTHU/HemaScopeR/">https://github.com/ZhenyiWangTHU/HemaScopeR/</a>	Min Li, Ju Xiang	[9]
	HematoMap	An R-based package for identifying and visualizing leukemia lineage aberrations from single-cell and bulk RNA-seq data	<a href="https://ngdc.cncb.ac.cn/biocode/tool/BT7617">https://ngdc.cncb.ac.cn/biocode/tool/BT7617</a>	Sajuan Chen, Kankan Wang, Xiao-Jian Sun, Hai Fang	[10]
	–	Identified G9a inhibitors as a promising therapeutic avenue for SETD2-mutant leukemia	<a href="https://ngdc.cncb.ac.cn/biocode/tool/7816">https://ngdc.cncb.ac.cn/biocode/tool/7816</a>	Qian-Fei Wang, Gang Huang	[5]
	–	Revealed the role of m <sup>6</sup> A in regulating cell state transition in normal hematopoiesis and leukemogenesis	<a href="https://ngdc.cncb.ac.cn/biocode/tool/DyNDG">https://ngdc.cncb.ac.cn/biocode/tool/DyNDG</a>	Haojian Zhang, Ying Cheng, Fuling Zhou	[6]
	–	Found elevated levels of lactate and H3K18la, as well as their association and roles in T-ALL	<a href="https://ngdc.cncb.ac.cn/biocode/tool/7581">https://ngdc.cncb.ac.cn/biocode/tool/7581</a>	Yu Liu, Yangyang Xie, Meng Yin	[7]
	–	Revealed ITGA2B as a key regulator of heart health in high-altitude settlers	<a href="https://github.com/NRCTM-bioinfo/HematoMap">https://github.com/NRCTM-bioinfo/HematoMap</a>	Yue Gao, Wei Zhou, Haitao Lu	[8]
	–	Identified B-cell clone types and established a clonal threshold specific for each patient clonality profile	<a href="https://doi.org/10.1093/gpbjnl/qzaf030">https://doi.org/10.1093/gpbjnl/qzaf030</a>	María José Terol, Ana-Bárbara García-García	
–	Demonstrated the specificity of genetic regulation of platelet traits in the context of $\beta$ thalassemia	<a href="https://doi.org/10.1093/gpbjnl/qzaf041">https://doi.org/10.1093/gpbjnl/qzaf041</a>	Xiangmin Xu, Zilin Li		
–	Developed a non-invasive liquid biopsy assay for early-stage diagnosis of breast cancer with cfDNA fragmentomics	<a href="https://doi.org/10.1093/gpbjnl/qzaf028">https://doi.org/10.1093/gpbjnl/qzaf028</a>	Jianzhong Su, Zhihua Liu, Xiang Wang		
–	Discovered clonal hematopoietic mutations in plasma cell disorders	<a href="https://doi.org/10.1093/gpbjnl/qzaf027">https://doi.org/10.1093/gpbjnl/qzaf027</a>	Jian Li, Chunyan Sun		

Note: HLA, human leukocyte antigen; DEG, differentially expressed gene; CHIP, clonal hematopoiesis of indeterminate potential; RNA-seq, RNA sequencing; TCR, T-cell receptor; T-ALL, T-cell acute lymphoblastic leukemia; cfDNA, cell-free DNA; WGS, whole-genome sequencing; WES, whole-exome sequencing; AI, artificial intelligence.

modification affects hematopoietic stem cell differentiation and leukemogenesis. They found that ABCD2, as an m<sup>6</sup>A-regulated driver of AML progression, promotes leukemogenesis by modulating fatty acid metabolism and maintaining leukemia cell viability. These findings highlight the central role of RNA epigenetics in stem cell fate and malignant transformation, paving the way for novel epitranscriptomic therapeutic strategies. Wu et al. [7] reported that lactate generated by the Warburg effect induces histone H3K18 lactylation in T-cell acute lymphoblastic leukemia (T-ALL), marking super-enhancer regions and activating oncogenes such as *IGFBP2*. This metabolism–epigenetics–transcription axis redefines the oncogenic role of lactate through chromatin remodeling, and suggests that targeting lactate-driven H3K18 lactylation could represent a promising therapeutic avenue for T-ALL.

Interestingly, one study exemplifies the synergy between traditional Chinese medicine and modern omics technologies in generating actionable translational insights. Wang et al. [8] applied plasma proteomics to investigate high-altitude-induced myocardial injury and identified ITGA2B as a regulator of *IL-6* expression and metabolic reprogramming under hypoxia. Tanshinone IIa, a compound derived from traditional Chinese medicine, reversed ITGA2B-mediated myocardial injury.

Together, these studies underscore the power of integrative omics with detailed clinical phenotyping in decoding hematologic pathogenesis and guiding precision medicine. As the field evolves, sustained collaboration between omics researchers and clinicians remains essential to translate these discoveries into meaningful benefits for patients.

## AI empowering: advancing hematological research and clinical translation

The integration of AI into hematological research has enabled advances in both biological understanding and clinical decision-making. This special issue highlights cutting-edge applications of AI models that decode disease complexity and bridge computational predictions with clinical utility. Two contributions featured here present innovative AI-driven strategies for identifying regulatory drivers and refining disease classification.

A et al. [9] developed the DyNDG model, which integrates machine learning and dynamic network modeling. By constructing a time-series multilayer network and applying a random-walk-based propagation framework, DyNDG captures temporal changes in gene interactions and identifies leukemia-related genes with higher accuracy. This strategy enhances the discovery of stage-specific biomarkers and emphasizes the value of temporal dynamics in gene prioritization. Dai et al. [10] presented HematoMap, a platform combining AI and single-cell omics to quantify lineage aberrancy and infer leukemic origins. Using cosine similarity and LASSO regression, HematoMap maps transcriptional deviations from bulk RNA-seq data. This enables personalized risk stratification and links molecular subtypes to clinical outcomes, demonstrating the broad applicability of AI in translational hematology.

In summary, these studies exemplify the dual impact of AI: resolving biological complexity through advanced pattern recognition and promoting clinical innovation through predictive modeling.

## Outlook: a new era of mechanistic discoveries and intelligent decision-making

This special issue reflects rapid progress in hematology, driven by multi-omics data resources and integration as well as computational tools. Future progress will rely on three pillars: standardized data curation, interpretable intelligent algorithms, and deep investigation of biological mechanisms. To promote openness and reproducibility, we strongly encourage contributors to share their datasets and analytical tools through national infrastructure platforms like the National Genomics Data Center (NGDC) of China. Such efforts will promote data interoperability and foster collaborative innovation across the hematology research community.

## CRedit author statement

**Fuhong He:** Conceptualization, Writing – original draft, Writing – review & editing. **Zhaojun Zhang:** Conceptualization, Writing – original draft, Writing – review & editing. **Xiangdong Fang:** Conceptualization, Writing – review & editing. **Qian-Fei Wang:** Conceptualization, Writing – review & editing. All authors have read and approved the final manuscript.

## Competing interests

The authors have declared no competing interests.

## Acknowledgments

This work was supported by the National Key R&D Program of China (Grant Nos. 2022YFF1202004, 2022YFC2503304, and 2022YFC2406803), the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA0460403), the National Natural Science Foundation of China (Grant Nos. 82270126 and 81870097), the Beijing Natural Science Foundation (Grant No. 7252091). We thank all authors, reviewers, and editors for their contributions to this special issue.

## ORCID

0000-0002-3176-8309 (Fuhong He)  
 0000-0003-0490-6507 (Zhaojun Zhang)  
 0000-0002-6628-8620 (Xiangdong Fang)  
 0000-0002-0086-2626 (Qian-Fei Wang)

## References

- [1] Wang Z, Miao Y, Li H, Cheng W, Shi M, Lv G, et al. HemaScope: a tool for analyzing single-cell and spatial transcriptomics data of hematopoietic cells. *Genomics Proteomics Bioinformatics* 2025;23:qzaf002.
- [2] Kang Z, Zhu T, Zou D, Liu M, Zhang Y, Wang L, et al. HemAtlas: a multi-omics hematopoiesis database. *Genomics Proteomics Bioinformatics* 2025;23:qzaf026.
- [3] Zheng G, Wu S, Zhang Z, Xin Z, Zhang L, Zhao S, et al. EryDB: a transcriptomic profile database for erythropoiesis and erythroid-related diseases. *Genomics Proteomics Bioinformatics* 2025;23:qzae029.

- [4] Zhou W, Xiang W, Yu J, Ruan Z, Pan Y, Wang K, et al. NeoTCR: an immunoinformatic database of experimentally-supported functional neoantigen-specific TCR sequences. *Genomics Proteomics Bioinformatics* 2025;23:qzae010.
- [5] Zhang Y, Xia M, Yi Z, Sui P, He X, Wang L, et al. Integrated computational and functional screening identify G9a inhibitors for *SETD2*-mutant leukemia. *Genomics Proteomics Bioinformatics* 2025;23:qzaf035.
- [6] Liu W, Wang Y, Yao S, Han G, Hu J, Yin R, et al. Reprogramming of RNA m<sup>6</sup>A modification is required for acute myeloid leukemia development. *Genomics Proteomics Bioinformatics* 2025;23:qzae049.
- [7] Wu W, Zhang J, Sun H, Wu X, Wang H, Cui B, et al. Glycolysis induces abnormal transcription through histone lactylation in T-lineage acute lymphoblastic leukemia. *Genomics Proteomics Bioinformatics* 2025;23:qzaf029.
- [8] Wang Y, Shen P, Wu Z, Tu B, Zhang C, Zhou Y, et al. Plasma proteomic profiling reveals ITGA2B as a key regulator of heart health in high-altitude settlers. *Genomics Proteomics Bioinformatics* 2025;23:qzaf030.
- [9] A J, Xiang J, Meng X, Sheng Y, Peng H, Li M. Identifying leukemia-related genes based on time-series dynamic network by integrating differential genes. *Genomics Proteomics Bioinformatics* 2025;23:qzaf037.
- [10] Dai Y, Ouyang W, Jin W, Zhang F, Cheng W, Li J, et al. Resolving leukemia heterogeneity and lineage aberrations with HematoMap. *Genomics Proteomics Bioinformatics* 2025; 23:qzaf005.