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.

Integrative transcriptomic profiling reveals convergent coding and non-coding signatures in pediatric bone marrow failure syndromes

Jaeeun Yoo
The Catholic University of Korea
Jae Wook Lee
The Catholic University of Korea
Nack-Gyun Chung
The Catholic University of Korea
Jae Won Yoo
The Catholic University of Korea
Yonggoo Kim
The Catholic University of Korea
Myungshin Kim
The Catholic University of Korea
Hyunjoo Bae
The Catholic University of Korea
Jong-Mi Lee
jongmi1226@catholic.ac.kr
The Catholic University of Korea

Article

Keywords:

Posted Date: April 7th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9165159/v1

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

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Scientific Reports on August 27th, 2026. See the published version at https://doi.org/10.1038/s41598-026-68248-5.

Abstract

Pediatric bone marrow failure (BMF) comprises heterogeneous inherited and immune-mediated disorders, yet many cases remain genetically unclassified despite comprehensive genomic evaluation. To identify shared transcriptional features beyond protein-coding variation, we performed total RNA sequencing on bone marrow samples from pediatric BMF patients (n=12) and controls (n=2). Based on whole-genome sequencing, patients were stratified into genetically defined BMF (g-BMF, n=4) and genetically undefined BMF (u-BMF, n=8).

Comparative transcriptomic analyses revealed minimal differences in protein-coding genes and long non-coding RNAs (lncRNAs) between g-BMF and u-BMF, indicating strong transcriptional concordance. Compared with controls, 55.2% of differentially expressed protein-coding genes were shared between g-BMF and u-BMF and were predominantly upregulated. Pathway enrichment analyses consistently identified immune activation and cellular stress-response pathways, suggesting a shared inflammatory transcriptional state across pediatric BMF irrespective of genetic classification.

LncRNA profiling showed that 69.6% of differentially expressed lncRNAs were common to both BMF groups. *ATP1A1-AS1, USP3-AS1*, and *SNHG32* were reproducibly overexpressed and validated by RT–qPCR. Correlation-based pathway analyses associated *ATP1A1-AS1* and *USP3-AS1* with immune-related and biosynthetic programs, whereas *SNHG32*showed a distinct co-expression pattern.

Collectively, these findings indicate shared coding and non-coding transcriptional features in pediatric BMF that appear largely independent of genetic classification.

Introduction

Bone marrow failure (BMF) syndromes are rare but clinically significant disorders characterized by uni- or multi-lineage cytopenias and an increased risk of progression to myelodysplastic syndromes, leukemia, and other severe complications [1–3]. Advances in genomic technologies have identified a wide spectrum of inherited BMF syndromes caused by pathogenic germline variants across more than a hundred genes involved in DNA repair, telomere maintenance, and ribosome biogenesis [4,5].

Despite advances in comprehensive genomic profiling, approximately half of clinically suspected BMF cases remain genetically unexplained [6–8]. This gap in diagnostic yield is particularly notable in pediatric patients, in whom inherited forms of BMF are more prevalent than acquired etiologies [9]. These considerations suggest that pathogenic mechanisms beyond protein-coding mutations are involved in disease development [10,11]. Supporting this hypothesis, a recent transcriptomic study had shown shared expression patterns across different BMF subgroups, regardless of the presence or absence of identifiable genetic defects [12]. This observation raises the possibility that the dysregulation of non-coding elements plays a pivotal role in disease etiology.

Among these, long non-coding RNAs (lncRNAs) have attracted particular attention for their versatile regulatory roles in gene expression. Defined as transcripts longer than 200 nucleotides with little or no protein-coding potential, lncRNAs modulate signaling networks governing the proliferation, differentiation, and survival of hematopoietic cells. Increasing evidence suggests that their dysregulation underlies or reflects the mechanisms involved in normal and malignant hematopoiesis [13–16]. A recent review further highlighted that lncRNAs regulate immune cell specification, lineage determination during hematopoiesis, and immune activation — mechanisms that are centrally disrupted in BMF.[17]

Nevertheless, the contribution of lncRNAs to the pathobiology of pediatric BMF remains poorly understood. To address this gap, we performed integrative transcriptomic profiling encompassing mRNA and lncRNA expression. By comparing transcriptional landscapes across genetically defined and undefined BMF subgroups, the aim of this study was to uncover BMF-specific coding and non-coding RNA signatures and identify lncRNA-associated regulatory pathways, that may offer insights into the pathogenesis of BMF.

Materials and methods

Patients and samples

Bone marrow samples were obtained at the time of first clinical evaluation from 12 pediatric patients presenting with persistent cytopenia and bone marrow hypocellularity for their age. All patients underwent whole-genome sequencing to identify pathogenic germline variants associated with inherited BMF syndromes. Three patients were genetically diagnosed with Fanconi anemia and one with dyskeratosis congenita, whereas the remaining eight showed no disease-causing variants.

Two additional control samples were obtained from age-matched children exhibiting normal bone marrow morphology. Accordingly, the study cohort included four patients with genetically defined BMF (g-BMF), eight with genetically undefined BMF (u-BMF), and two controls.

This study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board (IRB) of Seoul St. Mary's Hospital (approval number: KC24SISI0553). Written informed consent was obtained from all participants or their legal guardians prior to inclusion in the study.

Sample collection, RNA isolation, and library preparation

Total RNA was extracted from bone marrow aspirates using the High Pure RNA Isolation kit (Roche, Mannheim, Germany). Libraries were prepared using the TruSeq Stranded Total RNA Library Prep Globin Kit with Ribo-Zero rRNA removal (Illumina, San Diego, CA, USA), with 800 ng of total RNA as input. Sequencing was performed on an Illumina NovaSeq-6000 sequencer (Illumina, San Diego, CA, USA). The sequencing data were analyzed using the CLC

Genomics Workbench 20.0.4 (Qiagen, Hilden, Germany). Raw sequencing data are deposited in the NCBI SRA repository under accession number PRJNA1406476.

RNA sequencing data processing and differential expression analysis

Data were processed through a comprehensive RNA-seq analysis pipeline implemented using the Snakemake workflow management system (v7.32.4) [18]. Preprocessing and quality control were performed using FastQC (v0.11.9) [19] and MultiQC (v1.14) [20], followed by read trimming with quality-based filtering when necessary. Reads were aligned using STAR (v2.6.1d) [21] against the human genome reference (hg38/GRCh38) and gene quantification was performed using feature Counts from the Subread package (v2.0.3) [22] with Gencode v44 annotation in Ensembl format.

Differential expression analysis was conducted using DESeq2 (v1.40.0) [23] with raw count data. Variance stabilizing transformation was applied for downstream analysis and visualization. Statistical significance was determined using thresholds of a false discovery rate (FDR)-adjusted p-value of less than 0.05 and absolute log2 fold change of greater than 1.0, with minimum read count filtering of 10 reads per gene applied prior to analysis.

Differentially expressed genes were subjected to pathway enrichment analysis using the Metascape web tool (https://metascape.org), which integrates multiple ontology sources, including Gene Ontology (GO) biological processes, KEGG pathways, and Reactome gene sets. The results were visualized as a heatmap showing -log10 (P-value) for each enriched term across different sample groups.

Selection strategy for lncRNAs of interest

Among the differentially expressed lncRNAs (DE-lncRNAs), we focused on those that were consistently dysregulated in both g-BMF and u-BMF groups compared with controls. To ensure biological relevance and technical robustness, we selected DE-lncRNAs that were annotated in NCBI and exhibited sufficient expression levels, defined as an average CPM of ≥ 10 and CPM ≥ 1 across all 14 samples. CPM was used for expression-level filtering and co-expression analyses as it provides a library-size-normalized metric suitable for comparing relative expression levels across samples; DESeq2's internal normalization was applied separately for differential expression analysis.

Pathway enrichment and co-expression network analysis of selected lncRNAs

To assess the functional relevance of the selected lncRNAs, we performed pathway enrichment and co-expression analyses. Spearman's rank correlation coefficients (ρ-values) were calculated between the selected DE-lncRNAs and all protein-coding genes. These correlation-ranked gene lists were used as the input for pre-ranked gene set enrichment analysis (GSEA) using the Hallmark gene sets from MSigDB (version 2023.1). For each lncRNA, genes were ranked by correlation coefficients (ρ). Pre-ranked GSEA was performed with 1,000 permutations. Enrichment was evaluated using normalized enrichment scores, nominal p-values, and FDRs (q-values), with pathways considered significantly enriched at a FDR < 0.25.

To construct the lncRNA–mRNA co-expression network, Pearson correlation coefficients were calculated between the expression levels (CPM values) of each lncRNA and mRNA pair. Gene pairs with $|r| > 0.4$, $P < 0.05$, and FDR < 0.05 (Benjamini-Hochberg correction) were considered significantly co-expressed. For visualization, the top 25 mRNAs with the highest absolute correlation for each lncRNA were retained. The threshold of $|r| > 0.4$ was selected to balance network connectivity with statistical stringency in the context of a small cohort; given the limited sample size, more stringent thresholds would substantially reduce network coverage and potentially obscure biologically relevant co-expression relationships.

Validation using real-time quantitative polymerase chain reaction (RT-qPCR)

RT-qPCR validation of the selected DE-lncRNAs was conducted using the cDNA samples from the original cohort. Primers were designed to target the common exon junctions if there were more than one ENST transcript. Primer sequences and targeted transcript isoforms for each lncRNA assay are provided in sTable 11. RT-qPCR results were analyzed using the ΔΔCt method. Statistical analysis was performed using a Wilcoxon rank-sum test. Differences with $P < 0.05$ were regarded as significant.

Results

Patients characteristics

Baseline characteristics are summarized in Table 1 and sTable 1. The median age at first hematological presentation was 6.5 years old. Among the three patients with Fanconi anemia, one carried a homozygous *FANCG* splice-site mutation (c.307 + 1G > C), one carried digenic pathogenic variants involving *FANCD2* (c.3817C > T) and *FANCD1* (c.5795_5799del), and one carried compound heterozygous mutations in *FANCA* (c.3520_3522del, c.4168-2A > G). The patient with dyskeratosis congenita harbored a heterozygous *TERT* missense variant (c.1892G > A). Detailed clinical presentations are described in the supplementary information.

In the u-BMF group, four of eight patients had either a family history of hematological abnormalities (n = 3) or extra-hematopoietic manifestations (n = 2). Four patients in this group underwent hematopoietic stem cell transplantation.

Differential gene expression between g-BMF and u-BMF

In total, 62,653 genes were detected across the entire cohort based on Ensembl gene IDs, comprising 19,911 (32%) protein-coding genes and 19,242 (31%) lncRNAs (sFigure 1). Principal component analysis of global transcriptomic profiles demonstrated a separation between BMF cases and controls, whereas g-BMF and u-BMF samples exhibited considerable overlap and intra-group heterogeneity (sFigure 2).

Differential expression analysis between g-BMF and u-BMF identified only a small number of differentially expressed transcripts, including four protein-coding genes (*SFT2D3, OR52K1, LRRC24*, and *TEN1-CDK3*) and four lncRNAs (*lnc-XPR1-26, lnc-GABRB1-22, ENSG00000289145*, and *lnc-PKM-65*) (sTable 2, 3). These results suggest that g-BMF and u-BMF share largely similar transcriptional profiles, with only subtle molecular differences at the gene expression level.

Protein-coding gene signatures of g-BMF and u-BMF

Given the limited transcriptomic divergence between g-BMF and u-BMF, we next compared each subgroup with controls to delineate disease-associated transcriptional signatures.

In total, 2,078 and 1,315 protein-coding genes were differentially expressed in g-BMF and u-BMF, respectively. The majority of differentially expressed protein-coding genes were upregulated, accounting for 72.2% (1,501/2,078) in g-BMF and 91.3% (1,201/1,315) in u-BMF, suggesting a global shift toward transcriptional activation in both subgroups (Fig. 1A, B, sTable 4, 5).

More than half of dysregulated protein-coding genes (55.2%, 1,207/2,186 total unique DE genes) was shared between the groups (Fig. 1C). In particular, 70.5% of upregulated differentially expressed protein-coding genes were shared. These results suggest that pediatric BMF exhibits a common transcriptomic landscape characterized by the widespread upregulation of gene expression and highly similar transcriptomic profiles irrespective of genetic etiology, consistent with a previous report [12].

Subsequent GO and pathway enrichment analysis revealed distinct functional trends between up- and down-regulated gene sets. Upregulated transcripts in both groups were strongly enriched in immune activation and stress-response pathways, suggesting the transcriptional activation of inflammatory processes. Conversely, the downregulated gene sets showed no distinct enrichment pattern and largely overlapped with pathways enriched in the upregulated genes, suggesting the absence of specific functional suppression signatures in BMF (Fig. 1D, sFigure 3, sTable 6).

LncRNA signatures of g-BMF and u-BMF

We next focused on lncRNAs to identify non-coding transcriptomic signatures specific to pediatric BMF. Among 19,242 detected lncRNAs, 13,110 were annotated in the NCBI database. Using the same differential expression criteria applied to protein-coding genes, we identified 1,167 and 992 DE-lncRNAs in g-BMF and u-BMF compared with controls, respectively (Fig. 2A, B, sTable 7–8). Of these, 69.6% (886/1,273) DE-lncRNAs were shared between g-BMF and u-BMF. Among the shared DE-lncRNAs, 11 candidates were prioritized based on consistent expression across all samples: *ATP1A1-AS1, LINC01036, HCG11, HCP5, SNHG32, PSMB8-AS1, TAGAP-AS1, MALAT1, FAM30A, USP3-AS1*, and *MIR22HG* (sTable 9). All of these lncRNAs were upregulated compared with controls, except for *MALAT1*, which was downregulated (Fig. 2C).

To explore the biological relevance of these lncRNAs, we performed Spearman's correlation-based pre-ranked GSEA using Hallmark gene sets (sTable 10). Recurrent enrichment of immune-related pathways, such as interferon responses and TNFA signaling via NF-κB, and metabolic processes, including glycolysis, heme metabolism, and oxidative phosphorylation, was observed across multiple lncRNAs (Fig. 2C). These results support a role for the selected lncRNAs in immune activation and metabolic remodeling.

RT-qPCR validation showed that *ATP1A1-AS1* ($P = 0.028$), *USP3-AS1* ($P = 0.044$), and *SNHG32* ($P = 0.022$) were significantly overexpressed in BMF compared with controls (Fig. 2D, sFigure 4, sTable 11).

Notably, *ATP1A1-AS1* and *USP3-AS1* exhibited highly similar enrichment profiles, including positive associations with multiple immune and signaling-related pathways, including IL2–STAT5 signaling, PI3K–AKT–mTOR signaling, TNFα signaling via NF-κB, and interferon response pathways. In contrast, these lncRNAs demonstrated coordinated negative correlations with gene sets associated with cell cycle progression (E2F targets), MYC-driven transcription, and heme metabolism. These two lncRNAs also showed partially overlapping sets of co-expressed mRNAs, suggesting that they are involved in related transcriptional programs. In contrast, *SNHG32* displayed a distinct co-expression network with the other two lncRNAs, indicating a more divergent regulatory role (Fig. 2E).

Discussion

Despite extensive genomic efforts, nearly half of pediatric BMF cases with a clinical suspicion of inherited syndromes remain genetically unexplained. This diagnostic gap raises the possibility that pathogenic mechanisms beyond protein-coding gene mutations may contribute to disease development. To systematically evaluate the contribution of known genetic defects, we performed whole-genome sequencing in all patients and stratified the cohort into genetically defined and undefined groups (g-BMF vs. u-BMF) based on the presence or absence of pathogenic or likely pathogenic variants previously implicated in BMF syndromes.

Transcriptomic profiling of bone marrow samples revealed that g-BMF and u-BMF groups shared convergent expression signatures characterized by global transcriptional upregulation, consistent with previous findings by Walne et al. [12]. This widespread activation, accompanied by the enrichment of immune- and stress-response pathways, suggests that inflammatory signaling represents a common pathogenic axis in pediatric BMF, regardless of underlying genetic etiology. This result not only aligns with known mechanisms of T-cell-mediated immune attack in acquired BMF [2,24] but also reinforces emerging evidence implicating chronic inflammation in the pathogenesis and progression of inherited BMF syndromes [25-28]. Collectively, these results suggest that inflammatory signaling represents a common pathogenic denominator across diverse BMF subtypes.

We next explored whether non-coding regulators—particularly lncRNAs—contribute to transcriptomic remodeling. Similar to mRNA expression patterns, lncRNA profiles were highly concordant between g-BMF and u-BMF groups, suggesting that non-coding transcriptional programs reflect shared biological processes. Despite their typically low expression in the cellular level, *ATP1A1-AS1*, *USP3-AS1*, and *SNHG32* were reproducibly validated using RT-qPCR. Although these lncRNAs have been previously reported as prognostic markers in various solid tumors, their functional roles in bone marrow failure remain largely unknown [29-34]. Specifically, ATP1A1-AS1 is an antisense RNA to the Na^+/K^+-ATPase gene ATP1A1, previously implicated in ion transport regulation and tumor progression[35]. USP3-AS1 is antisense to USP3, a deubiquitinase involved in DNA damage response and histone H2B monoubiquitination[36] — a pathway directly relevant to Fanconi anemia pathobiology[37]. SNHG32, a small nucleolar RNA host gene, has been associated with ribosome biogenesis and translational regulation[38]. Their consistent upregulation across both BMF subgroups in this study warrants functional investigation in the hematopoietic context. Notably, lncRNA dysregulation in the genetically defined group is particularly compelling in the context of the dyskeratosis congenita patient, whose TERT mutation directly impairs the telomerase ribonucleoprotein complex — a molecular machine dependent on a non-coding RNA component (TERC) for its function[39] — lending further biological plausibility to the observed non-coding transcriptional alterations in this study.

Using a guilt-by-association framework [40], *ATP1A1-AS1* and *USP3-AS1* demonstrated highly similar enrichment profiles and co-expression networks. Both lncRNAs showed positive correlations across multiple immune- and signaling-related pathways, while exhibiting coordinated negative correlations with gene sets associated with cell cycle progression, MYC-driven transcription, DNA repair, and metabolic processes.This pattern may reflect a transcriptional trade-off between inflammatory activation and proliferative restraint, as previously described in stress-prone hematopoietic environments [41,42]. In contrast, *SNHG32* showed a distinct expression signature marked by the co-activation of immune and biosynthetic pathways, implying a role in coordinating stress-induced transcriptional programs beyond the inflammatory response alone.

This study has some limitations. The rarity of pediatric BMF constrains sample availability, and functional interrogation of the identified lncRNAs remains to be undertaken. Furthermore, RT-qPCR validation was performed on the same discovery cohort, which does not constitute independent validation; Nevertheless, this study represents an early transcriptomic effort to jointly profile coding and non-coding RNAs from pediatric bone marrow. Our results offer a foundational dataset that supports further investigation into shared molecular mechanisms and lay the groundwork for future studies aimed at elucidating the pathogenic role of non-coding RNA in pediatric BMF.

Declarations

Author Contributions

JY conceptualized and directed the project, collected and analyzed data, interpreted the findings, and drafted the manuscript. JML initiated and designed the study and oversaw all aspects of its execution. All authors were engaged in discussing the results and contributed to the refinement of the final manuscript.

Data availability statement

The variant data supporting this study's findings are included within the supplementary table. The raw sequencing data will be openly available in the NCBI SRA repository under accession number PRJNA1406476. Data not available in the repository are available on reasonable request from the corresponding author, J.M.L.

Funding statement

This work was supported by a grant from the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education (RS-2023-00244625 to J-M.L.) and the Research Fund of Seoul St. Mary's Hospital, The Catholic University of Korea.

Conflict of interest disclosure

No conflicts of interest declared.

Ethics approval statement

This study was approved by the Institutional Review Board (IRB) of Seoul St. Mary's Hospital (approval number: KC24SISI0553).

Permission to reproduce material from other sources

Permission to use KEGG pathway content in Figure 1D was obtained from Kanehisa Laboratories.

Clinical trial registration

Not applicable.

References

1. Alter, B. P. et al. Malignancies and survival patterns in the National Cancer Institute inherited bone marrow failure syndromes cohort study. *Br. J. Haematol.* **150**, 179–188 (2010).
2. Young, N. S. Aplastic anemia. *N. Engl. J. Med.* **379**, 1643–1656 (2018).
3. Shimamura, A. & Alter, B. P. Pathophysiology and management of inherited bone marrow failure syndromes. *Blood Rev.* **24**, 101–122 (2010).

4. Bluteau, O. et al. A landscape of germ line mutations in a cohort of inherited bone marrow failure patients. *Blood* **131**, 717–732 (2018).
5. Dokal, I., Tummala, H. & Vulliamy, T. Inherited bone marrow failure in the pediatric patient. *Blood* **140**, 556–570 (2022).
6. Muramatsu, H. et al. Clinical utility of next-generation sequencing for inherited bone marrow failure syndromes. *Genet. Med.* **19**, 796–802 (2017).
7. Lee, J. M. et al. Genomic insights into inherited bone marrow failure syndromes in a Korean population. *Br. J. Haematol.* **205**, 1581–1589 (2024).
8. Blombery, P. et al. Utility of clinical comprehensive genomic characterization for diagnostic categorization in patients presenting with hypocellular bone marrow failure syndromes. *Haematologica* **106**, 64 (2020).
9. Alter, B. P. Diagnosis, genetics, and management of inherited bone marrow failure syndromes. *Hematology Am. Soc. Hematol. Educ. Program*, 29–39 (2007).
10. Herman, A. B., Tsitsipatis, D. & Gorospe, M. Integrated lncRNA function upon genomic and epigenomic regulation. *Mol. Cell* **82**, 2252–2266 (2022).
11. Wilkes, M. C., Repellin, C. E. & Sakamoto, K. M. Beyond mRNA: the role of non-coding RNAs in normal and aberrant hematopoiesis. *Mol. Genet. Metab.* **122**, 28–38 (2017).
12. Walne, A. J. et al. Genome-wide whole-blood transcriptome profiling across inherited bone marrow failure subtypes. *Blood Adv.* **5**, 5360–5371 (2021).
13. Lu, S., Song, X., Chen, J. & Qiao, X. Identification of differentially expressed lncRNAs and mRNAs in children with acquired aplastic anemia by RNA sequencing. *Biomed. Res. Int.* **2020**, 8962090 (2020).
14. Morlando, M., Ballarino, M. & Fatica, A. Long non-coding RNAs: new players in hematopoiesis and leukemia. *Front. Med.* **2**, 23 (2015).
15. Luo, M. et al. Long non-coding RNAs control hematopoietic stem cell function. *Cell Stem Cell* **16**, 426–438 (2015).
16. Schwarzer, A. et al. The non-coding RNA landscape of human hematopoiesis and leukemia. *Nat. Commun.* **8**, 1–17 (2017).
17. Chang, H. Y. & Yu, H. Long noncoding RNA regulation of immunity. *Nat. Immunol.* **27**, https://doi.org/10.1038/s41590-025-02355-9 (2026).
18. Molder, F. et al. Sustainable data analysis with Snakemake. *F1000Res.* **10**, 33 (2021).
19. Simon Andrews, F., Segonds-Pichon, A., Biggins, L., Krueger, C. & Wingett, S. FastQC: a quality control tool for high throughput sequence data (2010).
20. Ewels, P., Magnusson, M., Lundin, S. & Kaller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
21. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
22. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
23. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
24. Medinger, M., Drexler, B., Lengerke, C. & Passweg, J. Pathogenesis of acquired aplastic anemia and the role of the bone marrow microenvironment. *Front. Oncol.* **8**, 587 (2018).
25. Briot, D., Mace-Aime, G., Subra, F. & Rosselli, F. Aberrant activation of stress-response pathways leads to TNF-alpha oversecretion in Fanconi anemia. *Blood* **111**, 1913–1923 (2008).
26. Saadatzadeh, M. R., Bijangi-Vishehsaraei, K., Kapur, R. & Haneline, L. S. Distinct roles of stress-activated protein kinases in Fanconi anemia-type C-deficient hematopoiesis. *Blood* **113**, 2655–2660 (2009).
27. Neoman, N., Kim, H. N., Viduya, J., Goyal, A., Liu, Y. L. & Sakamoto, K. M. Inflammatory pathways and the bone marrow microenvironment in inherited bone marrow failure syndromes. *Stem Cells* **43** (2025).
28. Repczynska, A., Ciastek, B. & Haus, O. New insights into the Fanconi anemia pathogenesis: a crosstalk between inflammation and oxidative stress. *Int. J. Mol. Sci.* **25** (2024).
29. Xu, S. et al. Cuproptosis-associated lncRNA establishes new prognostic profile and predicts immunotherapy response in clear cell renal cell carcinoma. *Front. Genet.* **13**, 938259 (2022).
30. Rao, Y., Liu, H., Yan, X. & Wang, J. In silico analysis identifies differently expressed lncRNAs as novel biomarkers for the prognosis of thyroid cancer. *Comput. Math. Methods Med.* **2020**, 3651051 (2020).
31. Xu, S., Sui, J., Yang, S., Liu, Y., Wang, Y. & Liang, G. Integrative analysis of competing endogenous RNA network focusing on long noncoding RNA associated with progression of cutaneous melanoma. *Cancer Med.* **7**, 1019–1029 (2018).
32. Kyung, J., Kim, M., Shin, H. R., Kim, E. & Oh, H. J. Multi-layered gene regulation by long non-coding RNAs: from chromatin to genome architecture. *BMB Rep.* (2025).
33. Song, P., Han, R. & Yang, F. Super enhancer lncRNAs: a novel hallmark in cancer. *Cell Commun. Signal.* **22**, 207 (2024).
34. Chodary Khameneh, S., Razi, S., Shamdani, S., Uzan, G. & Naserian, S. Weighted correlation network analysis revealed novel long non-coding RNAs for colorectal cancer. *Sci. Rep.* **12**, 2990 (2022).
35. Sumiyoshi, S. et al. Functional analysis and clinical importance of ATP1A1 in colon cancer. *Ann. Surg. Oncol.* **30**, 6898–6910 (2023).
36. Nicassio, F. et al. Human USP3 is a chromatin modifier required for S phase progression and genome stability. *Curr. Biol.* **17**, 1972–1977 (2007).
37. Kee, Y. & D'Andrea, A. D. Expanded roles of the Fanconi anemia pathway in preserving genomic stability. *Genes Dev.* **24**, 1680–1694 (2010).
38. Ojha, S., Malla, S. & Lyons, S. M. snoRNPs: functions in ribosome biogenesis. *Biomolecules* **10**, 783 (2020).
39. Vulliamy, T. et al. TERC and TERT gene mutations in patients with bone marrow failure and the significance of telomere length measurements. *Blood* **113**, 309–316 (2009).
40. Rinn, J. L. & Chang, H. Y. Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **81**, 145–166 (2012).

41. Pietras, E. M. Inflammation: a key regulator of hematopoietic stem cell fate in health and disease. *Blood* **130**, 1693–1698 (2017).
42. Olson, O. C., Kang, Y.-A. & Passegué, E. Normal hematopoiesis is a balancing act of self-renewal and regeneration. *Cold Spring Harb. Perspect. Med.* **10**, a035519 (2020).
43. Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y. & Ishiguro-Watanabe, M. KEGG: biological systems database as a model of the real world. *Nucleic Acids Res.* **53**, D672–D677 (2025).
44. Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**, 1947–1951 (2019).
45. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).

Table

Table 1. Baseline characteristics of the patients

ID	Group	Gender/age	Etiology	Bone marrow cellularity	Cytogenetics	Chromosome breakage test	Hg (g/dl)	ANC x10^9/L	ALC x10^9/L	ARC x10^9/L	PLT x10^9/L	Treatment
G1	g-BMF	M/3	FA	20%	46,XY[20]	Not tested	8.3	4.7	3.1	31.5	20.0	HSCT
G2	g-BMF	F/7	FA	<5%	46,XX[20]	2.17 (>0.5)	11.0	3.3	1.9	73.9	87.0	waiting for
G3	g-BMF	M/8	FA	<5%	46,XY[20]	Not tested	7.7	4.3	3.9	39.2	21.0	HSCT
G4	g-BMF	M/11	DKC	<5%	46,XY[20]	Not tested	9.2	4.3	2.0	84.0	15.0	HSCT
U1	u-BMF	F/8	Idiopathic	30%	46,XX[20]	Not tested	9.4	2.7	1.8	100.7	46.0	Monitoring
U2	u-BMF	F/11	Idiopathic	10%	46,XX[20]	0.25	8.4	3.16	2.1	79.4	12.0	HSCT
U3	u-BMF	M/17	Idiopathic	5%	46,XY[20]	0.27	8.9	2.59	1.3	59.3	19.0	HSCT
U4	u-BMF	F/15	Idiopathic	5%	46,XX[20]	0.46	7.6	2.51	1.7	49.5	38.0	HSCT
U5	u-BMF	M/3	Idiopathic	10%	46,XY[20]	Not tested	9.8	6.45	4.4	66.5	51.0	Monitoring
U6	u-BMF	M/8	Idiopathic	20%	Not tested	Not tested	9.4	4.35	3.0	68.6	24.0	Monitoring
U7	u-BMF	M/18	Idiopathic	<5%	46,XY[12]	0.17	8.6	4.09	6.9	47.3	16.0	HSCT
U8	u-BMF	M/11	Idiopathic	5%	46,XY[20]	0.17	12.4	3.37	1.9	87.1	84.0	Monitoring
C1	Control*	M/1	-	40%	46,XY[20]	Not tested	8.2	11.16	6.4	143.2	390.0	
C2	Control*	M/1	-	60%	46,XY[20]	Not tested	8.6	6.11	4.1	223.5	120.0	

Hg, hemoglobin; ANC, absolute neutrophil count; ALC, absolute lymphocyte count; ARC, absolute reticulocyte count; PLT, platelet; g-BMF, genetically defined bone marrow failure; u-BMF, genetically undefined bone marrow failure; M, male; F, female; FA, Fanconi anemia; DKC, dyskeratosis congenita; HSCT, hematopoietic stem cell transplantation. *Control samples were obtained from patients with suspected iron deficiency anemia (C1) and hemolytic uremic syndrome (C2), all of whom showed normal bone marrow morphology.

Figures

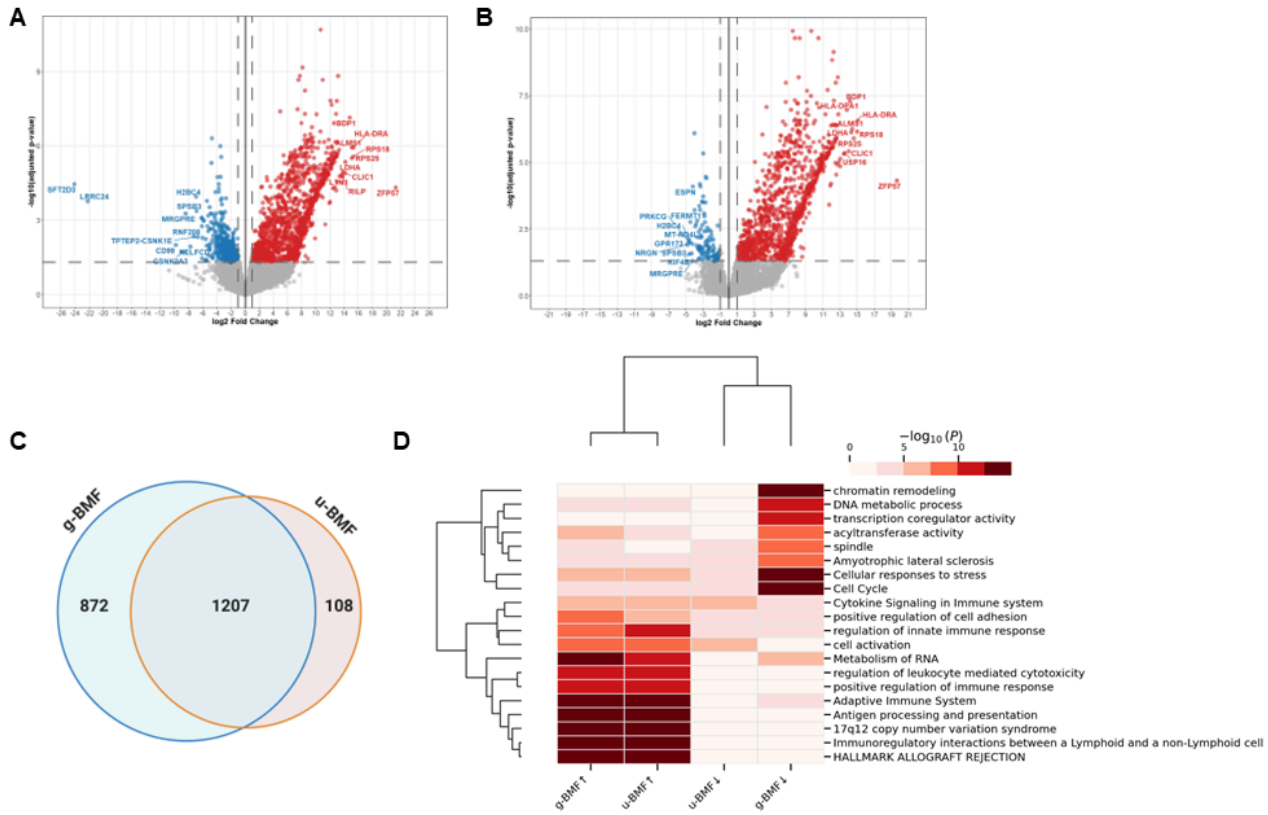


Figure 1

Figure 1

Shared protein-coding transcriptomic signatures in pediatric bone marrow failure. Global protein-coding gene expression landscape and functional pathways that are commonly dysregulated in genetically defined BMF (g-BMF) and genetically undefined BMF (u-BMF). (A, B) Volcano plots display protein-coding genes that are differentially expressed in g-BMF and u-BMF compared with controls ($|\log_2$ fold change$| > 1$, FDR < 0.05). (C) A Venn diagram illustrates the substantial overlap of differentially expressed protein-coding genes between g-BMF and u-BMF. (D) A heatmap of pathway enrichment demonstrates concordant activation of immune- and stress-response pathways across both BMF subtypes. Pathway terms were sourced in part from the KEGG database [43-45].

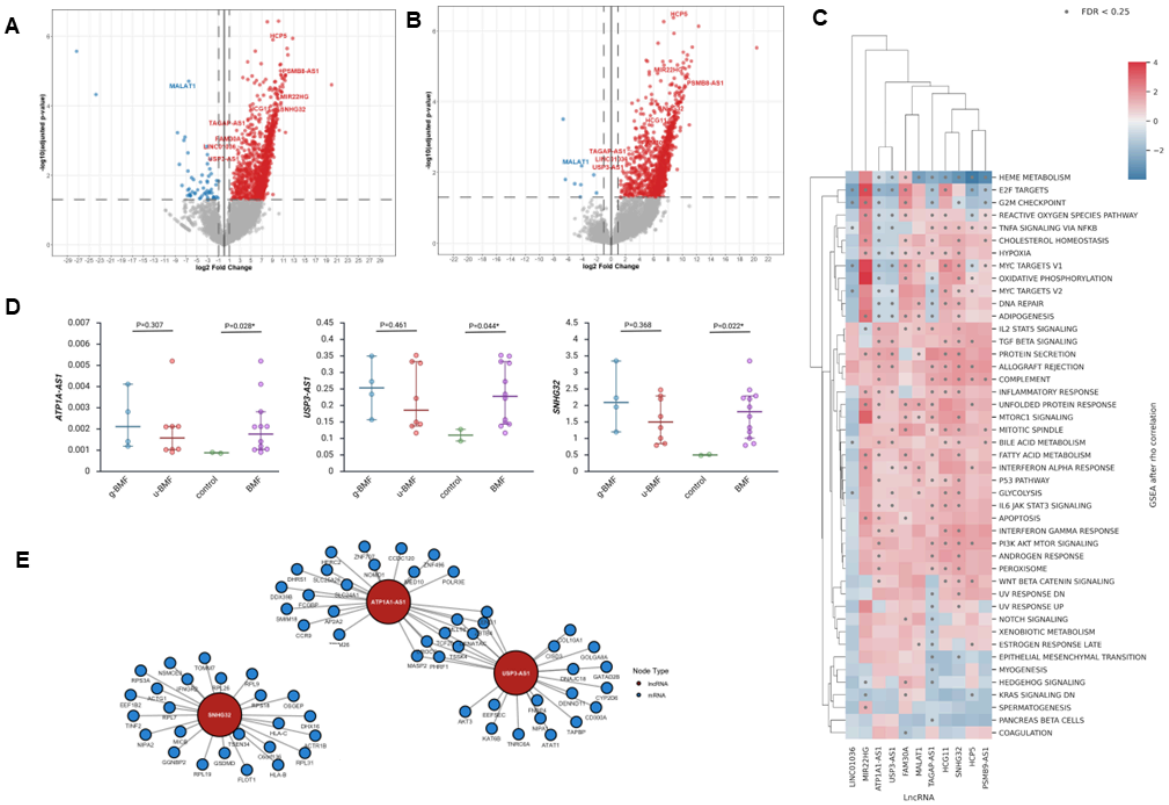


Figure 2

Figure 2

Shared lncRNA dysregulation and associated pathway signatures in pediatric bone marrow failure. Differential expression and functional associations of long non-coding RNAs (lncRNAs) in genetically defined BMF (g-BMF) and genetically undefined BMF (u-BMF) pediatric BMF. (A, B) Volcano plots show differentially expressed lncRNAs in g-BMF (A) and genetically u-BMF (B) compared with controls ($|\log_2$ fold change$| > 1$, FDR < 0.05). Significantly up- and down-regulated lncRNAs are highlighted. (C) Heatmap of Hallmark gene set enrichment analysis (GSEA) based on Spearman's correlation between each of the 11 selected lncRNAs and protein-coding genes, illustrating recurrent enrichment of immune- and metabolic-related pathways. (D) RT-qPCR validation of *ATP1A1-AS1, USP3-AS1,* and *SNHG32*shows significant overexpression (ΔΔCt, normalized to GUSB) in BMF samples compared with controls. (E) lncRNA-mRNA co-expression network, showing significant correlations between three selected lncRNAs and differentially expressed mRNAs. Edges indicate significant correlations ($|r| > 0.4$, $P < 0.05$, FDR < 0.05). The top 25 most strongly correlated mRNAs are displayed for each lncRNA.

Supplementary Files

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

- ScientificReportsPedBMFSupplementaryData.docx
- ScientificReportssupplementarytables.xlsx