



OPEN

Genotype-integrated single-cell transcriptome analysis reveals the role of DDX41 p.R525H in a patient with myelodysplastic neoplasms

Hirotaka Matsui^{1,2,3}✉, Akiko Nagamachi⁴, Minori Koizumi², Rei Kudo², Masahiko Ajiro², Hironori Harada^{5,6}, Yuka Harada⁷, Shigeyuki Shichino⁸ & Akihide Yoshimi²

DEAD-box helicase 41 (*DDX41*) is implicated in germline (GL)-predisposed myeloid neoplasms, where pathogenic GL variants often lead to disease following the acquisition of a somatic variant in trans, most commonly p.R525H. However, the precise molecular mechanisms by which *DDX41* variants contribute to the pathogenesis of myeloid neoplasms remain poorly understood, partly due to challenges in establishing cellular and animal models that faithfully recapitulate the human disease phenotype. This limitation highlights the necessity of directly analyzing primary human disease cells. In this case report, conducted to pursue this objective, we implemented single-cell RNA sequencing integrated with genotyping at the p.R525 locus in a myelodysplastic neoplasm (MDS) harboring both germline and somatic *DDX41* variants, leveraging highly efficient Terminator-Assisted Solid-phase cDNA amplification and sequencing. We found that acquiring p.R525H induced G2/M cell cycle arrest selectively in colony-forming unit-erythroid cells, accompanied by R-loop accumulation, which impaired erythropoiesis through DNA damage. In hematopoietic stem and myeloid progenitor populations, gene expression profiles were largely similar between p.R525H-positive and -negative cells. However, ligand-receptor interaction and transcriptional regulation analyses suggested a non-cell-autonomous influence from p.R525H-expressing cells on GL variant-only cells. This interaction drove convergence toward a shared expression profile, highlighting an intricate interplay shaping the patient's MDS phenotype.

Keywords Cell-cycle arrest, *DDX41*, Myeloid neoplasms with germline predisposition, Single-cell transcriptome, p.R525H variant

Myeloid neoplasms associated with germline (GL) predisposition are recognized as a distinct entity in the World Health Organization classification of myeloid neoplasms and as a provisional category in the International Consensus Classification^{1,2}. Among the most frequently implicated genes is DEAD-box helicase 41 (*DDX41*)³. Individuals harboring a GL pathogenic *DDX41* variant frequently acquire a somatic alteration in trans, typically p.R525H, during the development of myeloid neoplasms. This somatic alteration is present in more than 50% of cases involving *DDX41*-associated myeloid neoplasms^{4,5}.

DDX41 plays critical roles in small nucleolar RNA processing, RNA splicing, and the suppression of R-loop accumulation^{4,6,7}. However, the precise molecular mechanisms by which *DDX41* variants contribute to the pathogenesis of myeloid neoplasms remain poorly understood, partly due to the challenges in establishing cellular and animal models that recapitulate human disease⁶⁻⁸. In this study, we investigated the impact of the *DDX41* p.R525H variant on myelodysplastic neoplasms (MDS) hematopoietic stem and progenitor cells (HSPCs)

¹Department of Laboratory Medicine, National Cancer Center Hospital, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan. ²Division of Cancer RNA Research, National Cancer Center Research Institute, Tokyo, Japan. ³Department of Medical Oncology and Translational Research, Graduate School of Medical Sciences, Kumamoto University, Kumamoto, Japan. ⁴Department of Animal Experimentation, The Foundation for Biomedical Research and Innovation at Kobe, Hyogo, Japan. ⁵Laboratory of Oncology, Tokyo University of Pharmacy & Life Sciences, Tokyo, Japan. ⁶Hematology Division, Tokyo Metropolitan Cancer and Infectious Diseases Center, Komagome Hospital, Tokyo, Japan. ⁷Clinical Research Support Center, Tokyo Metropolitan Cancer and Infectious Diseases Center, Komagome Hospital, Tokyo, Japan. ⁸Research Institute of Biomedical Sciences, Tokyo University of Science, Chiba, Japan. ✉email: hmatsui@ncc.go.jp

through highly efficient single-cell RNA sequencing (scRNA-seq) combined with single-cell genotyping^{9,10} at the p.R525 locus. Our findings reveal cell type-specific effects of the p.R525H variant, including significant disruptions in erythroid differentiation and complex interactions between p.R525H-positive cells (R525H cells) and GL variant-only cells (GL cells) in myeloid lineages, leading to convergence in their gene expression patterns. These results underscore the multifaceted role of *DDX41* p.R525H in shaping this patient's MDS phenotype.

Results

Selection of a clinical sample aligned with the study objective

This study aimed to investigate an untreated MDS patient whose hematopoietic cells comprised a mixture of those harboring only a GL *DDX41* variant and those with an additional somatic variant. Based on these criteria, we selected a male patient in his late 50s, initially diagnosed with MDS with low blasts (MDS-LB) in 2021 (Fig. 1a,b). Comprehensive gene panel testing identified two pathogenic *DDX41* variants: p.R339C^{11–14} (NC_00005.10: g.177513768G>A, NM_016222.4: c.1015C>T) and p.R525H^{4,15} (g.177512369C>T, c.1574G>A), with variant allele frequencies (VAFs) of 50.16% and 6.97%, respectively (Fig. 1c). The patient was managed with observation alone until 2024, when progressive cytopenia and an increase in bone marrow blast percentage to 10.4% indicated progression to MDS with increased blasts (MDS-IB2) (Fig. 1a,b). Concurrently, the VAF of p.R525H increased to 11.7% (Fig. 1c), and a deletion of the long arm of chromosome 13 was newly detected in 2 out of 20 metaphase cells. The patient subsequently underwent azacitidine treatment followed by allogeneic stem cell transplantation from an unrelated donor.

The observed VAF of p.R525H reflected a hematopoietic architecture consisting of two distinct HSPC populations: GL cells and R525H cells. While nonsense variants are generally expected to exhibit reduced gene expression due to mechanisms such as nonsense-mediated decay¹⁶, leading to the significant bias toward a higher proportion of p.R525H transcripts, this case involves a missense GL variant with established pathogenic significance^{11–14}.

Although clinical phenotypic differences between cases harboring nonsense GL variants and those with missense variants remain to be clearly defined, this case provides a unique opportunity to investigate the molecular mechanisms driving clonal evolution of the disease and the pathophysiological impact of p.R525H variant acquisition. To explore these aspects, we performed scRNA-seq analysis on bone marrow cells collected immediately before treatment initiation. Specifically, the sample used for this study was obtained in February 2024; no suitable bone marrow samples collected prior to this time point were available for scRNA-seq analysis.

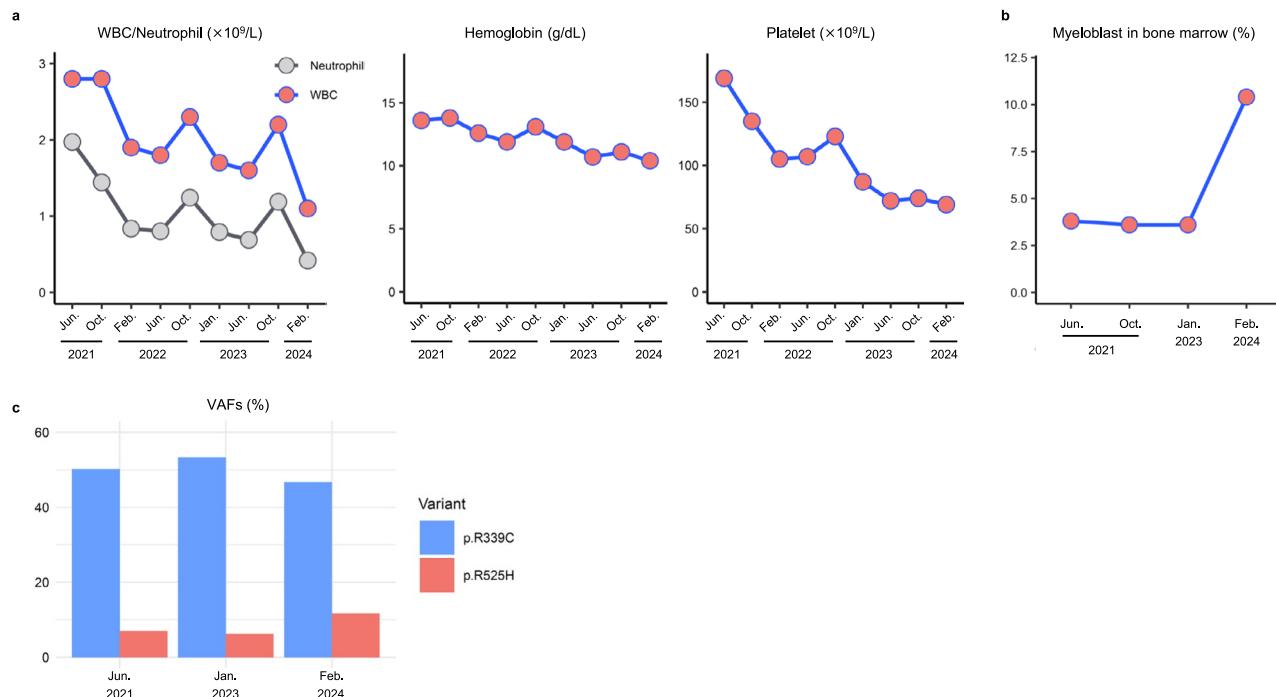


Fig. 1. Clinical course of the MDS case analyzed in this study. (a) Trends in hematological parameters over time, including white blood cell (WBC) counts, neutrophil counts, hemoglobin levels, and platelet counts. (b) Temporal changes in the percentage of myeloblasts observed in bone marrow. (c) Longitudinal analysis of VAFs for the *DDX41* variants identified in the analyzed case.

Increase of granulocyte-monocyte progenitors (GMPs) in the MDS HSPCs

The Terminator-Assisted Solid-phase cDNA amplification and Sequencing (TAS-Seq) procedure¹⁷ was applied for scRNA-seq. A total of 33,214 cells passed quality control filters, with a median of $5,657 \pm 1,627$ genes expressed per cell and a median of $34,333 \pm 19,786$ reads per cell.

Cell types were initially classified using the BoneMarrowMap¹⁸ R package applied to the Seurat object constructed from the sequencing data. Analysis of CD34-positive cells identified 41 distinct cell types (Fig. 2a). The classification was validated by the expression of characteristic marker genes for each identified cell type (Fig. 2b; Fig. S1a–c). The most abundant populations included early GMP, GMP-Neut, and GMP-Mono (Fig. 2c).

We compared the cell distribution in this case to that of CD34-positive HSPCs from three aged healthy donors (normal HSPCs) sourced from dataset GSE180298¹⁹. This comparison revealed striking differences in cell composition, with the MDS sample displaying a significantly reduced proportion of hematopoietic stem cells (HSCs) and erythroid progenitors (Fig. 2c,d).

Aberrant transcriptional regulation of MDS HSPCs leads to increased self-renewal activity in early GMPs

Gene Set Enrichment Analysis (GSEA) revealed largely consistent patterns across MDS cell types (Fig. 3a; Fig. S2a). Genes involved in ribosome biosynthesis and translational regulation were negatively enriched in MDS HSPCs compared to their normal counterparts. Although the proportion of cells classified as early GMP was considerably elevated (Fig. 2c,d), GSEA did not indicate enrichment of gene sets linked to hyperproliferative phenotypes or cell cycle arrest in this population.

RNA velocity analysis²⁰ demonstrated dynamic state transitions originating from early GMPs (Fig. 3b). The directional arrows in Fig. 3b represent inferred cell state transitions based on gene expression dynamics. In normal HSPCs, cell state transitions are observed to initiate mostly from HSCs, whereas in the analyzed MDS case, early GMPs serve as the starting point, suggesting enhanced self-renewal activity of early GMPs. SCENIC-based gene regulatory network analysis²¹ identified upregulation of transcriptional regulation driven by Kruppel-like factor 2 (KLF2) and KLF4, key transcription factors involved in hematopoiesis²² in MDS GMPs (Fig. 3c,d; Fig. S2b), accompanied by an increase in the transcriptional levels of these factors (Fig. 3e). Additionally, JUN/FOS, which plays a crucial role in normal HSC function and is involved in stress-induced hematopoiesis and aging^{23–25}, showed altered regulatory patterns in this MDS case; JUN/FOS activity extended into more differentiated progenitor cells. Trajectory analysis²⁶ further highlighted these regulatory disruptions (Fig. S2c): JUN expression was notably elevated in MDS HSCs compared to normal HSCs, with sustained high expression across differentiation stages. In contrast, FOS expression was slightly lower in MDS HSCs than in normal HSCs but remained relatively stable throughout differentiation, leading to higher levels in MDS myeloid progenitors. KLF2 and KLF4 expression, initially low at early stages, was upregulated as cells progressed along the differentiation trajectory.

Collectively, these findings suggest that multiple dysregulated transcriptional programs converge to reinforce the self-renewal capacity of early GMPs, potentially contributing to the impaired differentiation observed in this MDS case.

Transcriptional divergence and convergence between R525H and GL cells

We performed an amplicon sequencing for single-cell genotyping (Fig. S3a) and integrated the genotype information with gene expression data. Sufficient sequencing depth for genotyping was obtained, with approximately 100 to 10,000 reads per cell (Fig. S3b). Based on the read count of p.R525H versus wild-type transcripts, 6,575 cells were classified as GL cells and 10,843 cells as R525H cells. Note that this genotyping approach is designed to classify cells as either carrying wild-type alleles on both chromosomes or harboring the p.R525H variant on one allele. Since the classification relies on the absence of p.R525H to determine wild-type status, cells carrying p.R525H are inherently more likely to be identified, leading to a possible bias in genotyping. As a result, the observed ratio of GL to R525H cells based on the genotyping sequence may not accurately reflect their actual proportions. Given these limitations, we did not use scRNA-seq-based genotyping information for VAF estimation. Instead, we adopted the VAF values obtained from gene panel testing as the definitive measure (Fig. 1c). The genotyping was further validated using the more rigorous method described by Cortés-López et al.¹⁰, and demonstrated a remarkably high agreement rate of 96.5%, underscoring the robustness of our genotyping approach.

Compared to GL cells, R525H cells exhibited a further increase in the proportion of myeloid progenitors and a decrease in the proportions of HSC and lympho-myeloid primed progenitor cells (LMPPs) (Fig. 4a,b). GL cells displayed a GMP-centric distribution pattern distinct from normal cellular profiles (Fig. S3c), implying that GL cells might have served as the origin of MDS in this case. However, as discussed later, an alternative perspective on this point is also possible. The presence of HSC among R525H cells indicates that the p.R525H variant likely arose in an HSC. Intriguingly, Fig. S3b suggests variability in the relative expression of p.R525H transcripts among R525H cells, with some cells predominantly expressing p.R525H transcripts while others express both wild-type and p.R525H transcripts at comparable levels. While the influence of technical factors such as allele dropout cannot be entirely excluded, several lines of evidence collectively support the robustness of this observation: (1) the overall comparable *DDX41* expression levels between cells predominantly expressing p.R525H transcripts and those genotyped as wild-type ($p = 0.33$ by Wilcoxon rank sum test); (2) the substantial number of clearly heterozygous cells; and (3) the distinct cell type distributions observed among cells with predominant p.R525H transcript expression compared to heterozygous cells (Fig. S3d). However, given the primary objective of this study, we analyzed R525H cells as a single group to capture the overall impact of the p.R525H variant.

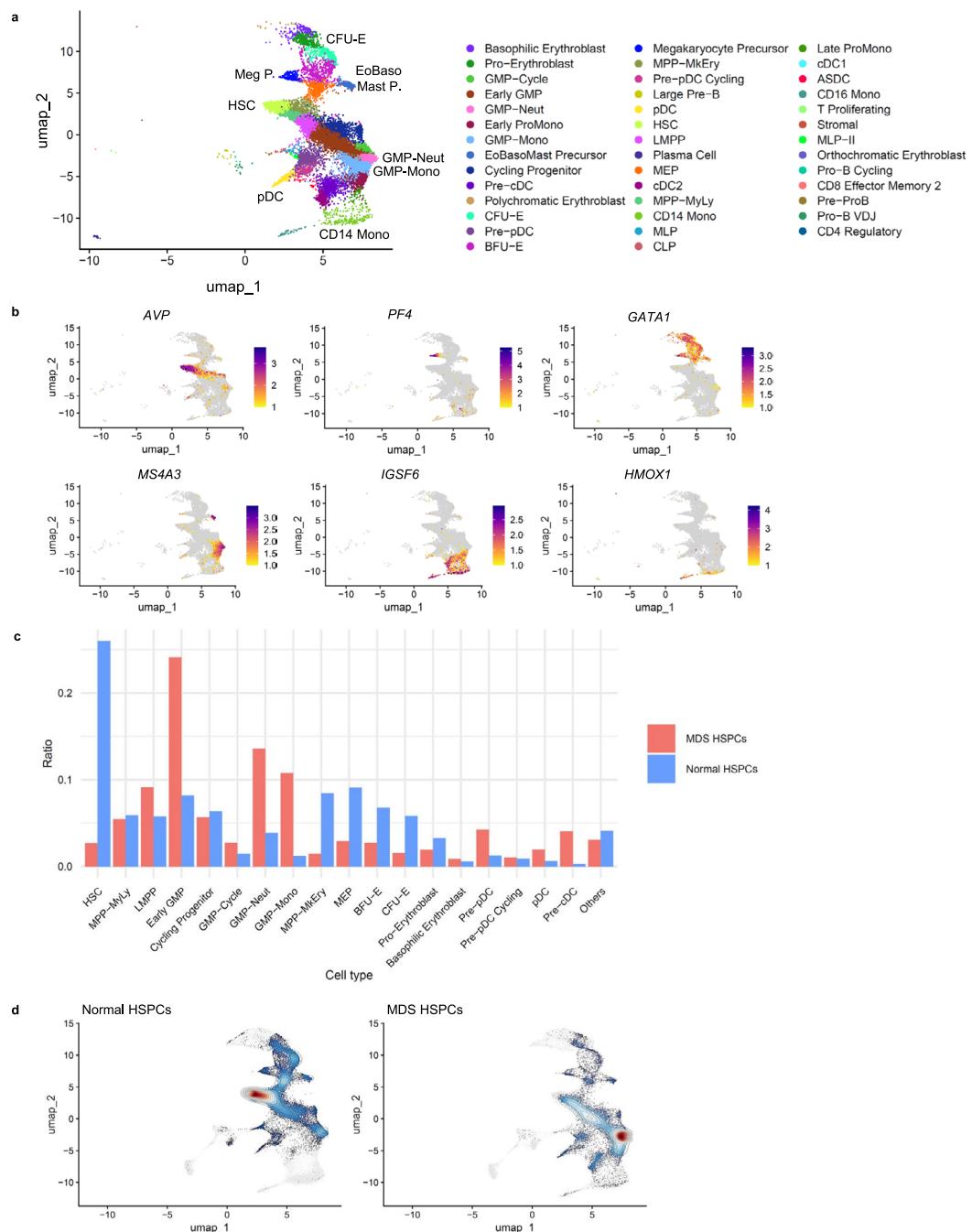


Fig. 2. Characterization of CD34-positive bone marrow cells and comparison with healthy donor samples. (a) UMAP plot of CD34-positive bone marrow cells, displaying 41 distinct cell types classified using the BoneMarrowMap R package. (b) UMAP distribution of cell type-specific gene expression. Representative expression patterns are shown for HSCs (AVP), megakaryocytic progenitors (PF4), erythroid progenitors (GATA1), monocytic progenitors and EoBosMast precursors (MS4A3), dendritic cells and monocytes (IGSF6), and monocytes (HMOX1). (c) Proportion of classified cell types within CD34-positive HSPCs. The plot displays the relative abundance of each cell type in the MDS case and three healthy donor samples. (d) Density distribution of cells visualized on UMAP. The MDS sample is contrasted with a combined dataset of three healthy donors, using the 'plot_Projection_byDonor' function from the BoneMarrowMap R package. Cell name abbreviations: ASDC, AXL + Siglec-6 + dendritic cells; BFU-E, burst forming unit-erythroid; cDC, conventional dendritic cell; CLP, common lymphoid progenitor; EoBosMast Precursor, eosinophil, basophil, and mast cell precursor; GMP-Mono, GMP-monocyte; GMP-Neut, GMP-neutrophil; MEP, megakaryocyte-erythroid progenitor; MLP, multi-lymphoid progenitor; Mono, monocyte; MPP-MkEry, multipotent progenitors with megakaryocyte-erythroid priming; MPP-MyLy, multipotent progenitors with myeloid-lymphoid priming; pDC, plasmacytoid dendritic cell; Pre-B, pre-B lymphocyte; Pro-B, pro-B lymphocyte; ProMono, pro-monocyte. Gene name abbreviations: AVP, arginine vasopressin; GATA1, GATA binding protein 1; HMOX1, heme oxygenase 1; IGSF6, immunoglobulin superfamily member 6; MS4A3, membrane spanning 4-domains A3; PF4, platelet factor 4.

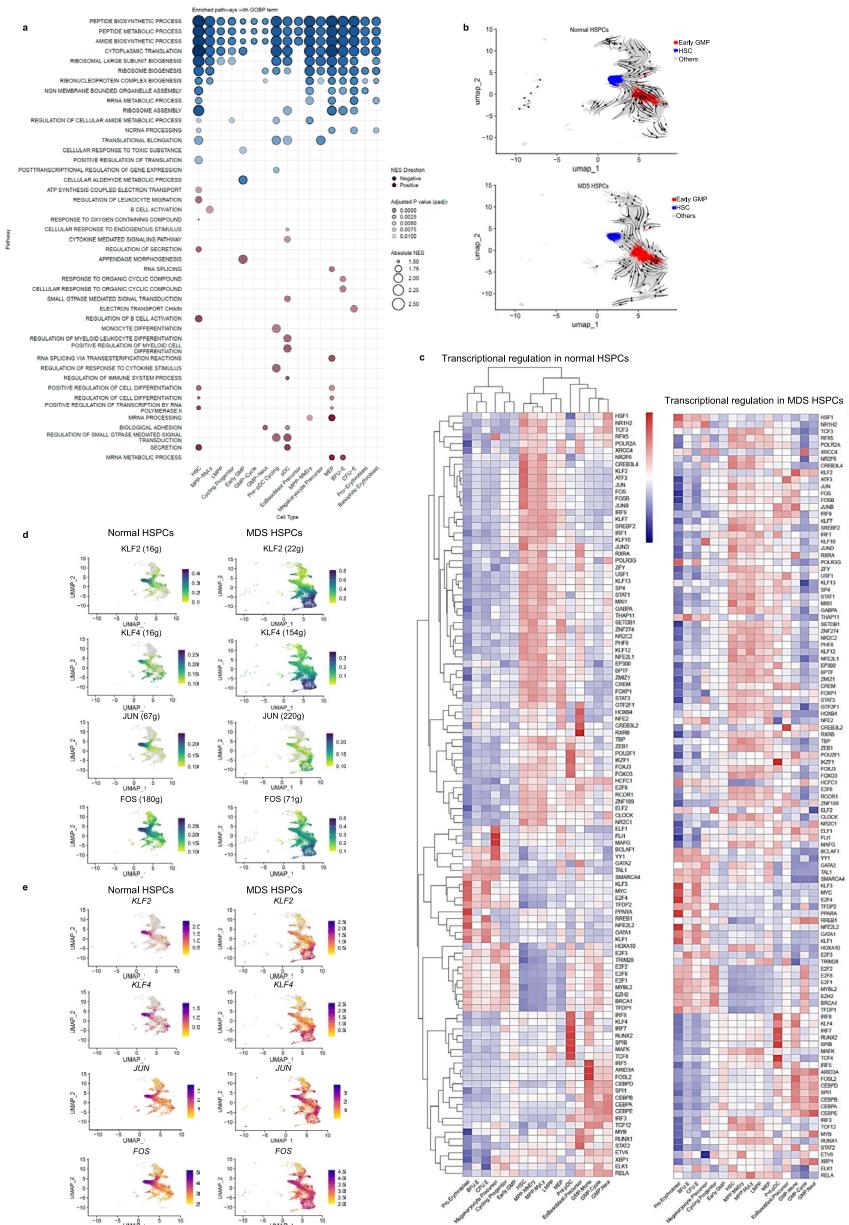


Fig. 3. Distinct transcriptional regulations of CD34-positive HSPCs between MDS case and healthy donors. (a) Bubble plot of GSEA results for each cell type with ≥ 100 cells in both normal and MDS HSPCs. Pathways classified under Gene Ontology Biological Processes (GOBP) categories with adjusted p -values ($padj$) < 0.01 are displayed. Positively enriched pathways in MDS HSPCs are shown in red, while negatively enriched pathways are in blue. Circle size corresponds to the Nominal Enrichment Score (NES), and color intensity reflects statistical significance ($padj$). (b) Cell state transitions inferred by scVelo. Differentiation trajectories based on RNA velocity (unspliced/spliced transcripts) are visualized as arrows overlaid on the UMAP, highlighting HSC (blue) and early GMP (red). (c) Heatmap of transcriptional regulation by cell type, depicting average area under the curve (AUC) scores for transcription factors calculated per cell. Only cell types with > 100 cells in both normal and MDS HSPCs are included. The MDS heatmap is aligned with the healthy donor heatmap. (d) UMAP plot visualizing transcriptional regulation at the single-cell level through AUC scores. (e) UMAP plot illustrating transcription factor expression at the single-cell level. Numbers beside gene names indicate the count of target genes regulated by each transcription factor.

Next, we investigated differential gene expression between the two genotypes across 19 distinct cell types, each containing ≥ 200 cells. In HSCs and early myeloid progenitors, no genes met the defined threshold for differential expression ($p < 0.05$), indicating that the acquisition of the p.R52H variant did not cause significant gene expression changes in these populations (Fig. 4c). However, remarkable changes were observed in erythroid progenitors, particularly beyond the colony-forming unit-erythroid (CFU-E) stage.

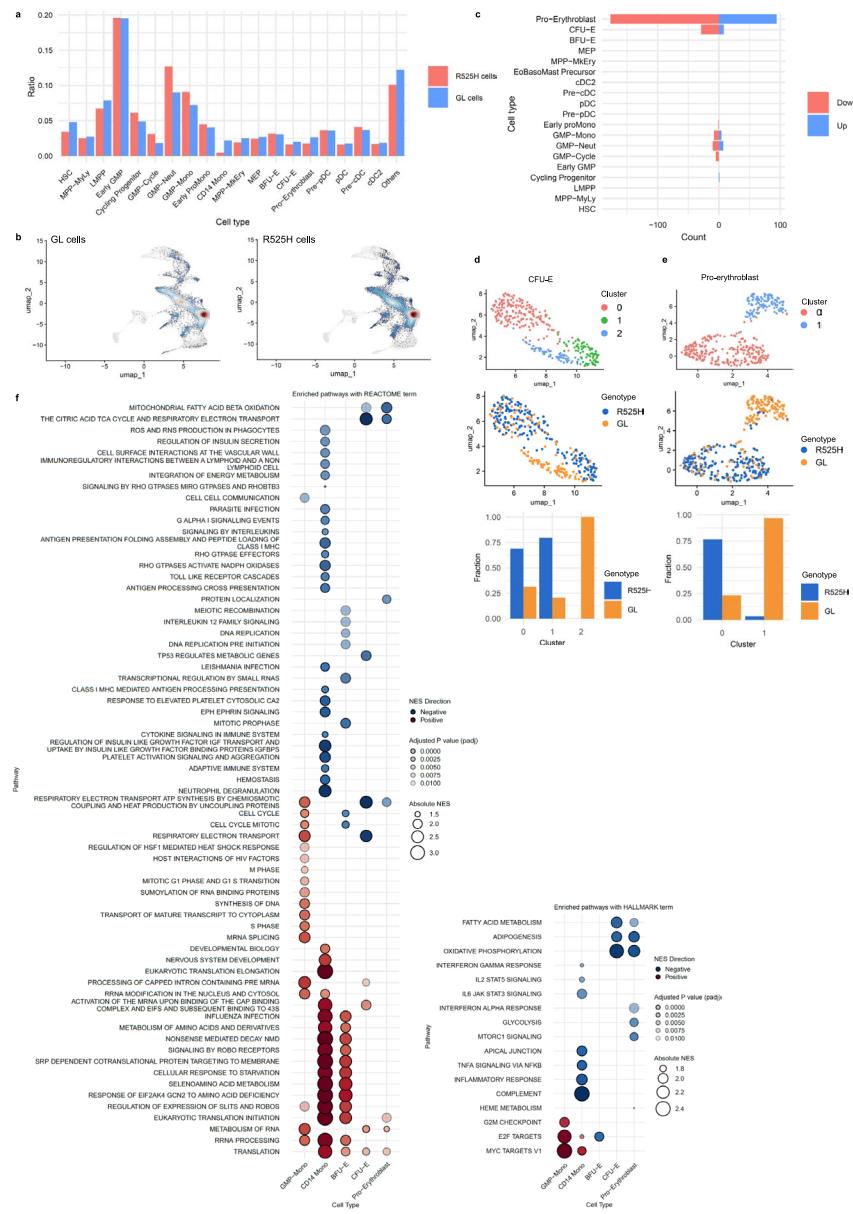


Fig. 4. Comparison of R525H and GL cells by scRNA-seq analysis combined with single-cell genotyping. **(a)** Proportion of classified cell types, illustrating the relative abundance of each cell type in R525H and GL populations. **(b)** Density distribution of R525H and GL cells, highlighting population differences across the UMAP projection. **(c)** Differential gene expression between R525H and GL cells across cell types. Bars indicate the number of differentially expressed genes ($p < 0.05$), with upregulated genes in R525H cells shown in blue and downregulated genes in red. **(d,e)** UMAP plots showing re-clustering results for CFU-E (d) and pro-erythroblast (e) populations. Upper panels: UMAP projections with clusters. Middle panels: Cells distinguished by genotype (R525H or GL), with color-coded labels. Lower panels: Bar graphs illustrating relative fractions of R525H and GL cells within each cluster. **(f)** GSEA comparison of R525H and GL cells, highlighting enriched REACTOME and HALLMARK pathways ($padj < 0.01$). Only cell types containing at least one pathway with $padj < 0.01$ are displayed. Pathways positively enriched in R525H cells are shown in red, and negatively enriched pathways in blue. Circle size represents the NES, while color intensity indicates statistical significance ($padj$).

Clustering by the self-assembling manifolds (SAM) algorithm²⁷ corroborated this observation. In the HSC and early myeloid progenitor populations, R525H and GL cells did not form distinct clusters in the uniform manifold approximation and projection (UMAP) visualizations generated by the SAM algorithm (Fig. S4a). However, in the CFU-E population, R525H and GL cells formed three distinct clusters: one comprising both genotypes, with a composition of 69% R525H and 31% GL cells, one predominantly composed of R525H cells (80% R525H and 20% GL), and another containing only GL cells (Fig. 4d). In the pro-erythroblast population,

cells were divided into two clusters: one primarily consisting of GL cells and another predominantly containing R525H cells (Fig. 4e).

GSEA comparing R525H and GL cells across cell types revealed genotype-dependent behavioral differences within erythroid and monocyte progenitors (Fig. 4f). Although the overall differences in transcriptional regulation between R525H and GL cells were more subtle than those observed between normal and MDS cells (Fig. S4b,c), notable variations emerged. For instance, CCCTC-binding factor (CTCF), a regulator essential for hematopoietic differentiation²⁸, exhibited increased relative activity in erythroid progenitor cells. Additionally, transcriptional regulators such as KLF16 and TEA domain transcription factor 4 (TEAD4), whose roles in hematopoiesis remain less well-defined, displayed distinct regulatory patterns between R525H and GL cells, suggesting potential genotype-specific transcriptional modulation.

Dysregulated cell cycle dynamics and mitochondrial dysfunction in R525H CFU-E impair erythropoiesis

In the CFU-E population, cells in the G2/M phase were specifically enriched within the cluster predominantly comprising R525H cells (“cluster 1” in Fig. 4d; Fig. S4a). Comparative pathway enrichment analysis revealed that cluster 1 exhibited higher enrichment of pathways related to cell cycle regulation and RNA splicing compared to cluster 2, which primarily contained GL cells (Fig. 5a). The distinct clustering within pro-erythroblasts (Fig. 4e) was attributed to enhanced DNA damage response pathways in R525H cells, in addition to altered cell

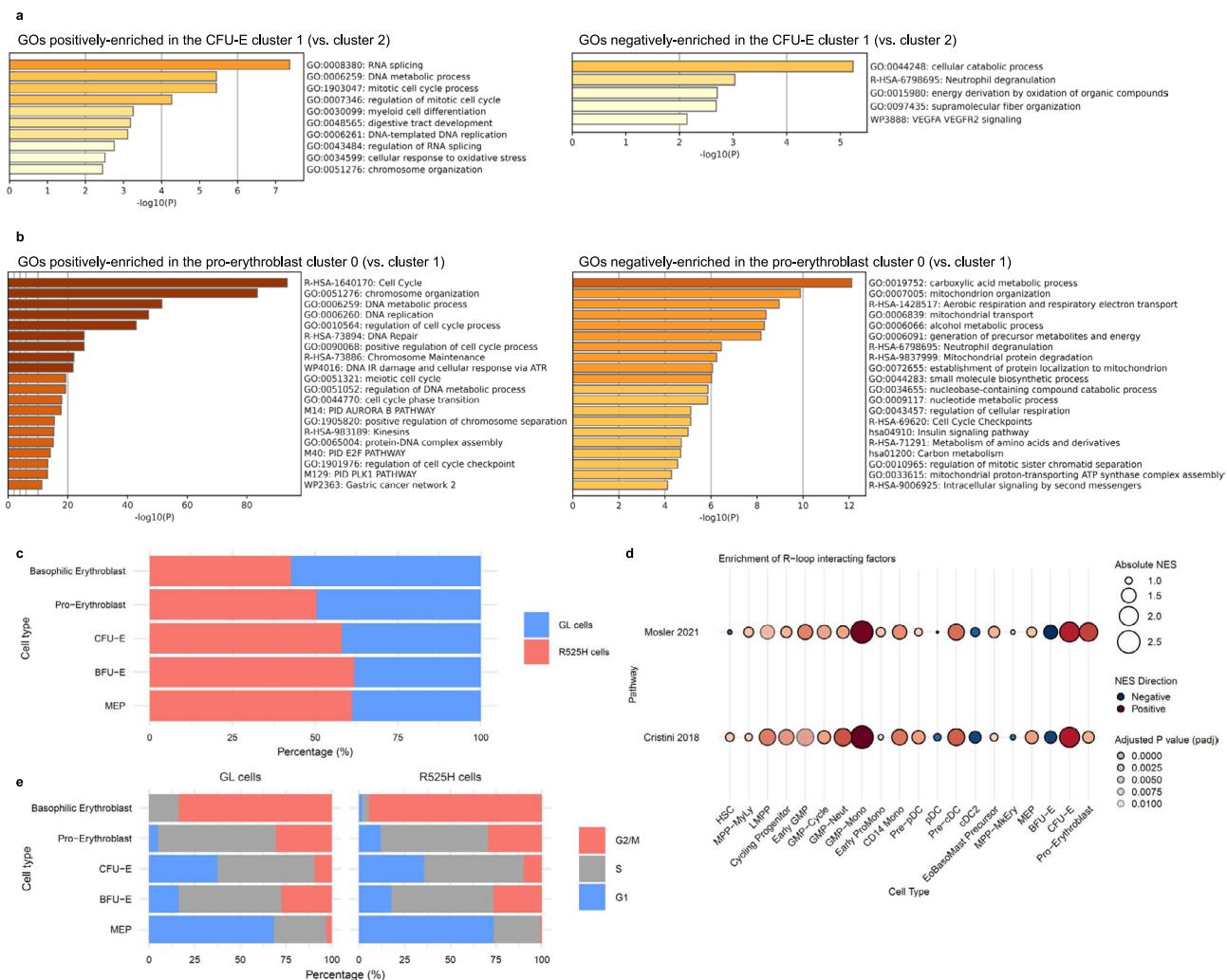


Fig. 5. Impaired erythroid differentiation by the acquisition of p.R525H. **(a,b)** Pathway enrichment analysis. **(a)** CFU-E cluster 1 (vs. cluster 2) pathway. **(b)** Pro-erythroblast cluster 0 (vs. cluster 1) pathway. Differentially expressed genes ($padj < 0.05$) were identified via Seurat's 'FindMarkers' function and analyzed in Metascape (<https://metascape.org>). **(c)** Proportion of R525H cells relative to GL cells in erythroid progenitors, visualized as a 100% stacked bar chart for each cell type. **(d)** Enrichment of R-loop-interacting factors^{29,30} in R525H and GL cells. GSEA used curated factors in GMT format, with positive and negative enrichments in R525H cells shown in red and blue, respectively. Circle size represents NES, and color intensity indicates $padj$. **(e)** Cell cycle status estimates in erythroid progenitors. Seurat's 'CellCycleScoring' function assessed phase distribution (G1, S, G2/M), presented as a 100% per-phase bar.

cycle dynamics (Fig. 5b). The proportion of R525H cells among erythroid progenitors declined progressively as erythroid maturation advanced (Fig. 5c), suggesting that R525H CFU-E undergo G2/M arrest, impairing proliferation and differentiation during erythropoiesis.

Reduced expression and compromised function of DDX41 disrupt the coordination of RNA splicing and transcriptional elongation, resulting in R-loop accumulation, increased replication stress, and subsequent G2 phase arrest⁷. Consistent with this, GSEA using a predefined list of R-loop binding factors^{29,30} showed significant enrichment in R525H cells, particularly within CFU-E and GMP-Mono (Fig. 5d). The pronounced mitotic defect in CFU-E compared to more cycling erythroid stages (Fig. 5e) remains unexplained. However, as discussed later, reduced mitochondrial function in R525H CFU-E cells (Fig. 4f) may contribute to the accumulation of excessive R-loops.

A recent study identified two distinct erythroid differentiation trajectories, one characterized by elevated heme synthesis, ultimately leading to cell death³¹. Diffusion map analysis of erythroid populations in this MDS case identified two trajectories: trajectory A, supporting normal erythropoiesis, and trajectory B, associated with cell death (Fig. S5a). Although a higher proportion of cells aligned with trajectory B, the distribution of R525H and GL cells was comparable across both trajectories (Fig. S5b). This suggests that the p.R525H variant promotes apoptosis through a mechanism independent of heme synthesis-associated cell death.

Influence of p.R525H on myeloid lineage phenotypes through non-cell-autonomous mechanisms

The VAF of p.R525H remained around 12% even during the advanced stages of the disease (Fig. 1c), consistent with previous reports indicating that p.R525H rate at diagnosis typically remains below 20%^{6,32}. Studies in mouse models have suggested that R525H cells act as a minor clone, exerting a phenotypic influence on GL cells⁶.

Despite the absence of notable differences in gene expression or transcriptional regulation between R525H and GL cells in myeloid lineages (Fig. 4c; Fig. S4b), GL cells already exhibited MDS-related altered cell population distributions (Fig. 4a,b). While this could imply that MDS originates primarily from the acquisition of a GL *DDX41* variant in this case, with p.R525H playing a relatively minor role in phenotypic changes within the myeloid lineage, we consider it more likely that interactions between R525H and GL cells contribute to disease development. Given that p.R525H is frequently associated with myeloid neoplasms^{4,5}, it is plausible that R525H cells influence GL cells, leading to a convergent gene expression profile that collectively shapes the disease phenotype. To explore this hypothesis, we conducted a detailed analysis of ligand-receptor interactions between GL and R525H cells. This approach was motivated by previous studies showing that cells harboring clonal hematopoiesis-related variants, as well as del(5q) MDS cells, can induce non-cell-autonomous gene expression changes in neighboring unaffected cells^{33,34}.

We analyzed scRNA-seq data using LIANA³⁵ to investigate ligand-receptor interactions within normal HSPCs, within GL cells, and between R525H and GL cells. This approach quantified interactions involving detected ligands and receptor combinations (Fig. S6a,b). We focused on HSCs, LMPPs, and early GMPs as target cells to examine lineage-specific interaction changes during early myeloid hematopoiesis (Fig. 6a; Fig. S6c,d). Our results revealed substantial alterations in interaction patterns between normal and MDS HSPCs: many interactions observed in normal cells were diminished in MDS cells, while novel interactions emerged. Notably, interactions mediated by midkine (MDK) and calreticulin (CALR), absent in normal HSPCs, became prominent in MDS cells (Fig. 6a; Fig. S6c). MDK is implicated in the expansion of HSC and multipotent progenitors³⁶, a role confirmed by a scRNA-seq study³⁷. CALR is associated with endoplasmic reticulum stress and recognized by natural killer cells and macrophages^{38,39}, playing a role in hematopoiesis.

Comparisons between R525H and GL cells demonstrated largely similar interaction patterns involving common cell types acting on HSCs, LMPPs, and early GMPs. However, some notable differences emerged: transforming growth factor beta 1 (TGFB1)-mediated interactions, observed in HSCs and LMPPs, were more pronounced in R525H cells, while intercellular adhesion molecule 4 (ICAM4)-mediated interactions were absent in R525H cells (Fig. 6a; Fig. S6a,c). *TGFB1* expression was widespread across a broader range of cell types, while *ICAM4*, primarily expressed in erythroid progenitors⁴⁰, was downregulated in MDS cells (Fig. 6b). These findings suggest that changes in *TGFB1* and *ICAM4* expression levels and the resulting alterations in cell-to-cell interactions in MDS contribute to the increased complexity of cellular communication, reinforcing the hypothesis that non-cell-autonomous effects of R525H cells on GL cells play a critical role in disease progression. Statistical analysis highlighted pathways that were differently regulated between R525H and GL cells, as well as between GL and normal cells, illuminating the potential impact of *DDX41* variants on the disease phenotype (Fig. 6c). Comparisons of interactions within GL cells with those in normal cells revealed shifts in adaptive immune regulation and cell adhesion, patterns echoed in R525H versus GL cell comparisons. These interactional changes likely drive the convergence of gene expression profiles and phenotypes between R525H and GL cells.

Discussion

This study presents a novel genotype-integrated scRNA-seq analysis of HSPCs from an MDS case with a GL pathogenic *DDX41* variant in conjunction with a somatic hotspot variant. Modeling human leukemogenesis with *DDX41* variants poses significant challenges, as its introduction—especially in the context of *DDX41* haploinsufficiency—often results in rapid growth arrest and cell death^{6,7}. This complicates establishing long-term experimental models, underscoring the critical need to utilize clinical specimens, despite the inherently limited availability of hematopoietic cells carrying *DDX41* variants. To address this challenge, we generated a high-quality scRNA-seq library from freshly isolated CD34-positive cells, collected and processed on the same day, using the TAS-Seq protocol¹⁷. This approach facilitated robust transcriptomic analysis from limited clinical material.

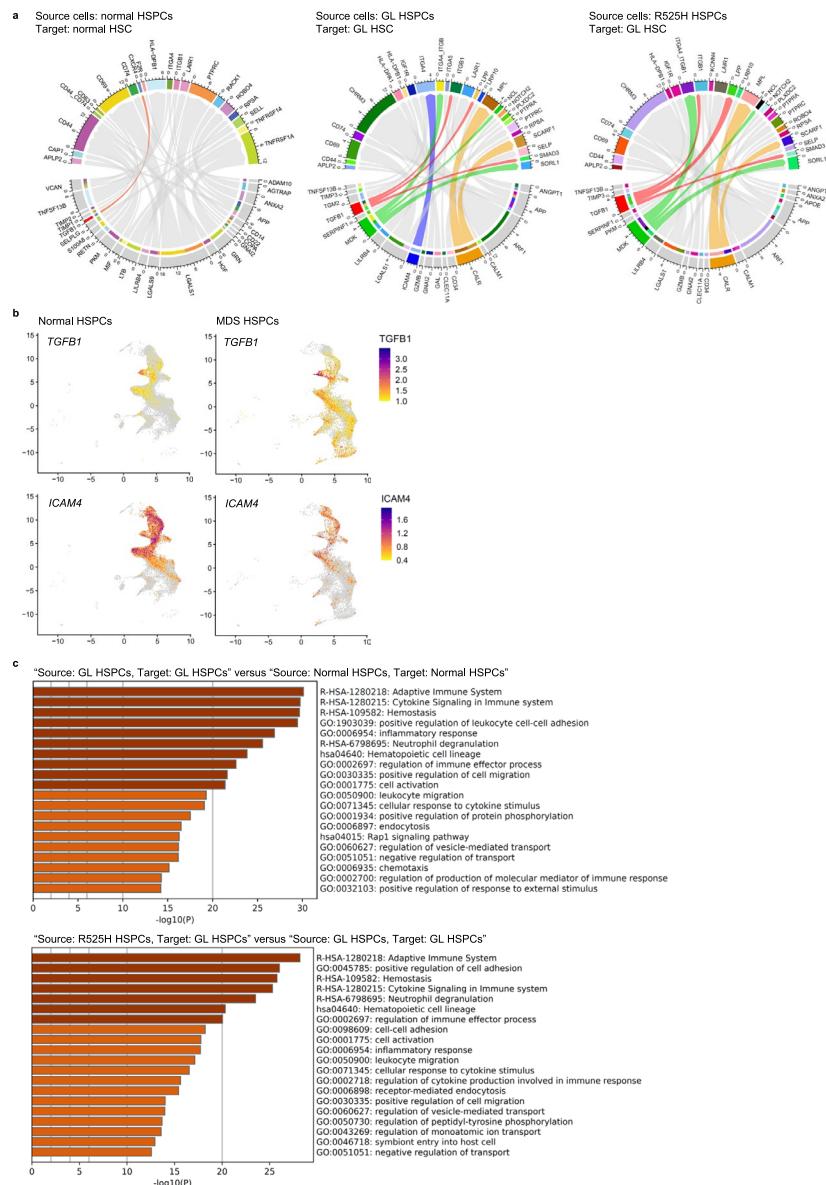


Fig. 6. Different ligand-receptor interactions between normal and MDS HSPCs and between R525H and GL cells. **(a)** Chord diagrams illustrating cell-to-cell interactions within healthy HSPCs, within GL cells, and from R525H cells to GL cells. Key ligand-receptor interactions are highlighted with color-coded connections referenced in the text. **(b)** Expression patterns of *TGFBI* and *ICAM4*, in normal and MDS HSPCs, visualized on UMAP plots. **(c)** Pathway enrichment analysis. Significant interaction differences ($p < 0.01$) between normal and GL cells (upper), and between R525H and GL cells (lower), analyzed using associated ligand and receptor names.

While our findings are based on a single case, they align with previous experimental studies, reinforcing the biological significance of DDX41 dysfunction in hematopoietic malignancies. Notably, our results demonstrate that G2/M cell cycle arrest, linked to increased DNA damage in CFU-E cells, drives defective erythropoiesis in the presence of the p.R525H variant, highlighting an intrinsic defect specific to these cells. This finding aligns with previous zebrafish studies showing that *ddx41* loss caused defective RNA splicing, G2/M phase accumulation, and DNA damage, ultimately leading to cell death⁴¹. In prior work, we have shown that DDX41 disruption results in R-loop accumulation, triggering severe DNA damage after mitosis⁷. In line with this, R-loop regulatory factors were elevated in R525H CFU-E, and DNA damage response genes were highly enriched in pro-erythroblasts, the immediate progeny of CFU-E. A recent study has also linked G-quadruplex accumulation, which stabilizes R-loops⁴², to impaired erythropoiesis in a *DDX41*-deficient context⁴³. The reason for the pronounced defect in CFU-E compared to more proliferative stages remains unclear. However, our data suggest that reduced mitochondrial oxidative phosphorylation in CFU-E cells harboring p.R525H may impair energy metabolism, limiting the resolution of excess R-loops. This metabolic vulnerability may exacerbate G2/M arrest, further compromising erythropoiesis in these progenitors.

Identifying the specific effects of p.R525H acquisition in the myeloid lineage proved challenging due to the overall similarity in gene expression profiles between R525H and GL cells. However, evidence suggests that R525H cells influence the behavior of co-existing GL cells⁶. Our study proposes that the p.R525H variant likely confers a selective advantage by enhancing self-renewal in early GMPs, while fostering convergence in gene expression profiles between R525H and GL cells. This interplay underscores the complex dynamics shaping the MDS phenotype. In MDS, altered innate immune responses to inflammatory states are considered to contribute to cellular fitness advantages⁴⁴. However, in the context of *DDX41* variants, the role of innate immune signaling appears limited, highlighting the distinct pathobiological features of this malignancy.

In R525H cells, *DDX41* dysfunction may cause RNA splicing alterations in critical genes without remarkable changes in overall gene expression. However, unlike canonical splicing-related factors such as *SF3B1*, which is frequently mutated in MDS, the role of *DDX41* in RNA splicing remains incompletely understood. While our previous studies have demonstrated that *DDX41* is essential for efficient splicing progression rather than splice site selection⁷, some reports suggest that *DDX41* homologs may be involved in 3' splice site recognition^{45,46}. The scRNA-seq method employed in this study was not suitable for detailed splicing analysis due to the inherent limitations of the data. Therefore, to further elucidate the role of *DDX41* in RNA splicing, future studies should incorporate long-read scRNA-seq approaches in conjunction with conventional scRNA-seq.

While the current study provides valuable insights into the transcriptional alterations associated with the *DDX41* p.R525H variant, several limitations warrant discussion. First, a time-series analysis and investigation of additional cases would offer deeper insights into the gene expression state of GL cells prior to the acquisition of the p.R525H variant and the mechanism by which this acquisition influences their transcriptional profile. Nonetheless, such analyses were not feasible due to the limited clinical material. Second, a chromosomal alteration, specifically del(13q), was detected in the sample used for scRNA-seq. Although the frequency of del(13q) was low (two out of 20 metaphase cells), it may have influenced the gene expression profiles in the MDS cells. Several algorithms were tested to infer copy number substructure from scRNA-seq data; however, the outputs were inconsistent, which hindered the reliable evaluation of chromosomal changes at single-cell resolution. Finally, this study primarily focused on transcriptional changes in CD34-positive HSPCs associated with *DDX41* variants and did not extensively examine ligand-receptor interactions involving lymphocytes or stromal cells. Future studies incorporating these cell types could offer a more comprehensive understanding of the cellular dynamics within the bone marrow microenvironment.

Materials and methods

Ethics declarations

This study was approved by the Ethical Review Committees of Tokyo Metropolitan Cancer and Infectious Diseases Center Komagome Hospital and National Cancer Center (approval numbers: Komagome Hospital, 2203; National Cancer Center, 2022-355) and was conducted in accordance with the Japanese Ethical Guidelines for Medical and Biological Research Involving Human Subjects, as well as the Declaration of Helsinki. The attending physician provided written and verbal explanations of the study to the patient, who provided written informed consent to publish the study findings in the manuscript.

Cell preparation

A minimal volume of bone marrow cells required for this study was collected during a routine clinical bone marrow examination. CD34-positive HSPCs were isolated using the EasySep™ Human CD34 Positive Selection Kit II (STEMCELL Technologies, Vancouver, Canada). Subsequently, approximately 40,000 cells (575 µL) were immediately loaded into the BD Rhapsody™ Express Single-Cell Analysis System (BD Biosciences, Franklin Lakes, NJ, USA).

Library preparation for TAS-Seq assay

scRNA-seq library was prepared using the TAS-Seq method¹⁷. Briefly, first-strand cDNA synthesis was performed on the BD Rhapsody™ Enhanced magnetic beads (BD Biosciences). Following exonuclease I treatment, a dC-tailing reaction was carried out at the 3' end of the cDNA using terminal deoxynucleotidyl transferase/RNase H enzymes. A mixture of dideoxycytidine 5'-triphosphate (ddCTP) and deoxycytidine triphosphate (dCTP) at a specific ratio was used to terminate the elongation reaction once ddCTP was incorporated. This approach effectively reduced the synthesis of by-products caused by excessive primers and promoted efficient cDNA synthesis. For second-strand synthesis, a 5' universal-dG12 primer was employed instead of the 5' universal-dG9 primer described in the original protocol, enhancing gene-detection sensitivity. After performing the first PCR with universal primers, size selection, and amplification via a second PCR, the library was ligated with TruSeq i7 adaptors (Illumina, San Diego, CA, USA) and further amplified through PCR to add P5/P7 sequencing adapters. Sequencing was performed using a NovaSeq 6000 sequencer and NovaSeq S4 v1.5 reagent kit (200 cycles) (Illumina).

DDX41 p.R525 genotyping

Amplicons for genotyping at the *DDX41* p.R525 site were generated using the whole transcript as a template, with primers listed in Table S1. Library preparation details are presented in Fig. S3a. Each PCR was performed with KAPA HiFi HotStart ReadyMix (NIPPON Genetics, Tokyo, Japan) under the following conditions: 95 °C for 3 min, followed by 10 cycles at 95 °C for 20 s, 65 °C for 20 s, and 72 °C for 1 min, with a final extension at 72 °C for 5 min. Amplified products were purified using AMPure XP (Beckman Coulter, Brea, CA, USA) and eluted in 22 µL of 10 mM Tris-HCl (pH 8.0). Sequencing was conducted on a NovaSeq 6000 with an S4 v1.5 reagent kit (200 cycles). The resulting data were mapped using a custom STAR index containing only *DDX41*

p.R525H and wild-type sequences. *DDX41* transcripts per cell were generated, and cell-specific genotyping was performed using FlowJo (BD Biosciences) (Fig. S3b), with results integrated into the Seurat object.

Data analysis

Detailed information, including data analysis procedures can be found in Doc. S1. The Supplementary Methods section in Doc. S1 describes the following processes: Preprocessing of scRNA-seq data, GSEA, RNA velocity analysis, Transcriptional regulatory network analysis, Cell trajectory analysis, Clustering of R525H and GL cells using the SAM algorithm, Ligand-receptor inference from cell-to-cell communication analysis, and Diffusion map analysis of erythroid cell differentiation trajectories.

Data availability

scRNA-seq data are available at Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE284647.

Received: 9 April 2025; Accepted: 9 June 2025

Published online: 02 July 2025

References

1. Khouri, J. D. et al. The 5th edition of the World Health Organization Classification of haematolymphoid tumours: Myeloid and histiocytic/dendritic neoplasms. *Leukemia* **36**, 1703–1719. <https://doi.org/10.1038/s41375-022-01613-1> (2022).
2. Arber, D. A. et al. International Consensus Classification of Myeloid Neoplasms and Acute Leukemias: integrating morphologic, clinical, and genomic data. *Blood* **140**, 1200–1228. <https://doi.org/10.1182/blood.2022015850> (2022).
3. Klco, J. M. & Mullighan, C. G. Advances in germline predisposition to acute leukaemias and myeloid neoplasms. *Nat. Rev. Cancer.* **21**, 122–137. <https://doi.org/10.1038/s41568-020-00315-z> (2021).
4. Polprasert, C. et al. Inherited and somatic defects in DDX41 in myeloid neoplasms. *Cancer Cell* **27**, 658–670. <https://doi.org/10.1016/j.ccr.2015.03.017> (2015).
5. Shinriki, S. & Matsui, H. Unique role of DDX41, a DEAD-box type RNA helicase, in hematopoiesis and leukemogenesis. *Front. Oncol.* **12**, 992340. <https://doi.org/10.3389/fonc.2022.992340> (2022).
6. Chlon, T. M. et al. Germline DDX41 mutations cause ineffective hematopoiesis and myelodysplasia. *Cell Stem Cell* **28**, 1966–1981. e1966. <https://doi.org/10.1016/j.stem.2021.08.004> (2021).
7. Shinriki, S. et al. DDX41 coordinates RNA splicing and transcriptional elongation to prevent DNA replication stress in hematopoietic cells. *Leukemia* **36**, 2605–2620. <https://doi.org/10.1038/s41375-022-01708-9> (2022).
8. Hwang, W. C. et al. Impaired binding affinity of YTHDC1 with METTL3/METTL14 results in R-loop accumulation in myelodysplastic neoplasms with DDX41 mutation. *Leukemia* **38**, 1353–1364. <https://doi.org/10.1038/s41375-024-02228-4> (2024).
9. Nam, A. S. et al. Somatic mutations and cell identity linked by genotyping of transcriptomes. *Nature* **571**, 355–360. <https://doi.org/10.1038/s41586-019-1367-0> (2019).
10. Cortés-López, M. et al. Single-cell multi-omics defines the cell-type-specific impact of splicing aberrations in human hematopoietic clonal outgrowths. *Cell Stem Cell* **30**, 1262–1281.e1268. <https://doi.org/10.1016/j.stem.2023.07.012> (2023).
11. Sébert, M. et al. Germline DDX41 mutations define a significant entity within adult MDS/AML patients. *Blood* **134**, 1441–1444. <https://doi.org/10.1182/blood.2019000909> (2019).
12. Li, P. et al. The genetic landscape of germline DDX41 variants predisposing to myeloid neoplasms. *Blood* **140**, 716–755. <https://doi.org/10.1182/blood.2021015135> (2022).
13. Cheloor Kovilakam, S. et al. Prevalence and significance of DDX41 gene variants in the general population. *Blood* **142**, 1185–1192. <https://doi.org/10.1182/blood.2023020209> (2023).
14. Maierhofer, A. et al. The clinical and genomic landscape of patients with DDX41 variants identified during diagnostic sequencing. *Blood Adv.* **7**, 7346–7357. <https://doi.org/10.1182/bloodadvances.2023011389> (2023).
15. Makishima, H. et al. Germ line DDX41 mutations define a unique subtype of myeloid neoplasms. *Blood* **141**, 534–549. <https://doi.org/10.1182/blood.2022018221> (2023).
16. Cheah, J. J. C., Hahn, C. N., Hiwase, D. K., Scott, H. S. & Brown, A. L. Myeloid neoplasms with germline DDX41 mutation. *Int. J. Hematol.* **106**, 163–174. <https://doi.org/10.1007/s12185-017-2260-y> (2017).
17. Shichino, S. et al. TAS-Seq is a robust and sensitive amplification method for bead-based scRNA-seq. *Commun. Biol.* **5**, 602. <https://doi.org/10.1038/s42003-022-03536-0> (2022).
18. Zeng, A. G. X. et al. Single-cell transcriptional atlas of human hematopoiesis reveals genetic and hierarchy-based determinants of aberrant AML differentiation. *Blood Cancer Discov.* <https://doi.org/10.1158/2643-3230.Bcd-24-0342> (2025).
19. Ainciburu, M. et al. Uncovering perturbations in human hematopoiesis associated with healthy aging and myeloid malignancies at single-cell resolution. *eLife* **12**, e79363. <https://doi.org/10.7554/eLife.79363> (2023).
20. Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, 1408–1414. <https://doi.org/10.1038/s41587-020-0591-3> (2020).
21. Aibar, S. et al. SCENIC: Single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086. <https://doi.org/10.1038/nmeth.4463> (2017).
22. Yang, Y. et al. Endothelial MEKK3-KLF2/4 signaling integrates inflammatory and hemodynamic signals during definitive hematopoiesis. *Blood* **139**, 2942–2957. <https://doi.org/10.1182/blood.2021013934> (2022).
23. Zhang, Y. et al. Temporal molecular program of human hematopoietic stem and progenitor cells after birth. *Dev. Cell.* **57**, 2745–2760.e2746. <https://doi.org/10.1016/j.devcel.2022.11.013> (2022).
24. Liebermann, D. A., Gregory, B. & Hoffman, B. AP-1 (Fos/Jun) transcription factors in hematopoietic differentiation and apoptosis. *Int. J. Oncol.* **12**, 685–700. <https://doi.org/10.3892/ijo.12.3.685> (1998).
25. Karakaslar, E. O. et al. Transcriptional activation of Jun and Fos members of the AP-1 complex is a conserved signature of immune aging that contributes to inflamaging. *Aging Cell* **22**, e13792. <https://doi.org/10.1111/ace.13792> (2023).
26. Qiu, X. et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982. <https://doi.org/10.1038/nmeth.4402> (2017).
27. Tarashansky, A. J., Xue, Y., Li, P., Quake, S. R. & Wang, B. Self-assembling manifolds in single-cell RNA sequencing data. *eLife* **8**, e48994. <https://doi.org/10.7554/eLife.48994> (2019).
28. Qi, Q. et al. Dynamic CTCF binding directly mediates interactions among cis-regulatory elements essential for hematopoiesis. *Blood* **137**, 1327–1339. <https://doi.org/10.1182/blood.2020005780> (2021).
29. Mosler, T. et al. R-loop proximity proteomics identifies a role of DDX41 in transcription-associated genomic instability. *Nat. Commun.* **12**, 7314. <https://doi.org/10.1038/s41467-021-27530-y> (2021).

30. Cristini, A., Groh, M., Kristiansen, M. S. & Gromak, N. RNA/DNA hybrid interactome identifies DXH9 as a molecular player in transcriptional termination and R-loop-associated DNA damage. *Cell Rep.* **23**, 1891–1905. <https://doi.org/10.1016/j.celrep.2018.04.025> (2018).
31. Doty, R. T. et al. The transcriptomic landscape of normal and ineffective erythropoiesis at single-cell resolution. *Blood Adv.* **7**, 4848–4868. <https://doi.org/10.1182/bloodadvances.2023010382> (2023).
32. Bruehl, F. K. et al. Peripheral blood and bone marrow findings in treatment-naïve patients with cytopenia(s)/myeloid neoplasms harboring both a germline and a somatic DDX41 mutation. *Appl. Immunohistochem. Mol. Morphol.* **32**, 371–381. <https://doi.org/10.1097/pai.00000000000001215> (2024).
33. Jakobsen, N. A. et al. Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging. *Cell Stem Cell* **31**, 1127–1144.e1117. <https://doi.org/10.1016/j.stem.2024.05.010> (2024).
34. Serrano, G. et al. Single-cell transcriptional profile of CD34+ hematopoietic progenitor cells from del(5q) myelodysplastic syndromes and impact of lenalidomide. *Nat. Commun.* **15**, 5272. <https://doi.org/10.1038/s41467-024-49529-x> (2024).
35. Dimitrov, D. et al. Comparison of methods and resources for cell-cell communication inference from single-cell RNA-Seq data. *Nat. Commun.* **13**, 3224. <https://doi.org/10.1038/s41467-022-30755-0> (2022).
36. Gao, S. et al. Identification of HSC/MPP expansion units in fetal liver by single-cell spatiotemporal transcriptomics. *Cell Res.* **32**, 38–53. <https://doi.org/10.1038/s41422-021-00540-7> (2022).
37. Ennis, S. et al. Cell-cell interactome of the hematopoietic niche and its changes in acute myeloid leukemia. *iScience* **26**, 106943. <https://doi.org/10.1016/j.isci.2023.106943> (2023).
38. Sen Santara, S. et al. The NK cell receptor NKp46 recognizes ecto-calreticulin on ER-stressed cells. *Nature* **616**, 348–356. <https://doi.org/10.1038/s41586-023-05912-0> (2023).
39. Chao, M. P. et al. Calreticulin is the dominant pro-phagocytic signal on multiple human cancers and is counterbalanced by CD47. *Sci. Transl. Med.* **2**, 63ra94. <https://doi.org/10.1126/scitranslmed.3001375> (2010).
40. Lee, G. et al. Novel secreted isoform of adhesion molecule ICAM-4: potential regulator of membrane-associated ICAM-4 interactions. *Blood* **101**, 1790–1797. <https://doi.org/10.1182/blood-2002-08-2529> (2003).
41. Weinreb, J. T., Gupta, V., Sharvit, E., Weil, R. & Bowman, T. V. Ddx41 inhibition of DNA damage signaling permits erythroid progenitor expansion in zebrafish. *Haematologica* **107**, 644–654. <https://doi.org/10.3324/haematol.2020.257246> (2022).
42. Wulfridge, P. & Sarma, K. Intertwining roles of R-loops and G-quadruplexes in DNA repair, transcription and genome organization. *Nat. Cell Biol.* **26**, 1025–1036. <https://doi.org/10.1038/s41556-024-01437-4> (2024).
43. Bi, H. et al. DDX41 dissolves G-quadruplexes to maintain erythroid genome integrity and prevent cGAS-mediated cell death. *bioRxiv* <https://doi.org/10.1101/2024.10.14.617891> (2024).
44. Trowbridge, J. J. & Starczynowski, D. T. Innate immune pathways and inflammation in hematopoietic aging, clonal hematopoiesis, and MDS. *J. Exp. Med.* <https://doi.org/10.1084/jem.20201544> (2021).
45. Osterhoudt, K., Bagno, O., Katzman, S. & Zahler, A. M. Spliceosomal helicases DDX41/SACY-1 and PRP22/MOG-5 both contribute to proofreading against proximal 3' splice site usage. *RNA* **30**, 404–417. <https://doi.org/10.1261/rna.079888.123> (2024).
46. Tsukamoto, T. et al. Insights into the involvement of spliceosomal mutations in myelodysplastic disorders from analysis of SACY-1/DDX41 in *Caenorhabditis elegans*. *Genetics* **214**, 869–893. <https://doi.org/10.1534/genetics.119.302973> (2020).

Acknowledgements

We wish to thank Ms. Manami Shimomura, Mr. Ryo Takahashi, and Mr. Yoshimasa Imamura at ImmunoGeneTeqs, Inc. for their technical assistance related to sequencing data acquisition and Ms. Michie Yoshikawa for her clerical work. The super-computing resource was provided by the Human Genome Center, the Institute of Medical Science, the University of Tokyo.

Author contributions

H.M. designed the study; H.M. and S.S. wrote the manuscript; H.M., A.N., M.K., and R.K. conducted experiments; H.M., M.A., and S.S. analyzed sequencing data; H.H. and Y.H. assisted with material and data collection; A.Y. and Y.H. supervised the study; all authors approved the final manuscript; all authors are accountable for all aspects of the work.

Funding

This study was supported by the Princess Takamatsu Cancer Research Fund (H.M.), the SENSIN Medical Research Foundation (H.M.), the Joint Usage/Research Center for Radiation Disaster Medical Science (H.M.), the Japanese Society of Hematology (H.M.), and Japan Society for the Promotion of Science (No. 25K11638; H.M, No. 22H05064; S.S.). The funders did not play a role in the study design, data collection, analysis, interpretation, or writing of this manuscript.

Declarations

Competing interests

S.S. reports an advisory role for ImmunoGeneTeqs, Inc.; stock for ImmunoGeneTeqs, Inc. but declares no non-financial competing interests. H.M. serves as a member of the journal's editorial board. The other authors declare no financial or non-financial competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-06477-w>.

Correspondence and requests for materials should be addressed to H.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025