Received: 30 August 2024 | Accepted: 8 December 2024

DOI: 10.1002/hem3.70073

ARTICLE

HemaSphere eha

Somatic mutations and DNA methylation identify a subgroup of poor prognosis within lower-risk myelodysplastic syndromes

David Rombaut[1,^] | Sarah Sandmann[2,^] | Tobias Tekath[2] | Simon Crouch[3] | Aniek O. de Graaf[4] | Alexandra Smith[3] | Daniel Painter[3] | Olivier Kosmider[1] | Magnus Tobiasson[5] | Andreas Lennartsson[6] | Bert A. van der Reijden[4] | Sophie Park[7] | Maud D'Aveni[8] | Borhane Slama[9] | Emmanuelle Clappier[10] | Pierre Fenaux[11] | Lionel Adès[11] | Arjan van de Loosdrecht[12] | Saskia Langemeijer[13] | Argiris Symeonidis[14] | Jaroslav Čermák[15] | Claude Preudhomme[16] | Aleksandar Savic[17] | Ulrich Germing[18] | Reinhard Stauder[19] | David Bowen[20] | Corine van Marrewijk[13] | Elsa Bernard[21] | Theo de Witte[22] | Julian Varghese[2] | Eva Hellström-Lindberg[5] | Martin Dugas[23] | Joost Martens[24] | Luca Malcovati[25] | Joop H. Jansen[4,^] | Michaela Fontenay[1,^] | MDS-RIGHT consortium

Correspondence: Joop H. Jansen (joop.jansen@radboudumc.nl); Michaela Fontenay (michaela.fontenay@aphp.fr)

Abstract

Lower risk (LR) myelodysplastic syndromes (MDS) are heterogeneous hematopoietic stem and progenitor disorders caused by the accumulation of somatic mutations in various genes including epigenetic regulators that may produce convergent DNA methylation patterns driving specific gene expression profiles. The integration of genomic, epigenomic, and transcriptomic profiling has the potential to spotlight distinct LR-MDS categories on the basis of pathophysiological mechanisms. We performed a comprehensive study of somatic mutations and DNA methylation in a large and clinically well-annotated cohort of treatment-naive patients with LR-MDS at diagnosis from the EUMDS registry (ClinicalTrials.gov.NCT00600860). Unsupervised clustering analyses identified six clusters based on genetic profiling that concentrate into four clusters on the basis of genome-wide methylation profiling with significant overlap between the two clustering modes. The four methylation clusters showed distinct clinical and genetic features and distinct methylation landscape. All clusters shared hypermethylated enhancers enriched in binding motifs for ETS and bZIP (C/EBP) transcription factor families, involved in the regulation of myeloid cell differentiation. By contrast, one cluster gathering patients with early leukemic evolution exhibited a specific pattern of hypermethylated promoters and, distinctly from other clusters, the upregulation of AP-1 complex members FOS/FOSL2 together with the absence of hypermethylation of their binding motif at target gene enhancers, which is of relevance for leukemic initiation. Among MDS patients with lower-risk IPSS-M, this cluster displayed a significantly inferior overall survival ($p < 0.0001$). Our study showed that genetic and DNA methylation features of LR-MDS at early stages may refine risk stratification, therefore offering the frame for a precocious therapeutic intervention.

INTRODUCTION

Lower-risk myelodysplastic syndromes (LR-MDS) are a group of heterogeneous disorders with variable severity of hematopoietic impairment and kinetics of disease progression, making clinical decision-making difficult in terms of timing and modality of therapeutic intervention.[1–4] While complications related to chronic cytopenias represent the most frequent causes of death, recognizing patients at higher risk of clonal progression is critical to inform therapeutic intervention.[2,3] Based on the prospective unbiased patient population recruited within the European MDS (EUMDS) Registry, we reported transfusion dependency, kinetics of cytopenia, and toxic iron species, as relevant prognostic factors.[5–7] Facing a spectrum of disorders comprising indolent conditions, we identified meaningful surrogate endpoints such as changes in Health-related Quality of Life (HRQoL) and RBC transfusion-free survival.[8,9]

Identification of somatic mutations in a high proportion of patients and recognition of genotype–phenotype correlations have paved the way for a molecular classification.[10–15] Genomic profiling can shed light on premalignant conditions,[16] identify subsets with a distinctive molecular basis,[17–19] and recognize patients at higher risk of clonal progression, requiring early disease-modifying treatments.[20,21] Mutations in epigenetic regulators involved in DNA methylation and histone modification are prevalent in more than half of MDS.[10,11,22,23] *TET2* loss-of-function mutations install early epigenetic changes and are associated with hypermethylation of regulatory regions.[24] Aberrant methylation profiles were associated with progression and response to treatment with hypomethylating agents (HMA).[25–30] Furthermore, MDS patients without epigenetic mutations show epigenetic alterations, suggesting convergent mechanisms.[31] The integration of genomic and epigenomic profiling has the potential to give insight into the pathogenic mechanisms of MDS.

Based on unsupervised clustering analyses of somatic mutations and DNA methylation in a large cohort of LR-MDS patients, we aimed to identify clinically and biologically relevant groups. Our study allowed recognizing divergent trajectories at early stages of the disease.

MATERIALS AND METHODS

Patients

The EUMDS Registry is recording consecutive newly diagnosed patients within three months of diagnosis. The data set TRIAGE comprised 543 LR-MDS cases based on IPSS risk low or intermediate-1 (Table 1). A French cohort of 175 LR-MDS patients at diagnosis consecutively enrolled after consent for biological sample collection and individual follow-up of 2 years (median 24 months [confidence interval [CI] 95%: 18.9–32.3]) was considered for methylome and transcriptome studies. This cohort included 28 cases who progressed to leukemia (MDSp) with a median of 15.5 months [CI 95%: 8.5–18.7]. Among them, 156 were included in the genetic study (Supporting Information S1: Figure 1). The procedures followed the statements of Helsinki's Declaration. EUMDS Registry (NCT00600860) was approved by the Institution's Ethics Committees according to national legislation. The French cohort was approved by IRB IdF-X 2010-A00033-36-2753.

Somatic mutation analysis

DNA was extracted from the bone marrow mononuclear cell (BMMNC) fraction using the DNA/RNA Kit (Qiagen). Sequencing of target regions (Supporting Information S1: Table 1) was performed using a custom 27-gene panel of single-molecule-tagged molecular inversion probes (smMIPs) on NextSeq500 platform and a 55-gene panel on NovaSeq for IPSS-M calculation[32] (Illumina). Variants were called with appreci8 using ≥5% VAF, a minimum of 2 mutated reads.[33]

DNA methylation profiling and bioinformatics analysis

DNA methylation was analyzed using Infinium MethylationEPIC 850 K BeadChip Array (Illumina; Supporting Information). The methylation level at each CpG was estimated as β-values using the ratio of intensities between methylated ($\beta = 1$) and unmethylated ($\beta = 0$) alleles. For unsupervised clustering, CpGs were filtered based on the

[1]Université Paris Cité, Institut Cochin, INSERM U1016, CNRS UMR8104, Assistance Publique-Hôpitaux de Paris.Centre, Laboratory of Hematology, Hôpital Cochin, Paris, France
[2]Institute of Medical Informatics, University of Münster, Münster, Germany
[3]Epidemiology and Cancer Statistics Group, Department of Health Sciences, University of York, York, UK
[4]Department of Laboratory Medicine, Laboratory of Hematology, Radboud University Medical Center, Nijmegen, The Netherlands
[5]Center for Hematology and Regenerative Medicine, Department of Medicine Huddinge, Karolinska Institute, Karolinska University Hospital, Stockholm, Sweden
[6]Department of Biosciences and Nutrition, Karolinska Institute, Stockholm, Sweden
[7]Department of Hematology, Université de Grenoble-Alpes, CHU, Grenoble, France
[8]Service d'Hématologie Clinique, University Hospital of Nancy and University of Lorraine, Nancy, France
[9]Service d'onco-hématologie, Centre Hospitalier Général d'Avignon, Avignon, France
[10]Université Paris Cité, Assistance Publique des Hôpitaux de Paris.Nord, Laboratoire d'Hématologie, Hôpital Saint-Louis, Paris, France
[11]Université Paris Cité, Assistance Publique des Hôpitaux de Paris.Nord, Service d'Hématologie Senior, Hôpital Saint-Louis, Paris, France
[12]Department of Hematology, Amsterdam UMC, Cancer Center Amsterdam, Amsterdam, The Netherlands
[13]Department of Hematology, Radboud University Medical Center, Nijmegen, The Netherlands
[14]Hematology Division, Department of Internal Medicine, University of Patras, Patras, Greece
[15]Department of Clinical Hematology, Institute of Hematology and Blood Transfusion, Prague, Czech Republic
[16]Laboratoire d'hématologie, Centre Hospitalier Régional Universitaire, Lille, France
[17]Clinic of Hematology, Clinical Center of Vojvodina, Faculty of Medicine, University of Novi Sad, Novi Sad, Serbia
[18]Department of Hematology, Oncology and Clinical Immunology, Heinrich-Heine-University, Medical Faculty, Düsseldorf, Germany
[19]Department of Internal Medicine V (Haematology and Oncology), Comprehensive Cancer Center Innsbruck, Medical University of Innsbruck, Innsbruck, Austria
[20]St. James's Institute of Oncology, Leeds Teaching Hospitals, Leeds, UK
[21]Department of Computational Biology, Institut Gustave Roussy, INSERM U981, Villejuif, France
[22]Department of Tumor Immunology, Radboud Institute of Molecular Life Sciences, Radboud University Medical Center, Nijmegen, The Netherlands
[23]Institute of Medical Informatics, University of Heidelberg, Heidelberg, Germany
[24]Department of Molecular Biology, Faculty of Science, Radboud University, Nijmegen, The Netherlands
[25]Department of Molecular Medicine and Department of Hematology Oncology, University of Pavia and Fondazione IRCCS Policlinico S. Matteo, Pavia, Italy
^David Rombaut and Sarah Sandmann are first co-authors. Joop H. Jansen and Michaela Fontenay are co-corresponding authors.

TABLE 1 Clinical and hematologic features of individuals included in the study.

Number of patients	543
Age (mean, range)	71.9 (20.9–97.5)
Sex (M/F)	305/238
Hemoglobin (g/dL), mean (range)	9.9 (3.8–15.0)
MCV (fL), median (range)	99 (58–134)
WBC ($\times 10^9$/L), mean (range)	5.2 (0.9–18.8)
ANC ($\times 10^9$/L), mean (range)	3.0 (0.0–11.9)
AMC ($\times 10^9$/L), mean (range)	0.4 (0.0–4.3)
Monocytes (%), mean (range)	8 (0.0–43.5)
Platelets ($\times 10^9$/L), mean (range)	204 (8–904)
BM blasts (%), mean (range)	2.6 (0.0–10.0)
RS (%), mean (range)	16 (0–92)
WHO category, *n* (%)	
MDS-SLD	56 (10.3)
MDS-MLD	196 (36.1)
MDS-RS-SLD	90 (16.6)
MDS-RS-MLD	53 (9.8)
MDS del(5q)	26 (4.7)
MDS-EB-1	84 (15.5)
MDS unclassified	21 (3.9)
Missing	17 (3.1)
IPSS risk group, *n* (%)	
Low	263 (48.5)
Intermediate-1	225 (41.4)
Missing	55 (10.1)
IPSS-R risk group, *n* (%)	
Very low	127 (23.4)
Low	240 (44.2)
Intermediate	81 (14.9)
High	16 (3.0)
Missing	79 (14.5)
Karyotype (normal/abnormal)	318/140

Abbreviations: AMC, absolute monocyte count; ANC, absolute neutrophil count; BM, bone marrow; EB1, excess of blasts 5%–9%; IPSS-R, International Prognostic Scoring System-Revised; MCV, median corpuscular volume; MLD, multiple lineage dysplasia; RS, ring sideroblasts; SLD, single lineage dysplasia; WBC, white blood cell.

standard deviation (SD) of normalized β-values. The 5000 CpGs with the highest SD values were retained (SD > 0.195 754) in order to realize unsupervised clustering with the K-means algorithm (100 initializations with 1000 max iterations each). A hierarchical clustering was then performed using a selection of the 500 highest mean CpG and 500 lowest mean CpG from each of the 4 clusters identified by unsupervised clustering. Differentially methylated CpGs were defined through pairwise comparisons between clusters and controls with a $\Delta\beta$-value >|0.20|. Differentially methylated regions (DMR)-CpG calling was performed with DMRcate package, each requiring ≥2 minProbes, with an adjusted $p < 0.05$ and a false discovery rate (FDR) < 0.001. The re-annotation of Illumina Infinium MethylationEPIC CpGs (including enhancer positions) by Bizet et al. was used.[34]

Transcription factor (TF) motif enrichment analyses of hypermethylated enhancer and bivalent enhancer/promoter regions were performed using HOMER's script findmotifsGenome. For enrichment analysis, sequences of 125-bp flanking each of the CpG sites located in these regions were defined (-size 250), background selection was set as default, -nomotif option was used (no search for de novo motif enrichment), and enrichment values were calculated as *p*-values of the Fischer's exact test with correction for the number of tested TF binding motifs in HOMER. The motifs representing at least 10% of the sequences and containing a cytosine preceding a guanine were considered. For specific methylated CpG in motifs search, a -size of 20 was used, with a motif file containing all 440 motifs from HOMER (v4.11.1), and the same parameters as above.

Messenger RNA (mRNA) profiling

Libraries were prepared from BMMNC using TruSeq Stranded mRNA Sample Preparation Kit (Illumina) and sequenced paired-end (2 × 75 bp) on NextSeq550 platform. FASTQ sequences were aligned on hg19/GRCh37 with STARv.2.7.3a. The mean depth of sequence coverage was 47 (IQR: 38–48) million reads. Abundance estimation of cell types for RNA-seq data (for samples with both methylation and RNA-seq data) was realized with CIBERSORT. The reference signature matrix was created with publicly available single-cell RNA-seq data of purified $CD34^+$ BM cells with the following parameters: minimum expression = 1, replicates = 30, and sampling = 0.5 for RNA input options, and $\kappa = 13$, $q = 10^{-5}$, and a number of barcode genes between 3 and 50 for additional options.[35,36] Normalized counts of 134 RNA-seq samples were used to impute cell fractions with B-mode batch correction, absolute mode, and 500 permutations for significance analysis. Differentially expressed genes were identified with DESeqv.1.3.0.1 ($S < 0.01$). Pathways were determined in GeneOntology using Panther16.0 and biological process data set.

Statistical analyses

Numerical variables were summarized by median, mean, and range; categorical variables were described with count and frequency (%). Comparison of numerical variables was carried out using nonparametric tests (Mann–Whitney, Kruskal–Wallis analysis of variance). Comparison of categorical variables was performed with Fisher's exact or Chi-square tests. Genetic hierarchical clustering based on the Jaccard distance was used with the Ward D.2 clustering method. Cluster optimal number was chosen using the Kelley–Gardner–Sutcliffe method.[37] For comparison of clustering methods, standardized residuals from the Chi-square test were used to determine the largest deviations from randomness. Relations between variables were studied by logistic regression to obtain odds ratios with confidence intervals. Multivariate analyses were performed by means of Cox proportional hazards regression. Survival analyses were performed with the Kaplan–Meier method. Endpoints included all-cause mortality, time to progression to higher-risk MDS and AML or to AML, and time to first treatment.

RESULTS

Unsupervised clustering analysis based on somatic mutation profiles

Four hundred twenty-two out of 543 patients harbored one or more somatic mutations (78%) (Supporting Information S1: Figure 2A). One hundred forty patients showed an abnormal karyotype with 6% del(5q)

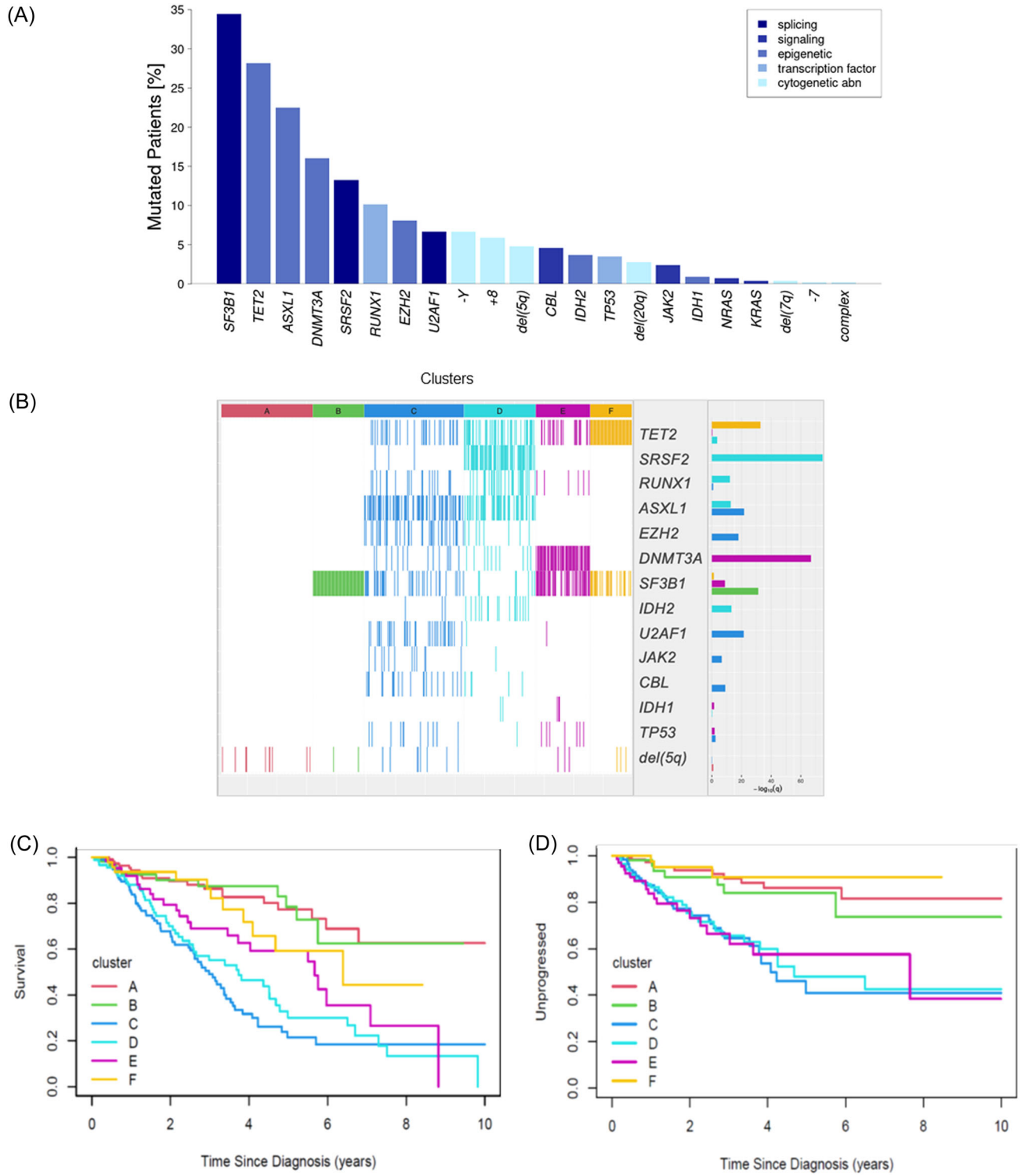


FIGURE 1 **Unsupervised genetic clustering and clinical outcomes of the six clusters of lower risk myelodysplastic syndromes (LR-MDS). (A)** Frequency of genetic lesions including somatic mutations and chromosomal anomalies. **(B)** Heatmap of the genetic clusters. The tightest clusters are A and B, followed by cluster F (mean Jaccard distance 0.25), cluster E (0.50), cluster D (0.57), and finally cluster C (0.81). Cluster tightness is strongly associated with the number of genetic features that determine the clusters. Mutation enrichment is indicated as $-\log_{10}(q)$. **(C)** Overall survival of LR-MDS patients stratified by cluster ($p < 0.001$). Compared to cluster A, clusters B and F did not show significantly different survival in multivariable analysis (hazard ratio [HR] 0.87, $p = 0.72$; HR 1.45, $p = 0.33$, respectively), while C, D, and E showed a significantly worse outcome (HR 4.14, $p < 0.001$; HR 3.57, $p < 0.001$; HR 2.54, $p = 0.002$, respectively). **(D)** Risk of disease progression analyzed as a composite endpoint of progression into higher-risk MDS or acute myeloid leukemia (AML). The six clusters showed significantly different time to progression ($p < 0.001$): clusters B and F did not show significantly different survival compared to cluster A (HR 1.57, $p = 0.36$; HR 0.78, $p = 0.70$, respectively), while C, D, and E showed a significantly worse outcome (HR 4.93, $p < 0.001$; HR 4.47, $p < 0.001$; HR 4.43, $p < 0.001$, respectively).

TABLE 2 Characteristics of LR-MDS patients according to genetic clusters.

Variables	Genetic cluster A	B	C	D	E	F	*p* Value
Number of patients	121	68	132	95	72	55	–
Age, mean (range)	69.4 (20.9–89.5)	72.4 (41.4–92.7)	73.8 (51.0–90.2)	72.1 (39.1–91.6)	71.5 (44.5–97.5)	72.4 (45.6–93.1)	0.058
Sex (M/F)	61/60	38/30	83/49	61/34	32/40	30/25	0.059
Hemoglobin (g/dL), mean (range)	10.2 (5.0–14.2)	9.5 (3.8–14.3)	9.5 (4.4–14.6)	10.4 (6.4–15.0)	9.5 (6.4–14.1)	10.2 (7.3–14.4)	<0.001
MCV (fL), median (range)	97 (59–122)	103 (68–125)	99 (72–129)	96 (58–124)	100 (74–134)	101 (68–127)	0.004
WBC ($\times 10^9$/L), mean (range)	4.8 (0.9–9.5)	5.4 (2.0–12.4)	5.5 (1.0–18.8)	5.1 (1.0–18.5)	5.4 (1.2–13.6)	4.7 (1.1–15.7)	0.251
ANC ($\times 10^9$/L), mean (range)	2.9 (0.1–7.2)	3.0 (0.1–8.9)	3.4 (0.2–11.9)	2.8 (0.0–11.2)	3.2 (0.1–11.8)	2.6 (0.1–9.8)	0.225
AMC ($\times 10^9$/L), mean (range)	0.4 (0.0–1.3)	0.4 (0.0–1.1)	0.4 (0.0–1.5)	0.5 (0.0–4.3)	0.3 (0.0–1.6)	0.4 (0.0–1.4)	0.193
Monocytes (%), mean (range)	7.7 (0.1–26.2)	6.8 (0.4–15.3)	7.5 (0.0–43.5)	9.2 (0.0–33.0)	6.5 (0.0–19.0)	11.0 (0.3–33.0)	0.001
Platelets ($\times 10^9$/L), mean (range)	196 (16–714)	281 (46–706)	198 (8–904)	149 (21–557)	235 (51–587)	195 (11–496)	<0.001
Serum Ferritin (ng/mL), mean (range)	461 (13–2261)	700 (6–1799)	758 (31–3775)	777 (16–8000)	769 (17–6420)	592 (38–2924)	0.156
Transferrin saturation (%), mean (range)	34.7 (5.0–98.0)	59.3 (29.0–95.0)	51.6 (7.0–111)	42.3 (8.0–95.0)	46.7 (12.0–92.0)	46.2 (12.0–91.0)	0.001
BM blasts (%), mean (range)	2.1 (0.0–9.0)	2.0 (0.0–8.6)	3.0 (0.0–9.0)	3.6 (0.0–9.0)	2.3 (0.0–9.0)	2.6 (0.0–9.0)	<0.001
RS (%), mean (range)	5.2 (0.0–74.0)	37.8 (0.0–85.0)	17.5 (0.0–92.0)	7.6 (0.0–87.0)	22.1 (0.0–90.0)	17.1 (0.0–70.0)	<0.001
WHO 2016, *n* (%)							<0.001
MDS-SLD	26 (21.5)	1 (1.5)	11 (8.3)	8 (8.4)	6 (8.3)	4 (7.3)	
MDS-MLD	62 (51.2)	11 (16.2)	53 (40.2)	36 (37.9)	14 (19.4)	20 (36.4)	
MDS-RS-SLD	8 (6.6)	33 (48.5)	13 (9.8)	3 (3.2)	24 (33.3)	9 (16.4)	
MDS-RS-MLD	2 (1.7)	12 (17.6)	13 (9.8)	4 (4.2)	11 (15.3)	11 (20.0)	
MDS del(5q)	10 (8.3)	2 (2.9)	8 (6.1)	0 (0.0)	3 (4.2)	3 (5.5)	
MDS-EB-1	10 (8.3)	5 (7.4)	27 (20.5)	30 (31.5)	6 (8.4)	6 (10.9)	
MDS unclassified	3 (2.5)	1 (1.5)	1 (0.8)	10 (10.5)	4 (5.6)	2 (3.6)	
Missing	-	3 (4.4)	6 (4.5)	4 (4.2)	4 (5.6)	-	
IPSS-R risk group, *n* (%)							<0.001
Very low	40 (33.1)	18 (26.5)	24 (18.2)	13 (13.7)	16 (22.2)	16 (29.1)	
Low	60 (49.6)	32 (47.1)	57 (43.2)	34 (35.8)	31 (43.1)	26 (47.3)	
Intermediate	9 (7.4)	8 (11.8)	22 (16.7)	29 (30.5)	5 (6.9)	8 (14.5)	
High	1 (0.8)	1 (1.5)	10 (7.6)	3 (3.2)	1 (1.4)	0 (0.0)	
Missing	11 (9.1)	9 (13.2)	19 (14.4)	16 (16.8)	19 (26.4)	5 (9.1)	
Number of mutations, median (range)	0 (0–1)	1 (1–2)	2 (1–5)	3 (1–5)	2 (1–4)	2 (1–3)	<0.001

Abbreviations: AMC, absolute monocyte count; ANC, absolute neutrophil count; BM, bone marrow; EB1, excess of blasts 5%–9%; IPSS-R, International Prognostic Scoring System-Revised; MCV, median corpuscular volume; MLD, multiple lineage dysplasia; RS, ring sideroblasts; SLD, single lineage dysplasia; WBC, white blood cell.

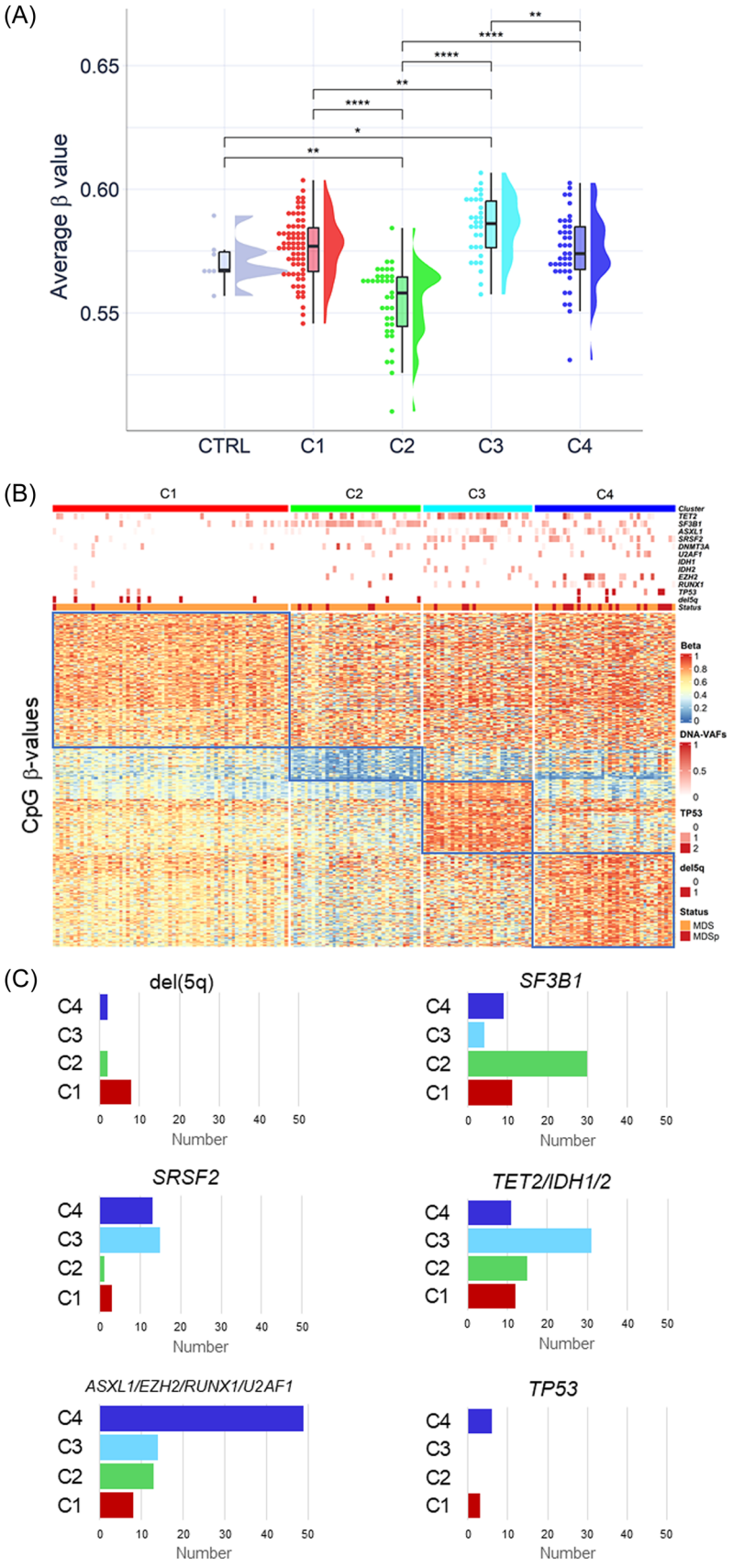


FIGURE 2 (See caption on next page).

FIGURE 2 **Unsupervised methylation-based clustering and patterns of genetic lesions**. Unsupervised clustering of lower-risk myelodysplastic syndromes (LR-MD) was obtained by the K-means method using the 5000 unique CpGs with the highest standard deviation (SD) of their normalized β-values and identified 4 clusters. **(A)** Raincloud plots representing the distribution of average β-value of samples in the 4 methylation clusters (C1, C2, C3, C4) and controls (CTRL). Box plots show the median ± SD and the whiskers represent the maximal and minimal values. Mann–Whitney test for *p*-values (*<0.05, **<0.01, ****<0.0001). **(B)** Heatmap representing the normalized β-value of selected CpG (row) including the 500 highest mean CpGs and the 500 lowest mean CpGs from each cluster after removal of CpG duplicates and hierarchical clustering. Columns represent patient samples. Disease and mutation status are indicated. DNA variant allele frequencies (VAFs) are shown as color shades. Presence or absence of del(5q) is indicated by 1 or 0, respectively, and the number of *TP53* mutations (0, 1, 2) per patient is indicated by distinct colors. **(C)** Bar plots representing the number of the indicated mutation or cytogenetic abnormality in each cluster.

(Figure 1A and Supporting Information S1: Figure 2B). At least one mutation and/or cytogenetic abnormality was found in 86.4%. As expected for LR-MDS, the spectrum of genetic lesions was biased compared to other unselected MDS populations, *SF3B1* being the most frequently mutated gene (34%) while *TP53* or *RUNX1* being underrepresented (Figure 1A).[14,15,33,36]

Unsupervised hierarchical clustering analysis based on genetic profile with the Jaccard distance measure identified an optimal cluster number of six (Supporting Information S1: Figure 3A,B). The six clusters showed a distinct genetic profile (Figure 1B). Cluster A gathered MDS-del(5q), those associated with *SF3B1* mutation segregating in other clusters. Cluster A comprised cases without proof of clonal hematopoiesis by mutation and cytogenetic analysis. A primary *SF3B1* cluster first divided into a cluster B characterized by isolated *SF3B1* mutation and a *SF3B1*/*DNMT3A*/*TET2*, which then divided into *SF3B1*/*DNMT3A* cluster E and *SF3B1*/*TET2* cluster F (Supporting Information S1: Figure 4A–D). We identified a cluster D almost invariably characterized by *SRSF2* mutation, mainly associated with *TET2* or *ASXL1* mutations, and a cluster C enriched in high-risk mutations including *ASXL1*, *EZH2*, *U2AF1*, and *TP53*. Notably, a distinct homogeneous subcluster, driven by *U2AF1* mutation, emerged from this cluster at a lower hierarchical level (Supporting Information S1: Figure 4D).

Clinical features, health-related quality of life, and outcomes of genetic clusters

The six genetically defined clusters showed distinct clinical features and outcomes. *SF3B1*-mutant clusters (B, E, F) showed lower hemoglobin, higher MCV, ferritin, and transferrin saturation, higher proportion of ring sideroblasts, and lower blast count. Cluster C showed higher BM blasts and cases with multilineage dysplasia. Cluster D showed higher BM blasts and monocyte counts and lower platelets compared to others (Table 2).

We analyzed the time to first treatment or specific treatment modalities, adjusted for age and sex, including erythropoiesis-stimulating agents (ESA), RBC transfusion, HMA, and lenalidomide. Relative to cluster A, cluster C showed a shorter time to first treatment ($p = 0.011$), RBC transfusion ($p = 0.001$), and HMA treatment ($p = 0.040$). A shorter time to HMA treatment was also noticed for cluster D ($p = 0.016$) and a shorter time to lenalidomide treatment in clusters B ($p = 0.003$), C ($p = 0.009$), E ($p = 0.031$), and F ($p = 0.008$). Conversely, time to ESA treatment was not different among the six clusters ($p = 0.380$) (Supporting Information S1: Table 2). A thorough analysis of patterns of correlation between mutations and HRQoL in a subgroup of 185 LR-MDS patients showed no significant effect on the EQ.5D summary scores (Supporting Information S1: Table 3). Focusing on EQ.5D components, patterns suggested a relationship between *SF3B1* or *SRSF2* mutations and mobility and usual actions arguing for divergent effect on HRQoL of these mutations.

We analyzed clinical outcomes including overall survival (OS) and risk of disease progression (Supporting Information S1: Figure 5A–C). The six genetic clusters showed different OS ($p < 0.001$) (Figure 1C). Compared to cluster A, clusters B, and F did not show different OS in multivariable analysis (HR 0.87, $p = 0.72$; HR 1.45, $p = 0.33$, respectively), while clusters C, D, and E showed a worse outcome (HR 4.14, $p < 0.001$; HR = 3.57, $p < 0.001$; HR = 2.54, $p = 0.002$, respectively). Notably, clusters B (isolated *SF3B1*) and F (*SF3B1*/*TET2*) showed similar outcome (HR 1.66, $p = 0.25$), while cluster E (*SF3B1*/*DNMT3A*) and cluster C (*SF3B1*/*ASXL1*/*EZH2*/*RUNX1*) showed a worse OS (HR 2.90, $p = 0.005$). The 6 clusters showed different times to progression, analyzed as a composite endpoint of progression into higher-risk MDS or AML ($p < 0.001$) (Figure 1D). In multivariable analysis, clusters B and F did not show different time to any progression compared to cluster A (HR 1.57, $p = 0.36$; HR 0.78, $p = 0.70$, respectively), while clusters C, D, and E showed a worse outcome (HR 4.93, $p < 0.001$; HR 4.47, $p < 0.001$; HR 4.43, $p < 0.001$, respectively). Finally, clusters C, D, and E showed a shorter time to progression into higher-risk MDS (HR 3.58, $p = 0.006$; HR 4.19, $p = 0.002$; HR 3.1, $p = 0.019$, respectively) (Supporting Information S1: Figure 5D). The genetic clustering stratified patients with poor outcome in three subgroups with distinct OS.

Identification of four methylation clusters with distinct mutations by unsupervised clustering

To improve our understanding of the variability of LR-MDS trajectories, we investigated the epigenetic landscape of 175 LR-MDS patients in comparison to 7 age-matched healthy controls. The methylation cohort displayed no differences from the TRIAGE cohort in terms of age, sex, and mutation frequencies (Supporting Information S1: Table 4). Fifty-nine patients (33.7%) had an abnormal karyotype. At least one mutation and/or cytogenetic abnormality was found in 150 (85.7%) cases. Mutations in *TET2* were detected in 58 (33%), *SF3B1* in 54 (31%), and *SRSF2* in 32 (18%) patients. The median mutation number per patient was 1 and common co-mutation patterns were observed (Supporting Information S1: Figure 6A–C). IPSS-M scoring[33] identified 6 groups of patients with different OS ($p < 0.0001$) and a distinct risk of AML evolution between very low (VL), low (L), and moderate low (ML) IPSS-M ≤ 0 and moderate high (MH), high (H), and very high (VH) IPSS-M > 0 ($p < 0.0001$; Supporting Information S1: Figure 6D,E). More precisely, IPSS-M refined the prognosis of patients with the most favorable outcome (Supporting Information S1: Table 5). The methylation cohort comprised 28 patients with early progression to AML (MDSp) within the 2-year follow-up. These patients displayed a lower mean platelet count ($p = 0.013$), a higher mean blast count ($p = 0.001$), a higher median number of mutations ($p < 0.0001$), and a significant enrichment in IPSS-R intermediate- and high-risk categories ($p = 0.0001$). Mutations in *ASXL1*, *EZH2*, *RUNX1*, and *TP53* were more frequent ($p = 0.009$) and IPSS-M was worse ($p < 0.0001$) in MDSp than in MDS (Supporting Information S1: Table 6).

The EPIC array allowed interrogating the methylation of 723,612 sites representing ~2% of the genome CpGs with an enrichment at promoters and enhancers. Genome-wide methylation level computed as mean normalized CpG β-values was not different between MDS,

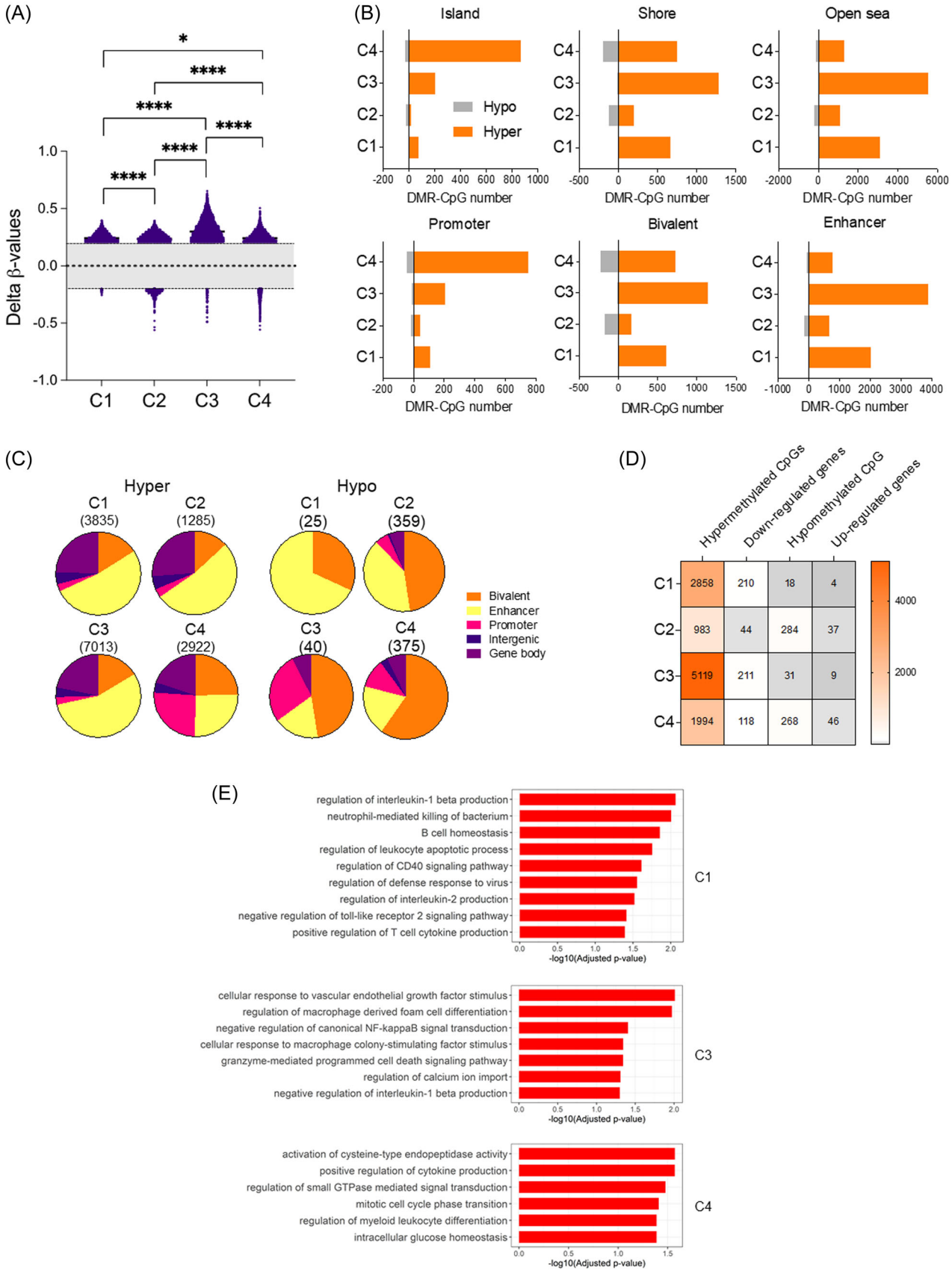


FIGURE 3 (See caption on next page).

FIGURE 3 **Differentially methylated CpGs within differentially methylated regions (DMR-CpG).** DMR-CpGs were defined as regions containing at least 2 CpGs with delta β >|0.20| and a Benjamini–Hochberg (BH)-adjusted $p < 0.05$. **(A)** Scatter plots representing the delta β-values of each hypermethylated and hypomethylated DMR-CpG in clusters C1, C2, C3, and C4 compared to controls. Mann–Whitney test for p-values. **<0.01; ****<0.0001. **(B)** Numbers of hypermethylated and hypomethylated DMR-CpGs in CpG islands, open seas, shores, bivalent enhancers/promoters, enhancers, and promoters. **(C)** Pie charts showing the proportion of hypermethylated (hyper) or hypomethylated (hypo) DMR-CpGs at bivalent enhancer and promoter, enhancer, gene body, intergenic region, and promoter identified in the comparisons of C1, C2, C3, and C4 to controls. Total numbers of DMR-CpGs are indicated in brackets. **(D)** Heatmap representing the numbers of hypermethylated or hypomethylated individual CpGs near referenced genes and the numbers of downregulated genes close to hypermethylated CpGs and of upregulated genes close to hypomethylated CpGs with $\log_2$ fold-change > |1| and $p < 0.05$ in each cluster. **(E)** Gene Ontology analyses of downregulated and upregulated gene sets near hypermethylated and hypomethylated CpGs, respectively, identified in **(D)**. Bar plots showing pathways significantly over-represented in clusters C1, C3, and C4 compared to controls. No significant terms were found in the comparison of cluster C2 with controls.

MDSp, and control group, while it was increased in *TET2*-mutated compared to *TET2*-wild-type samples or controls in correlation with mutation number (Supporting Information S1: Figure 7A–C). Unsupervised clustering using the 5000 CpGs with the highest standard deviation (SD) of their normalized β-values identified four clusters (Supporting Information S1: Figure 8A,B). Genome-wide methylation level of each cluster was similarly elevated in clusters 1 and 4 compared to controls, lower in cluster 2 compared to clusters 1, 3, and 4 ($p < 0.0001$) and controls ($p < 0.01$), and higher in cluster 3 compared to other clusters ($p < 0.001$) and controls ($p < 0.05$; Figure 2A). After selection of the highest mean CpGs and lowest mean CpGs from each cluster, a hierarchical clustering showed that a subset of hypomethylated CpG defined cluster 2, while two distinct subsets of hypermethylated CpGs defined clusters 3 and 4, respectively (Figure 2B). LR-MDS patients allocated to each cluster showed distinct genetic profiles (Figure 2B,C and Supporting Information S1: Table 7). Methylation cluster 1 was enriched in MDS-del(5q) (12%), or no evidence of clonality (42%). Cluster 2 was enriched in *SF3B1* mutation (81%) with *TET2* (38%) or *DNMT3A* mutation (22%). By logistic regression, *SF3B1* was predicted in cluster 2 with an odds ratio of 25.61 [95% CI: 9.00–84.44]; $p < 10^{-7}$). Hypermethylated cluster 3 was enriched in *TET2*/*IDH1/2* (74%/26%) and *SRSF2* (48%) mutations, which were predicted in cluster 3 with odds ratios of 16.66 ([95% CI: 5.91–52.81]; $p = 4.10^{-7}$) and 18.67 ([95% CI: 5.31–88.91]; $p = 3.10^{-7}$), respectively. Cluster 4 gathered cases with at least one mutation, mainly high-risk *RUNX1*, *EZH2*, *ASXL1*, *U2AF1*, and multi*TP53* mutations. *ASXL1* and *RUNX1* mutations were predicted in cluster 4 with odds ratios of 11.53 ([95% CI: 4.03–37.24]; $p = 10^{-5}$) and 16.76 ([95% CI: 4.08–114.58]; $p = 5.10^{-5}$), respectively.

A cross-tabulation of the methylation-based and the mutation-based unsupervised clustering methods demonstrated a strong correlation ($X^2 = 127.52$; $p < 10^{-5}$). The standardized residuals from the Chi-square test established that methylation clusters 1 and 3 highly correlated with genetic clusters A and D, respectively, methylation cluster 2 with both genetic clusters B and E, and methylation cluster 4 with genetic cluster C (Supporting Information S1: Table 8A). Furthermore, a cross-tabulation of the methylation-based unsupervised clustering and IPSS-M categorization demonstrated also a strong correlation ($X^2 = 60.80$; $p < 10^{-6}$) with cluster 1 correlated with the very low category, cluster 2 correlated with low category, and cluster 4 correlated with moderate high, high, and very high categories (Supporting Information S1: Table 8B). This shows that methylation-based clustering integrated specific molecular patterns and distinct risk categories.

DMR and differentially expressed gene sets in methylation clusters

To explore methylation changes between clusters, we analyzed DMR using DMRcate, as sequences containing at least 2 CpGs with a delta β-value >|0.20| and a BH-adjusted $p < 0.05$ further called DMR-CpGs (Supporting Information S1: Tables 9A–E). We identified 3860, 1642, 7053, and 3296 DMR-CpGs in the comparisons of clusters 1, 2, 3, and 4 with controls, respectively (Figure 3A), and 1880 DMR-CpGs in the comparison of MDSp with controls (Supporting Information S1: Tables 9A–E). Most of the DMR-CpGs were hypermethylated, and cluster 3, in which almost all samples were *TET2* or *IDH1/2* mutated, retained the highest number and the most elevated positive delta $\beta > 0.20$ of the hypermethylated DMR-CpGs compared to the 3 others, while clusters 2 and 4 had significantly more negative delta $\beta < -0.20$ corresponding to hypomethylated DMR-CpGs (Figure 3A and Supporting Information S1: Figure 9).

We then annotated the lists of hypermethylated or hypomethylated DMR-CpGs with CG-rich regions (CpG islands), shores (at up to 2 kb from CpG islands), and open seas (>4 kb from CpG islands) and with gene features including bivalent promoter/enhancer, enhancer, gene body, intergenic, and promoter (Supporting Information S1: Table 9A–D). To avoid underestimating enhancers, we used a refined probe annotation.[34] We found that the hypermethylated DMR-CpGs were predominant at CpG islands in cluster 4. In agreement with CpG islands occupying more than 90% of promoters, the proportion of hypermethylated DMR-CpGs at promoters was also higher in cluster 4 compared to controls. The TET2 hydroxylase is known to orchestrate the demethylation of enhancers through its recruitment to the DNA by interacting with TF, and conversely, inactivation of *TET2* gene leads to enhancer hypermethylation.[24] Consistently, *TET2* mutant-enriched cluster 3 displayed the highest number of hypermethylated enhancers. Hypermethylated DMR-CpGs overlapping with enhancers were not abundant in clusters 2 and 4, which, in contrast, displayed hypomethylated DMR-CpGs mainly in bivalent regulatory regions and in the shores of CpG islands (Figure 3B). Remarkably, the proportion of hypermethylated DMR-CpGs at promoters and the proportion of hypomethylated DMR-CpGs at bivalent regulatory regions were increased in cluster 4 compared to the others (Figure 3C).

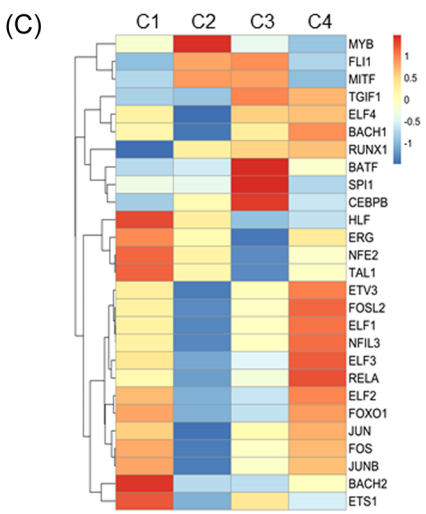
To investigate the impact of methylation changes on gene expression, RNA-sequencing data of 127/175 LR-MDS including 25 MDSp and 7 controls were generated. We studied the repartition of hematopoietic cell types among BM mononuclear cells using a reference signature matrix publicly available from single-cell RNA-seq data of BM $CD34^+$ cells and CIBERSORT algorithm, as previously (Supporting Information S1: Table 10).[35,36,38] We found similar abundances of common myeloid, megakaryocyte/erythroid, and granulocyte/monocyte progenitors across methylation clusters and controls while multi-lymphoid progenitors (MLPs) were less abundant in cluster 2 compared to cluster 4 and increased in clusters 1, 3, and 4 compared to controls (Supporting Information S1: Figure 10). Using DESeq2, we identified differentially downregulated or upregulated genes between clusters and controls (with log2 fold-change > |1| and $p < 0.05$). Then, the DMR-CpGs were annotated with the nearest differentially downregulated or upregulated genes (Supporting Information S1: Table 9A–D). Clusters 1 and 3 retained the highest

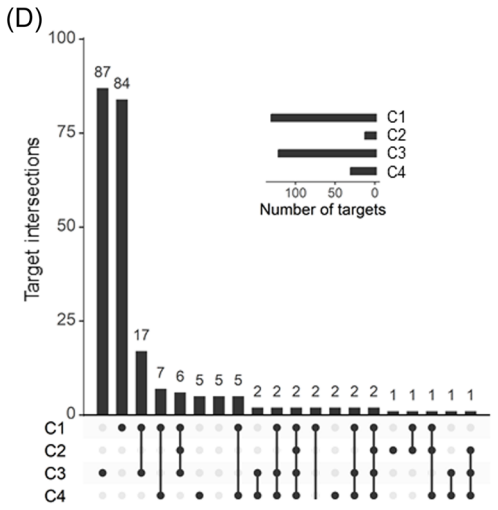
(A)

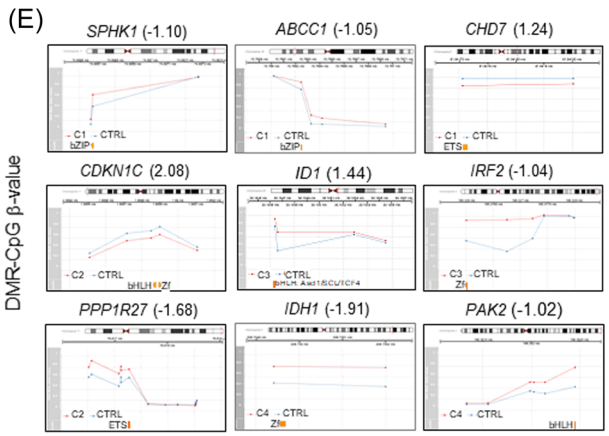
TF	Family	C1	C2	C3	C4
ERG	ETS				
ETV1/2/4	ETS				
ETS1	ETS				
ELF1/3/4/5	ETS				
EHF/Elk4	ETS				
FLI1	ETS				
PU.1/SpiB	ETS				
GABPA	ETS				
CEBP	bZIP				
HLF	bZIP				
NFIL3	bZIP				
CHOP	bZIP				
ATF4	bZIP				
FOS	bZIP				
FOSL2	bZIP				
BACH1/2	bZIP				
BATF	bZIP				
JUN	bZIP				
JUNB	bZIP				
ATF1/3	bZIP				
MITF	bHLH				
HIF2A/1B	bHLH				
SCL	bHLH				
TGIF2	Homeobox				
MYB	HTH				
RUNX1	Runt				
FOXO1	Forkhead				
NFκB-p65-Rel	RHD				

(B)

TF	Motif	C1	%	C2	%	C3	%	C4	%
ERG		1.00E-204	36	1.00E-45	35	1.00E-283	39	1.00E-22	30
ETV1		1.00E-207	32	1.00E-41	31	1.00E-251	34	1.00E-22	26
ETS1		1.00E-193	25	1.00E-32	24	1.00E-245	28	1.00E-24	21
Fli1		1.00E-180	26	1.00E-37	26	1.00E-202	27	1.00E-23	23
CEBP		1.00E-171	20	1.00E-70	25	1.00E-124	18	1.00E-32	18
NFIL3		1.00E-144	16	1.00E-48	18	1.00E-121	15	1.00E-27	14
HLF		1.00E-111	18	1.00E-43	20	1.00E-97	17	1.00E-19	16







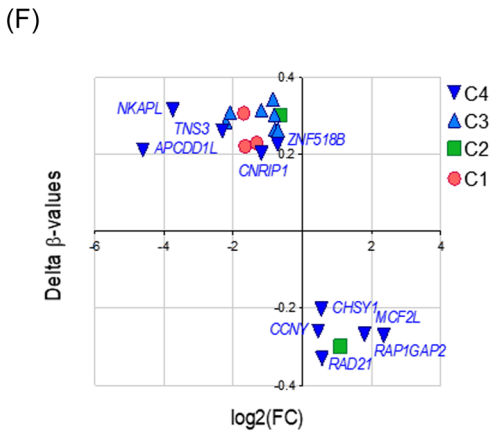


FIGURE 4 (See caption on next page).

FIGURE 4 **Differentially methylated enhancer and promoter gene target expression. (A)** Transcription factor (TF) binding motif enrichment analyses within stretches of 125-bp flanking on both sides of the hypermethylated CpG (delta β-value >|0.20| and Benjamini–Hochberg (BH)-adjusted $p < 0.05$) found in enhancers in the comparison of clusters C1, C2, C3, and C4 to control samples using HOMER. Shared and differential TF families which binding motifs were enriched ($q < 10^{-4}$) are shown. **(B)** Major TF binding motifs at hypermethylated enhancers shared between clusters. *p*-Values generated using Fischer's exact test with correction for the number of tested TF binding motifs and percentage of target sequences with motif are indicated. **(C)** Heatmap representing the expression levels shown as DESeq2 normalized counts transformed with a variance stabilizing transformation of transcription factors in each cluster compared to controls. **(D)** Upset plot showing intersections between clusters of the gene targets of differentially methylated enhancers. **(E)** Examples of genes with a TF motif (yellow squares) occupying the enhancers with hypermethylated or hypomethylated CpGs within differentially methylated regions (DMR-CpGs). DMR-CpG β-values are shown. Dots represent CpGs. The red line represents the β-values of CpGs in clusters C1, C2, C3, or C4 and the blue line represents the β-values of CpGs in controls. The $\log_2$ fold-change (FC) of gene expression between cluster and controls is indicated in brackets. **(F)** Scatter plots showing differentially expressed genes [$\log_2$(FC) >|0.5|; BH-adjusted $p < 0.05$] overlapping differentially methylated promoters (delta β > |0.20|; BH-adjusted $p < 0.05$) specific to clusters C1, C2, C3, or C4 in comparison with controls. Cluster C4 genes are annotated.

number of downregulated genes close to hypermethylated CpGs and clusters 2 and 4 retained the highest numbers of upregulated genes in the vicinity of hypomethylated CpGs (Figure 3D). We performed a Gene Ontology analysis of these gene sets. The main pathways over-represented in cluster 1 involved B cell homeostasis and the regulation of innate immune response. We cannot exclude a relationship between the emergence of B cell homeostasis or CD40 signaling pathways and the increased abundancy of MLPs in cluster 1. No specific pathways were found in cluster 2. Pathways emerging in cluster C3 involved the cellular response to M-CSF, a key factor for monocytic differentiation, macrophage differentiation, and the negative regulation of IL1β production and NFκB signaling. In cluster 4, the main pathways involved the regulation of cell cycle and differentiation, caspase activation, and glucose metabolism (Figure 3E).

Altogether, our unsupervised clustering unveiled distinct profiles of methylation aberrancies. MDS with short time to AML mostly gathered in cluster 4 were characterized by an enrichment in hypermethylated promoters and hypomethylated bivalent regulatory regions and a specific transcriptomic profile.

Landscape of TF binding motifs occupying differentially methylated enhancers

Aberrant methylation of the binding sites for TF overlapping enhancers may perturb their function that is required for hematopoietic cell development.[39] To explore this, we carried out motif enrichment analyses within stretches of 125-bp flanking differentially methylated CpG on both sides using HOMER. The enrichment in motifs for the ETS family and bZIP TFs was highly significant in the four clusters. Among a large set of 440 unique TF motifs in HOMER, 55, 30, 93, 28, and 29 binding motifs were significantly enriched ($q < 0.0001$) at hypermethylated enhancers (delta $\beta > 0.20$) in clusters 1, 2, 3, 4, and MDSp, respectively (Supporting Information S1: Table 11A). Hypermethylated binding motifs for the ETS TF family including ETS1 and SPI1, several bZIP TF families including C/EBP involved in myeloid cell differentiation, and HTH TF MYB were the most frequently enriched in all samples (Figure 4A,B). However, clusters might display distinct patterns of enrichment for other bZIP TF binding motifs in hypermethylated enhancers. Notably, the hypermethylated motifs for FOS and FOSL2 were enriched in clusters 1, 2, and 3, but not in cluster 4. The motifs for BACH1/2, BATF, JUN, and JUNB were enriched in clusters 1 and 3, but not in clusters 2 and 4. A hypermethylated motif for bHLH TF SCL/Tal1 involved in hematopoietic lineage specification was enriched in clusters 1 and 3. Moreover, cluster 2 displayed a specific pattern of hypomethylated motifs for zinc finger (Zf) GATA1 TF family that was consistent with the maintenance of an erythroid commitment (Supporting Information S1: Table 11B). Altogether, the hypermethylation of TF motifs in clusters 1 and 3 illustrates the uncoupling between cell proliferation and differentiation processes characterizing the hematopoiesis of myeloid malignancies. Conversely, in cluster 4, the absence of hypermethylated motifs at enhancers for AP-1 complex members FOS, FOSL2, JUN, and JUNB, together with the upregulation of *FOS*, *FOSL2*, *JUN*, and *JUNB* transcripts which has been implicated in AML initiation[39] suggest that downstream pathways could be activated (Supporting Information S1: Table 11A and Figure 4C). The profile of TFs which binding motifs at enhancers were hypermethylated, was almost similar between cluster 4 and MDSp, notably, the lack of hypermethylated motifs at enhancers for AP-1 complex members (Supporting Information S1: Table 11A and Supporting Information S1: Figure 11). Finally, the absence of hypermethylated motifs for AP-1 complex members at enhancers, together with the downregulation of AP-1 transcripts in cluster 2, suggests a distinct cell fate (Figure 4A,C).

Regulation of gene expression by enhancer and promoter methylation

As a proxy of enhancer activity, we analyzed differentially expressed target genes of enhancers that contained one differentially methylated CpG (delta $\beta > 0.2$) within a TF binding motif. Globally, we identified 131, 14, 122, 32, and 51 differentially expressed target genes downstream of enhancers with hypermethylated motifs in clusters 1, 2, 3, 4, and MDSp, respectively (Supporting Information S1: Table 12A). These genes were involved in hematopoiesis, metabolism regulation, and cancer development. Clusters 1 and 3's identity was defined by specific target gene sets not overlapping with gene sets of other clusters (Figure 4D). Deregulated gene expression linked to abnormal methylation of binding motifs for ETS and C/EBP TF families was observed in all clusters. As examples, in cluster 1, two hypermethylated bZIP motifs were associated with the downregulation of antiapoptotic sphingosine kinase 1 (*SPHK1*), and of ATB-binding cassette subfamily C member 3 (*ABCC3*) genes, and a hypomethylated ETS motif was associated with upregulation of chromodomain helicase DNA binding protein 7 (*CHD7*). Of note, in cluster 2, eight genes were upregulated downstream of enhancers with hypomethylated motifs, including the cyclin-dependent kinase inhibitor 1C (*CDKN1C*), the DNA-binding TF Lim homeobox protein 6 (*LHX6*), a GATA target, and the sprouty RTK signaling antagonist (*SPRY4*), an inhibitor of MAPK signaling pathway, while the downregulation of the protein phosphatase 1 regulatory subunit 27 (*PPP1R27*) was linked to a hypermethylated ETS motif (Figure 4E and Supporting Information S1: Table 12B). Moreover, in cluster 3, hypermethylated Zf and bHLH motifs were associated with the downregulation of interferon regulatory factor 2 (*IRF2*) and inhibitor of DNA binding 1 (*ID1*), and in cluster 4, and MDSp as well, hypermethylated Zf and bHLH motifs were associated with the downregulation of

TABLE 3 Characteristics of 175 LR-MDS patients according to methylation clusters. (continued on next page)

Variables	Cohort	Methylation cluster				*p* value
		C1	C2	C3	C4	
Number of patients	175	67	37	31	40	
Age, mean (range)	73.7 (20.9–94.3)	71.8 (20.9–92.5)	75.2 (44.5–93.1)	75.5 (55.5–94.3)	73.9 (30.5–90.2)	0.763
Sex (M/F)	87/88	24/43	17/20	20/11	26/14	0.008
Hemoglobin (g/dL), mean (range)	10.1 (6.2–14.9)	10.0 (6.2–14.2)	9.7 (6.6–14.1)	11.1 (7.1–14.9)	9.8 (6.7–14.2)	0.012
MCV (fL), median (range)	101 (79–129)	100 (79–124)	105 (89–129)	97 (81–127)	103 (85–134)	0.040
WBC ($\times 10^9$/L), mean (range)	5.2 (0.9–16.3)	5.1 (0.9–13.0)	5.6 (2.2–12.4)	4.1 (1.5–14.0)	6.0 (1.6–16.3)	0.003
ANC ($\times 10^9$/L), mean (range)	3.1 (0.1–11.2)	3.1 (0.1–9.7)	3.4 (0.4–8.9)	2.0 (0.5–9.8)	3.8 (0.4–11.2)	<0.001
AMC ($\times 10^9$/L), mean (range)	0.5 (0.1–2.6)	0.4 (0.1–1.2)	0.5 (0.1–0.6)	0.6 (0.1–2.6)	0.5 (0.1–1.6)	0.107
Monocytes (%), mean (range)	9.4 (1.0–35.8)	7.8 (2.0–21.6)	9.4 (1.0–15.0)	16.2 (2.6–70.3)	8.3 (2.0–17.8)	0.003
Platelets ($\times 10^9$/L), mean (range)	205 (10–714)	201 (19–714)	270 (46–532)	133 (38–295)	205 (10–580)	<0.001
Serum ferritin (ng/mL), mean (range)	437 (13–1969)	361 (13–1769)	616 (56–1969)	365 (32–1320)	523 (49–1476)	0.018
Transferrin sat (%), mean (range)	46.2 (5.4–95.0)	34.5 (5.4–80.0)	55.5 (23.0–95.0)	42.1 (21.0–81.0)	56.9 (31.0–92.0)	0.01
BM blasts (%), mean (range)	2.9 (0.0–9.0)	2.4 (0.0–9.0)	2.0 (0.0–9.0)	3.9 (0.0–9.0)	3.7 (0.0–9.0)	<0.001
RS (%), mean (range)	18.6 (0–87)	9.2 (0–68)	39.0 (0–80)	8.3 (0–70)	18.7 (0–87)	<0.001
WHO 2016, *n* (%)						<0.001
MDS-SLD	17 (9.7)	12 (17.9)	0 (0.0)	4 (12.9)	1 (2.5)	
MDS-MLD	58 (33.1)	30 (44.8)	4 (10.8)	9 29.0)	15 (37.5)	
MDS-RS-SLD	21 (12.0)	6 (9.0)	9 (24.3)	1 (3.2)	5 (12.5)	
MDS-RS-MLD	32 (18.3)	6 (9.0)	20 (54.1)	2 (6.4)	4 (10.0)	
MDS del(5q)	5 (2.9)	3 (4.5)	1 (2.7)	0 (0.0)	1 (2.5)	
MDS-EB1	40 (22.9)	9 (13.4)	3 (8.1)	15 (48.4)	13 (32.5)	
MDS unclassified	2 (1.1)	1 (1.5)	0 (0.0)	0 (0.0)	1 (2.5)	
IPSS-R risk group, *n* (%)						<0.001*
Very low	34 (19.4)	17 (25.4)	6 (16.2)	7 (22.6)	4 (10.0)	
Low	95 (54.3)	42 (62.7)	29 (78.4)	13 (41.9)	11 (27.5)	
Int	30 (17.1)	5 (7.5)	1 (2.7)	8 (25.8)	16 (40.0)	
High	7 (4.0)	1 (1.5)	0 (0.0)	2 (6.4)	4 (10.0)	
Missing	9 (5.1)	2 (3.0)	1 (2.7)	1 (3.2)	5 (12.5)	
Number of mutations, median (range)	2 (0–6)	0 (0–6)	2 (0–4)	3 (1–6)	3 (1–6)	<0.001
IPSS-M, *n* (%)						<0.0001
Very low/low/moderate low	129 (73.7	59 (88.1)	33 (89.2)	23 (74.2)	14 (35.0)	

TABLE 3 (Continued)

Variables	Cohort	Methylation cluster				*p* value
		C1	C2	C3	C4	
Very high/high/moderate high	37 (21.1)	6 (8.9)	3 (8.1)	7 (22.6)	21 (52.5)	
Missing	9 (5.2)	2 (3.0)	1 (2.7)	1 (3.2)	5 (12.5)	
AML transformation, *n* (%)	28 (16.0)	3 (4.5)	5 (13.5)	4 (12.9)	16 (40.0)	<0.001

Abbreviations: AMC, absolute monocyte count; AML, acute myeloid leukemia; ANC, absolute neutrophil count; BM, bone marrow; EB1, excess of blasts 5%–9%; IPSS-R, International Prognostic Scoring System-Revised; IPSS-M, IPSS-molecular; MCV, median corpuscular volume; MLD, multiple lineage dysplasia; RS, ring sideroblasts; SLD, single lineage dysplasia; WBC, white blood cell.

isocitrate dehydrogenase 1 (*IDH1*) and p21(RAC1)-activated kinase 2 (*PAK2*) (Figure 4E and Supporting Information S1: Table 12A).

We also annotated differentially methylated promoters with the nearest differentially expressed gene (Supporting Information S1: Table 9A–E). The number of genes overlapping promoters with an aberrant methylation level was higher in cluster 4 than in the others. Specific to cluster 4, we identified a set of 10 differentially expressed genes, all involved in cancer development, including 5 upregulated genes with a hypomethylated promoter, such as cyclin Y (*CCNY*), chondroitin sulfate synthase 1 (*CHSY1*), guanine nucleotide exchange factor (*MCF2L*), cohesin complex component (*RAD21*), and RAP1 GTPase activating protein (*RAP1GAP2*), and *5* downregulated genes with a hypermethylated promoter, such APC downregulated 1 like (*APCDD1L*), cannabinoid receptor interacting protein 1 (*CNRIP1*), NFKB activating protein like (*NKAPL*), tensin 3 (*TNS3*), and zinc finger protein 518B (*ZNF518B*) (Figure 4F).

Altogether, these results indicate that methylation imprinting at distal promoter regulatory regions drives gene expression changes in LR-MDS and that the methylation-deregulated proximal promoters may confer an additional level of regulation to gene expression among the LR-MDS which rapidly evolve to AML.

Clinical features and outcomes of methylation-based clusters within LR-MDS

Methylation-based clusters showed distinct clinical features (Table 3). Males were more represented than females in clusters 3 and 4. Clusters 2 and 4 showed lower hemoglobin and blast percentage and higher MCV, ferritin, and transferrin saturation. Cluster 3 showed higher hemoglobin, monocyte, and blast percentages and lower platelet count compared to other clusters. Cluster 1 was enriched in MDS-del(5q) and MDS-SLD/MLD, while clusters 2 and 3 were enriched in MDS-RS and MDS-EB1, respectively. Cluster 4 gathered miscellaneous WHO categories with the highest proportion of MDSp.

We analyzed the clinical outcomes of each cluster, including OS and risk of disease progression to AML. The four methylation-defined clusters showed different OS ($p < 0.0001$) and different times to AML progression ($p < 0.0001$) (Figure 5A,B). Compared to clusters 1, 2, and 3, cluster 4 showed the worst OS with HR 5.69 (97.5% CI [2.94–11.00], $p < 0.0001$), HR 4.50 (97.5% CI [2.18–9.28], $p < 0.0001$) and HR 3.02 (97.5% CI [1.51–6.06], $p = 0.002$), respectively, and the highest risk of AML evolution with HR 14.12 (97.5% CI [4.03–49.45], $p < 0.0001$), HR 6.32 (97.5% CI [2.07–19.32], $p = 0.001$), and HR 4.76 (97.5% CI [1.57–14.40], $p = 0.006$), respectively (Supporting Information S1: Table 13). We then investigated whether cluster 4 might contain information that could refine prognostication in the lower-risk IPSS-M group (IPSS-M ≤ 0). Methylation cluster 4 retained a significantly inferior OS compared to clusters 1, 2, and 3 with HR 5.80 (97.5% CI [2.48–13.55], $p = 0.0001$), HR 5.66 (97.5% CI [2.10–15.21], $p = 0.0006$), and HR 3.97 (97.5% CI [1.43–11.06], $p = 0.008$), respectively. By contrast, the risk of AML progression was similar between clusters (Figure 5C,D).

DISCUSSION

In this study, unsupervised clustering approaches based on mutations or methylation profiling identified relevant clinical and biological clusters in a cohort of LR-MDS. Methylation-based clustering integrated specific genetic patterns and identified MDS patients with inferior OS and at risk of AML progression. Furthermore, it refined

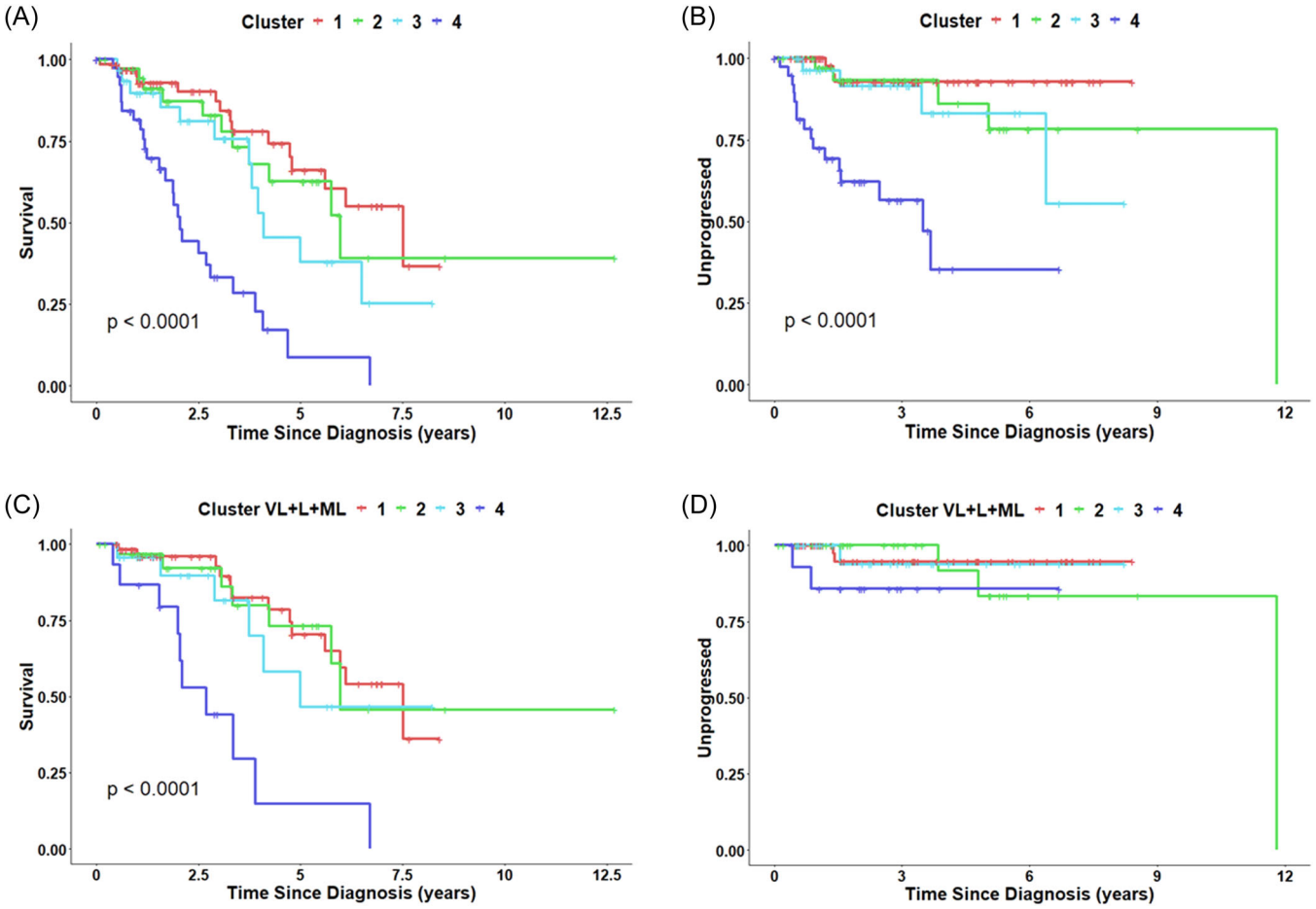


FIGURE 5 **Overall survival and risk of disease progression to acute myeloid leukemia (AML) stratified on methylation clustering.** **(A)** Overall survival (OS) of lower-risk myelodysplastic syndromes (LR-MDS) patients stratified on methylation clusters. The four clusters showed significantly different OS ($p < 0.0001$). **(B)** Risk of disease progression to AML of LR-MDS patients stratified on methylation clusters. The four clusters showed significantly different times to AML progression ($p < 0.0001$). **(C)** OS of MDS patients with very low (VL), low (L), moderate low (ML) IPSS-M stratified on methylation clusters (hazard ratio 3.10, 95% CI [1.71–5.63]; $p < 0.0001$). **(D)** Risk of AML progression of MDS patients with VL, L, ML IPSS-M stratified on methylation clusters.

the prediction of death from other causes than AML within lower-risk IPSS-M patients.

We identified six mutation-based clusters presenting similarities with those reported in unselected MDS.[40] Our strategy identified three *SF3B1*-mutated clusters with distinct clinical outcomes, *SF3B1/DNMT3A* showing worse OS and higher risk of progression than isolated *SF3B1* and *SF3B1/TET2*. In addition, it identified a cluster enriched in high-risk mutations associated with reduced OS, higher risk of progression, and shorter time to first treatment and a cluster characterized by *SRSF2* mutation mainly associated with *TET2* mutations, a co-mutation pattern highly predictive of chronic myelomonocytic leukemia (CMML) which is consistent with their early interception preceding the development of an overt CMML.[41] Finally, this analysis identified a group of LR-MDS without proof of clonality. Although we did not analyze all possible gene mutations and minor cytogenetic lesions might be missed by standard cytogenetics, this highlights that current morphology-based diagnostic criteria might not be sufficiently accurate to always discriminate myeloid neoplasms from other non-neoplastic conditions. In this cohort of LR-MDS patients, the genetic clustering identified subgroups of patients with shared molecular patterns suggesting distinct disease trajectories. Although its objective was different from outcome prediction, the genetic clustering informed the prognosis of patients with poor survival. On the contrary, the IPSS-M stratified better patients with more favorable outcomes.

Unsupervised analysis of genome-wide methylation profiles concentrated LR-MDS into four clusters that were enriched in specific genetic lesions such as del5q, *SF3B1*, *TET2/SRSF2*, and high-risk mutations. Most cases gathering in cluster 3 demonstrated frequent *TET2/SRSF2* or multi*TET2* mutations, a hypermethylated profile, a transcriptomic signature including M-CSF pathway activation and innate immune response and an intermediate prognosis, all features of MDS which may evolve to CMML.[42,43] Our analysis also identified a unique cluster of patients at risk of early AML evolution and inferior OS, characterized by a global hypermethylation predominating at promoters and frequent hypomethylated regions in shores and bivalent regulatory regions. Previous studies based on the methylome analysis of paired MDS/secondary AML samples have shown an increased methylation level at the advanced stage and distinctive baseline methylation patterns between stable MDS and MDS at risk of progression.[28,44] In agreement with our findings, hypomethylation extends across the genome at disease progression and is an important feature of de novo AML.[44,45] Using different computational methods, the interest for methylation-based clustering was strengthened by

the identification of groups of poor prognosis in an unselected population of LR-MDS and HR-MDS.[46,47] Selectively focusing on LR-MDS has the potential to increase the sensitivity in recognizing distinct disease trajectories and LR-MDS patients with poor outcome in the earliest stages.

The poor redundancy of methylation aberrancies at promoters or enhancers and transcriptomic profiles may reflect MDS heterogeneity. Seminal studies in human *TET2* or *IDH1/2* mutated MDS or AML samples reported events of promoter DNA hypermethylation,[25] when extended genomic coverage of CpGs showed that hypermethylation at enhancer regions appeared as a hallmark of *TET2* invalidation both in human CMML monocytes and in peripheral blood cells of clonal hematopoiesis of indeterminate potential and of clonal cytopenia of undetermined significance and in cell lines or mice models.[24,48,49] We confirmed here that enhancers were heavily hypermethylated, when *TET2* or *IDH1/2* genes were mutated (cluster 3) and also when *IDH1* gene was downregulated by hypermethylation of its enhancer TF binding motif (clusters 1 and 4). Other mutations in genes may indirectly affect the DNA methylation such as *SRSF2* and *ZBTB33*.[31,50] More surprisingly, the *SF3B1* mutation was strongly associated with a global hypomethylation of the genome even in the absence of a *DNMT3A* co-mutation. Altogether, the gene expression patterns driven by methylation aberrancies may capture the underlying biology and reveal the divergence of trajectories occurring during clonal progression.

The strong correlation between genetic and epigenetic classifications emphasizes the usefulness of multi-omics approach to LR-MDS patient risk stratification. The identification of a methylation subgroup of patients with inferior OS indicates that DNA methylation features may refine risk evaluation. Future studies circumscribed to MDS patients in lower-risk categories according to IPSS-M are required to strengthen this strategy and offer a frame for early therapeutic interventions.

ACKNOWLEDGMENTS

The authors acknowledge MDS-RIGHT consortium members, the physicians of the Groupe Francophone des Myélodysplasies: Pr. Bruno Quesnel, CHRU Lille, Pr. Mathilde Hunault-Berger, CHU Angers, Dr. Lise Willems, AP-HP.Centre Cochin, Dr. Aspasia Stamatoulas, Centre H Becquerel Rouen, Pr. Odile Beyne-Rauzy, CHU Toulouse, Dr. Carole Soussain, Institut Curie, Saint-Cloud; Dr. Nathalie Droin of the genomic platforms of Institut Gustave Roussy, and Dr. Franck Letourneur of Institut Cochin Genom'IC platform, and the engineers of the Genome Technology Center, Radboud University Medical center, and the Mutation Analysis Facility, Karolinska Institute.

AUTHOR CONTRIBUTIONS

David Rombaut, Tobias Tekath, and Sarah Sandmann performed bioinformatic analyses. Sarah Sandmann, Simon Crouch, Alexandra Smith, and Daniel Painter performed biostatistical analyses. Aniek O. de Graaf and Olivier Kosmider performed DNA-sequencing and analyzed the data. Magnus Tobiasson, Andreas Lennartsson, and Bert A. van der Reijden analyzed the data. Pierre Fenaux, Lionel Adès, Argiris Symeonidis, Jaroslav Čermák, Claude Preudhomme, Aleksandar Savic, Ulrich Germing, Reinhard Stauder, and David Bowen provided human samples and clinical annotations. Reinhard Stauder analyzed the HQoLdata. Corine van Marrewijk provided organizational support. Theo de Witte lead the consortium. Elsa Bernard critically reviewed the data. Eva Hellström-Lindberg provided platform for methylation study and critically reviewed the manuscript. Simon Crouch and Maud D'Aveni supervised the statistical analyses. Joost Martens analyzed the data and critically reviewed the manuscript. Joop H. Jansen provided platform for DNA-sequencing and analyzed the data. Luca Malcovati, Joop H. Jansen, and Michaela Fontenay supervised the study and wrote the manuscript.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Proposals for access to EUMDS de-identified data can be addressed to L.M. Reference genome and annotations are available on Gencode: https://www.gencodegenes.org/human/release_19.html. Data are available on NCBI's Gene Expression Omnibus (GEO) for methylation array data and RNA-seq data: DNA methylation profiling of 175 lower-risk MDS patients: GSE256394: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE256394. Transcriptome profiling of 185 BM MNC RNA lower-risk MDS patients: GSE220518: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE220518.

FUNDING

This work was part of the MDS-RIGHT project, which has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N°634789, "Providing the right care to the right patient with myelodysplastic syndrome at the right time." Disclaimer: This work reflects only the authors' view and that the European Commission is not responsible for any use that may be made of the information it contains. This work was supported by the Programme Hospitalier de Recherche Clinique MDS-04 funded by INCa-DGOS under grant agreement N°5480, the Site de Recherche Intégrée sur le Cancer (SIRIC) CAncer Research for PErsonalized Medicine (CARPEM), and the Fondation pour la Recherche Médicale (équipe labellisée FRM202003010191). L.M. was supported by the Associazione Italiana per la Ricerca sul Cancro (AIRC), Milan, Italy (Investigator Grant #20125; AIRC 5×1000 project #21267); Cancer Research UK, FC AECC, and AIRC under the International Accelerator Award Program (project #C355/A26819 and #22796).

ORCID

David Rombaut http://orcid.org/0000-0001-8910-0945
Michaela Fontenay https://orcid.org/0000-0002-5492-6349

SUPPORTING INFORMATION

Additional supporting information can be found in the online version of this article.

REFERENCES

1. de Swart L, Smith A, Johnston TW, et al. Validation of the revised international prognostic scoring system (IPSS-R) in patients with lower-risk myelodysplastic syndromes: a report from the prospective European LeukaemiaNet MDS (EUMDS) registry. *Br J Haematol*. 2015;170(3):372-383. doi:10.1111/bjh.13450
2. Malcovati L, Hellström-Lindberg E, Bowen D, et al. Diagnosis and treatment of primary myelodysplastic syndromes in adults: recommendations from the European LeukemiaNet. *Blood*. 2013;122(17):2943-2964. doi:10.1182/blood-2013-03-492884
3. de Witte T, Bowen D, Robin M, et al. Allogeneic hematopoietic stem cell transplantation for MDS and CMML: recommendations from an international expert panel. *Blood*. 2017;129(13):1753-1762. doi:10.1182/blood-2016-06-724500

4. Stojkov K, Silzle T, Stussi G, et al. Guideline-based indicators for adult patients with myelodysplastic syndromes. *Blood Adv*. 2020; 4(16):4029-4044. doi:10.1182/bloodadvances.2020002314
5. Itzykson R, Crouch S, Travaglino E, et al. Early platelet count kinetics has prognostic value in lower-risk myelodysplastic syndromes. *Blood Adv*. 2018;2(16):2079-2089. doi:10.1182/bloodadvances.2018020495
6. Hoeks M, Bagguley T, van Marrewijk C, et al. Toxic iron species in lower-risk myelodysplastic syndrome patients: course of disease and effects on outcome. *Leukemia*. 2021;35(6):1745-1750. doi:10.1038/s41375-020-01022-2
7. De Witte T, Malcovati L, Fenaux P, et al. Novel dynamic outcome indicators and clinical endpoints in myelodysplastic syndrome; the European LeukemiaNet MDS Registry and MDS-RIGHT project perspective. *Haematologica*. 2020;105(11):2516-2523. doi:10.3324/haematol.2020.266817
8. Stauder R, Yu G, Koinig KA, et al. Health-related quality of life in lower-risk MDS patients compared with age- and sex-matched reference populations: a European LeukemiaNet study. *Leukemia*. 2018;32(6):1380-1392. doi:10.1038/s41375-018-0089-x
9. Garelius HKG, Johnston WT, Smith AG, et al. Erythropoiesis-stimulating agents significantly delay the onset of a regular transfusion need in nontransfused patients with lower-risk myelodysplastic syndrome. *J Intern Med*. 2017;281(3):284-299. doi:10.1111/joim.12579
10. Delhommeau F, Dupont S, Valle VD, et al. Mutation in TET2 in myeloid cancers. *N Engl J Med*. 2009;360(22):2289-2301. doi:10.1056/NEJMoa0810069
11. Langemeijer SMC, Kuiper RP, Berends M, et al. Acquired mutations in TET2 are common in myelodysplastic syndromes. *Nat Genet*. 2009;41(7):838-842. doi:10.1038/ng.391
12. Papaemmanuil E, Cazzola M, Boultwood J, et al. Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *N Engl J Med*. 2011;365(15):1384-1395. doi:10.1056/NEJMoa1103283
13. Yoshida K, Sanada M, Shiraishi Y, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature*. 2011;478(7367):64-69. doi:10.1038/nature10496
14. Papaemmanuil E, Gerstung M, Malcovati L, et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood*. 2013;122(22):3616-3627. doi:10.1182/blood-2013-08-518886
15. Haferlach T, Nagata Y, Grossmann V, et al. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia*. 2014;28(2):241-247. doi:10.1038/leu.2013.336
16. Gallì A, Todisco G, Catamo E, et al. Relationship between clone metrics and clinical outcome in clonal cytopenia. *Blood*. 2021;138(11):965-976. doi:10.1182/blood.2021011323
17. Malcovati L, Karimi M, Papaemmanuil E, et al. SF3B1 mutation identifies a distinct subset of myelodysplastic syndrome with ring sideroblasts. *Blood*. 2015;126(2):233-241. doi:10.1182/blood-2015-03-633537
18. Arber DA, Orazi A, Hasserjian RP, et al. International Consensus Classification of Myeloid Neoplasms and Acute Leukemias: integrating morphologic, clinical, and genomic data. *Blood*. 2022;140(11):1200-1228. doi:10.1182/blood.2022015850
19. Khoury JD, Solary E, Abla O, et al. The 5th edition of the World Health Organization classification of haematolymphoid tumours: myeloid and histiocytic/dendritic neoplasms. *Leukemia*. 2022;36(7):1703-1719. doi:10.1038/s41375-022-01613-1
20. Bernard E, Nannya Y, Hasserjian RP, et al. Implications of TP53 allelic state for genome stability, clinical presentation and outcomes in myelodysplastic syndromes. *Nature Med*. 2020;26(10):1549-1556. doi:10.1038/s41591-020-1008-z
21. Duncavage EJ, Bagg A, Hasserjian RP, et al. Genomic profiling for clinical decision making in myeloid neoplasms and acute leukemia. *Blood*. 2022;140(21):2228-2247. doi:10.1182/blood.2022015853
22. Gelsi-Boyer V, Trouplin V, Adélaïde J, et al. Mutations of polycomb-associated gene ASXL1 in myelodysplastic syndromes and chronic myelomonocytic leukaemia. *Br J Haematol*. 2009;145(6):788-800. doi:10.1111/j.1365-2141.2009.07697.x
23. Ernst T, Chase AJ, Score J, et al. Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. *Nat Genet*. 2010;42(8):722-726. doi:10.1038/ng.621
24. Rasmussen KD, Jia G, Johansen JV, et al. Loss of TET2 in hematopoietic cells leads to DNA hypermethylation of active enhancers and induction of leukemogenesis. *Genes Dev*. 2015;29(9):910-922. doi:10.1101/gad.260174.115
25. Jiang Y, Dunbar A, Gondek LP, et al. Aberrant DNA methylation is a dominant mechanism in MDS progression to AML. *Blood*. 2009;113(6):1315-1325. doi:10.1182/blood-2008-06-163246
26. Figueroa ME, Skrabanek L, Li Y, et al. MDS and secondary AML display unique patterns and abundance of aberrant DNA methylation. *Blood*. 2009;114(16):3448-3458. doi:10.1182/blood-2009-01-200519
27. Shen L, Kantarjian H, Guo Y, et al. DNA methylation predicts survival and response to therapy in patients with myelodysplastic syndromes. *J Clin Oncol*. 2010;28(4):605-613. doi:10.1200/JCO.2009.23.4781
28. Qin T, Sotzen J, Rampal RK, et al. Risk of disease progression in low-risk MDS is linked to distinct epigenetic subtypes. *Leukemia*. 2019;33(11):2753-2757. doi:10.1038/s41375-019-0518-5
29. Itzykson R, Kosmider O, Cluzeau T, et al. Groupe Francophone des Myelodysplasies (GFM). Impact of TET2 mutations on response rate to azacitidine in myelodysplastic syndromes and low blast count acute myeloid leukemias. *Leukemia*. 2011;25(7):1147-1152. doi:10.1038/leu.2011.71
30. Bejar R, Lord A, Stevenson K, et al. TET2 mutations predict response to hypomethylating agents in myelodysplastic syndrome patients. *Blood*. 2014;124(17):2705-2712. doi:10.1182/blood-2014-06-582809
31. Beauchamp EM, Leventhal M, Bernard E, et al. ZBTB33 is mutated in clonal hematopoiesis and myelodysplastic syndromes and impacts RNA splicing. *Blood Cancer Discovery*. 2021;2(5):500-517. doi:10.1158/2643-3230.BCD-20-0224
32. Bernard E, Tuechler H, Greenberg PL, et al. Molecular international prognostic scoring system for myelodysplastic syndromes. *NEJM Evid*. 2022;1(7):EVIDo a2200008. doi:10.1056/EVIDoa2200008
33. Sandmann S, Karimi M, de Graaf AO, et al. appreci8: a pipeline for precise variant calling integrating 8 tools. *Bioinformatics*. 2018;34(24):4205-4212. doi:10.1093/bioinformatics/bty518
34. Bizet M, Defrance M, Calonne E, et al. Improving Infinium MethylationEPIC data processing: re-annotation of enhancers and long noncoding RNA genes and benchmarking of normalization methods. *Epigenetics*. 2022;17(13):2434-2454. doi:10.1080/15592294.2022.2135201
35. Newman AM, Steen CB, Liu CL, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol*. 2019;37(7):773-782. doi:10.1038/s41587-019-0114-2
36. Pellin D, Loperfido M, Baricordi C, et al. A comprehensive single cell transcriptional landscape of human hematopoietic progenitors. *Nat Commun*. 2019;10(1):2395. doi:10.1038/s41467-019-10291-0
37. Kelley LA, Gardner SP, Sutcliffe MJ. An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally-related subfamilies. *Protein Eng Des Select*. 1996;9(11):1063-1065. doi:10.1093/protein/9.11.1063
38. Todisco G, Creignou M, Bernard E, et al. Integrated genomic and transcriptomic analysis improves disease classification and risk stratification of MDS with ring sideroblasts. *Clin Cancer Res*. 2023;29(20):4256-4267. doi:10.1158/1078-0432.CCR-23-0538
39. Assi SA, Imperato MR, Coleman DJL, et al. Subtype-specific regulatory network rewiring in acute myeloid leukemia. *Nat Genet*. 2019;51(1):151-162. doi:10.1038/s41588-018-0270-1

40. Bersanelli M, Travaglino E, Meggendorfer M, et al. Classification and Personalized Prognostic assessment on the basis of clinical and genomic features in myelodysplastic syndromes. *J Clin Oncol*. 2021;39(11):1223-1233. doi:10.1200/JCO.20.01659
41. Todisco G, Creignou M, Gallì A, et al. Co-mutation pattern, clonal hierarchy, and clone size concur to determine disease phenotype of SRSF2P95-mutated neoplasms. *Leukemia*. 2021;35(8):2371-2381. doi:10.1038/s41375-020-01106-z
42. Droin N, Jacquel A, Hendra JB, et al. Alpha-defensins secreted by dysplastic granulocytes inhibit the differentiation of monocytes in chronic myelomonocytic leukemia. *Blood*. 2010;115(1):78-88. doi:10.1182/blood-2009-05-224352
43. Zhao LP, Boy M, Azoulay C, et al. Genomic landscape of MDS/CMML associated with systemic inflammatory and autoimmune disease. *Leukemia*. 2021;35(9):2720-2724. doi:10.1038/s41375-021-01152-1
44. Zhou J, Zhang T, Xu Z, et al. Genome-wide methylation sequencing identifies progression-related epigenetic drivers in myelodysplastic syndromes. *Cell Death Dis*. 2020;11(11):997. doi:10.1038/s41419-020-03213-2
45. Glass JL, Hassane D, Wouters BJ, et al., Epigenetic Identity in AML depends on disruption of nonpromoter regulatory elements and is affected by antagonistic effects of mutations in epigenetic modifiers. *Cancer Discovery*. 2017;7(8):868-883. doi:10.1158/2159-8290.CD-16-1032
46. Reilly B, Tanaka TN, Diep D, et al. DNA methylation identifies genetically and prognostically distinct subtypes of myelodysplastic syndromes. *Blood Adv*. 2019;3(19):2845-2858. doi:10.1182/bloodadvances.2019000192
47. Siamoglou S, Boers R, Koromina M, et al. Genome-wide analysis toward the epigenetic aetiology of myelodysplastic syndrome disease progression and pharmacoepigenomic basis of hypomethylating agents drug treatment response. *Hum Genomics*. 2023;17(1):37. doi:10.1186/s40246-023-00483-7
48. Merlevede J, Droin N, Qin T, et al. Mutation allele burden remains unchanged in chronic myelomonocytic leukaemia responding to hypomethylating agents. *Nat Commun*. 2016;7:10767. doi:10.1038/ncomms10767
49. Tulstrup M, Soerensen M, Hansen JW, et al. TET2 mutations are associated with hypermethylation at key regulatory enhancers in normal and malignant hematopoiesis. *Nat Commun*. 2021;12(1):6061. doi:10.1038/s41467-021-26093-2
50. Yoshimi A, Lin KT, Wiseman DH, et al. Coordinated alterations in RNA splicing and epigenetic regulation drive leukaemogenesis. *Nature*. 2019;574(7777):273-277. doi:10.1038/s41586-019-1618-0