



**HAL**  
open science

## Accelerated DNA replication fork speed due to loss of R-loops in myelodysplastic syndromes with SF3B1 mutation

David Rombaut, Carine Lefèvre, Tony Rached, Sabrina Bondu, Anne Letessier, Raphael Mangione, Batoul Farhat, Auriane Lesieur-Pasquier, Daisy Castillo-Guzman, Ismael Boussaid, et al.

### ► To cite this version:

David Rombaut, Carine Lefèvre, Tony Rached, Sabrina Bondu, Anne Letessier, et al.. Accelerated DNA replication fork speed due to loss of R-loops in myelodysplastic syndromes with SF3B1 mutation. Nature Communications, 2024, 15 (1), pp.3016. 10.1038/s41467-024-46547-7 . hal-04591955

**HAL Id: hal-04591955**

**<https://u-paris.hal.science/hal-04591955>**

Submitted on 4 Sep 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License































# Accelerated DNA replication fork speed due to loss of R-loops in myelodysplastic syndromes with *SF3B1* mutation

Received: 7 December 2022

Accepted: 29 February 2024

Published online: 08 April 2024


 Check for updates

David Rombaut <sup>1,2,3,4,18</sup>, Carine Lefèvre <sup>1,2,3,18</sup>, Tony Rached <sup>1,2</sup>, Sabrina Bondu<sup>1,2</sup>, Anne Letessier <sup>1</sup>, Raphael M. Mangione <sup>5</sup>, Batoul Farhat <sup>1,2</sup>, Auriane Lesieur-Pasquier<sup>1,2</sup>, Daisy Castillo-Guzman<sup>6</sup>, Ismael Boussaid <sup>1,2,4</sup>, Chloé Friedrich <sup>1,2,4</sup>, Aurore Tourville <sup>1,2</sup>, Magali De Carvalho<sup>1,2</sup>, Françoise Levavasseur <sup>1,2</sup>, Marjorie Leduc <sup>1,7</sup>, Morgane Le Gall <sup>1,7</sup>, Sarah Battault<sup>1,2</sup>, Marie Temple <sup>4</sup>, Alexandre Houy <sup>8</sup>, Didier Bouscary<sup>1,9</sup>, Lise Willems<sup>1,9</sup>, Sophie Park <sup>10</sup>, Sophie Raynaud<sup>11</sup>, Thomas Cluzeau <sup>12</sup>, Emmanuelle Clappier<sup>13</sup>, Pierre Fenaux<sup>14</sup>, Lionel Adès <sup>14</sup>, Raphael Margueron <sup>15</sup>, Michel Wassef<sup>15</sup>, Samar Alsafadi <sup>8</sup>, Nicolas Chapuis <sup>1,4</sup>, Olivier Kosmider <sup>1,2,4</sup>, Eric Solary <sup>16</sup>, Angelos Constantinou <sup>17,19</sup>, Marc-Henri Stern <sup>8</sup>, Nathalie Droin <sup>16</sup>, Benoit Palancade <sup>5</sup>, Benoit Miotto <sup>1</sup>, Frédéric Chédin <sup>6</sup> & Michaela Fontenay <sup>1,2,3,4</sup> 

Myelodysplastic syndromes (MDS) with mutated *SF3B1* gene present features including a favourable outcome distinct from MDS with mutations in other splicing factor genes *SRSF2* or *U2AF1*. Molecular bases of these divergences are poorly understood. Here we find that *SF3B1*-mutated MDS show reduced R-loop formation predominating in gene bodies associated with intron retention reduction, not found in *U2AF1*- or *SRSF2*-mutated MDS. Compared to erythroblasts from *SRSF2*- or *U2AF1*-mutated patients, *SF3B1*-mutated erythroblasts exhibit augmented DNA synthesis, accelerated replication forks, and single-stranded DNA exposure upon differentiation. Importantly, histone deacetylase inhibition using vorinostat restores R-loop formation, slows down DNA replication forks and improves *SF3B1*-mutated erythroblast differentiation. In conclusion, loss of R-loops with associated DNA replication stress represents a hallmark of *SF3B1*-mutated MDS ineffective erythropoiesis, which could be used as a therapeutic target.

Myelodysplastic syndromes (MDS) are hematopoietic stem cell neoplasms with heterogeneous outcome and limited therapeutic options in which somatic mutation in splicing factor (SF) genes is a cardinal feature<sup>1</sup>. MDS with splicing factor3b subunit 1 mutation (*SF3B1*<sup>MUT</sup>) are commonly associated with bone marrow (BM) ring sideroblasts and ineffective erythropoiesis<sup>2,3</sup>. The response rate to erythropoiesis-stimulating agents (ESA) is achieved in as many as 50-60%, of low-

risk MDS, but patients with *SF3B1*<sup>MUT</sup> MDS could be more often primary resistant with a trend to shorter median duration of response<sup>4,5</sup>. The transforming-growth factor beta family ligand trap, luspatercept has been initially approved to treat the anemia of transfusion-dependent patients with MDS with ring sideroblasts after ESA failure. Forty-five percent achieve transfusion independency with a 30-week median duration of response and patients could continue to benefit from long-

A full list of affiliations appears at the end of the paper.  e-mail: [michaela.fontenay@inserm.fr](mailto:michaela.fontenay@inserm.fr)

term treatment<sup>6–8</sup>. Deciphering the mechanisms of anemia is needed to generate treatments.

*SF3B1* mutation causes multiple alterations in mRNA processing. The use of alternative 3' or 5' splice site produces transcripts containing short intronic sequences that are degraded by the non-sense mediated decay (NMD) or are translated into a variant protein<sup>9–11</sup>. These splicing changes drive transcriptional reprogramming that shapes disease phenotype. For example, down-regulation of Fe-S cluster transporter *ABC7* by transcript isoform-specific degradation, and reduced translation efficiency of mitochondrial iron transporter *TMEM14C*, contribute to mitochondrial iron accumulation<sup>12,13</sup>. Overproduction of alternative and canonical transcripts of *ERFE* gene encoding hepcidin transcriptional repressor erythroferrone, leads to systemic iron overload<sup>14</sup>. *SF3B1* mutation also targets mitochondrial respiration and serine synthesis pathway<sup>14</sup>. While mutations in serine and arginine-rich splicing factor 2 (*SRSF2*) and U2 small nuclear RNA auxiliary factor 1 (*U2AF1*), predominantly alter cassette exon<sup>15,16</sup>, *SF3B1*<sup>MUT</sup> splicing pattern is dominated by intron retention reduction (IRR)<sup>17</sup>.

An increasing number of genomic alterations associated with MDS progression to acute myeloid leukemia (AML) suggests a genomic instability of stem and progenitor cells<sup>18</sup>. DNA damage and activation of ataxia telangiectasia and Rad3-related protein (ATR) pathway were detected in *SF*-mutated MDS<sup>19,20</sup>. More specifically, *SRSF2* and *U2AF1* mutations induce the formation of unscheduled RNA:DNA hybrids or R-loops, triggering DNA replication stress<sup>19–21</sup>. Mechanistically, mutant *SRSF2* impairs the RNA polymerase II transcription pause release, allowing nascent RNA forming a R-loop at promoter<sup>21</sup>. *SF3B1* has been involved in the pathways of DNA repair<sup>22,23</sup>. However, *SF3B1*<sup>MUT</sup> MDS patients have a lower risk of AML than other MDS<sup>24,25</sup> suggesting that *SF3B1*<sup>MUT</sup> MDS are less prone to genomic instability.

In the present study, we report that, on the contrary to *SRSF2*<sup>MUT</sup> or *U2AF1*<sup>MUT</sup> cells, *SF3B1*<sup>MUT</sup> erythroblasts demonstrate a significant loss of R-loops. These cells endure a DNA replication stress consisting in accelerated fork progression and single-stranded (ss)DNA exposure, and correlating with increased erythroid cell proliferation and impaired differentiation. The ability of low doses of histone deacetylase inhibitor (HDACi) vorinostat to restore R-loops without DNA damage, and to improve erythroid differentiation could serve as a therapeutic approach.

## Results

### Reduction of intron retention correlates with transcriptomic changes of *SF3B1*-mutated bone marrow mononuclear cells

A total of 143 subjects were enrolled in this study including 70 MDS with *SF3B1* mutation, 49 *SF3B1*<sup>WT</sup> MDS and 24 healthy controls (Table 1). To investigate the molecular pathways whose deregulation drives the phenotype of *SF3B1*<sup>MUT</sup>-MDS, BM mononuclear cells (MNC) RNA-sequencing data available from 21 *SF3B1*<sup>MUT</sup>-MDS and 6 *SF3B1*<sup>WT</sup>-MDS were re-analyzed<sup>11</sup>. With a mean number of 87 to 97 million reads per sample, DESeq2 analysis identified 1764 differentially expressed genes (DEGs) including 812 up- and 952 down-regulated genes ( $\log_2$  fold-change (FC) > |1|, Benjamini-Hochberg (BH)-adjusted *P* value < 0.05) (Fig. 1a; Supplementary Data 1). Gene ontology (GO) enrichment analysis showed that up- and down-regulated genes were involved in several pathways such as DNA replication, DNA repair, chromatid segregation, and cell cycle checkpoint signaling (Fig. 1b). These pathways were over-represented among up-regulated genes (Supplementary Fig. 1a). Eighty genes associated with these pathways allowed the clustering of *SF3B1*<sup>MUT</sup>-samples (Fig. 1c; Supplementary Data 1). Using KisSplice with a variation of percent splice in ( $\Delta$ PSI) > |0.10| and a BH-adjusted *P* value < 0.05, we detected 3937 differential splicing events (DSEs) in *SF3B1*<sup>MUT</sup> samples, including 1256 abnormal intron retention events, consisting in a majority of IRRs (*n* = 1027) in *SF3B1*<sup>MUT</sup>-samples (Fig. 1d). IRRs were the most frequent event in *SF3B1*<sup>MUT</sup>-MDS,

**Table 1 | Clinical and biological characteristics of patients according to *SF3B1* gene status**

Parameters	<i>SF3B1</i> <sup>WT</sup> <i>n</i> = 49	<i>SF3B1</i> <sup>MUT</sup> <i>n</i> = 70	<i>P</i> value
<b>Age</b>			
Median. years (range)	73 (44–94)	76 (46–91)	0.210
<b>Sex</b>			
Male no. (%)	29 (59.2)	38 (54.3)	0.708
<b>WHO-Subtype 2016 no. (%)</b>			
5q- syndrome	1 (2.0)	1 (0.1)	<0.0001
MDS-SLD	8 (16.3)	0 (0.0)	
MDS-MLD	23 (46.9)	0 (0.0)	
MDS-RS-SLD	0 (0.0)	25 (35.7)	
MDS-RS-MLD	1 (2.0)	40 (57.1)	
MDS-EB1	16 (32.6)	4 (5.7)	
<b>Hemogram, median (IQR 25–75)</b>			
Hemoglobin (g/dL)	9.7 (8.7–10.6)	8.8 (8.1–9.6)	0.001
Neutrophil count (G/L)	2.3 (1.4–3.9)	2.8 (1.8–4.0)	0.617
Platelet count (G/L)	136 (82–243)	258 (205–575)	<0.0001
<b>Bone marrow, median (IQR 25–75)</b>			
Blasts (%)	4 (2–6)	2 (1–3)	<0.0001
Erythroblasts (%)	24 (18–33)	40 (29–49)	<0.0001
Ring sideroblasts (%)	0 (0–4)	41 (25–54)	<0.0001
<b>Karyotype IPSS-R no. (%)</b>			
Very good	1 (2.0)	6 (8.6)	0.033
Good	29 (59.2)	51 (72.9)	
Intermediate	8 (16.3)	9 (12.9)	
Poor	1 (2.0)	0 (0.0)	
NA	10 (20.4)	4 (5.7)	
<b>IPSS-R no. (%)</b>			
Very low	7 (14.3)	12 (15.9)	0.0002
Low	13 (26.5)	44 (62.9)	
Intermediate	11 (22.4)	8 (11.4)	
High	6 (12.2)	2 (2.9)	
Very high	2 (4.1)	0 (0.0)	
NA	10 (20.4)	4 (5.7)	

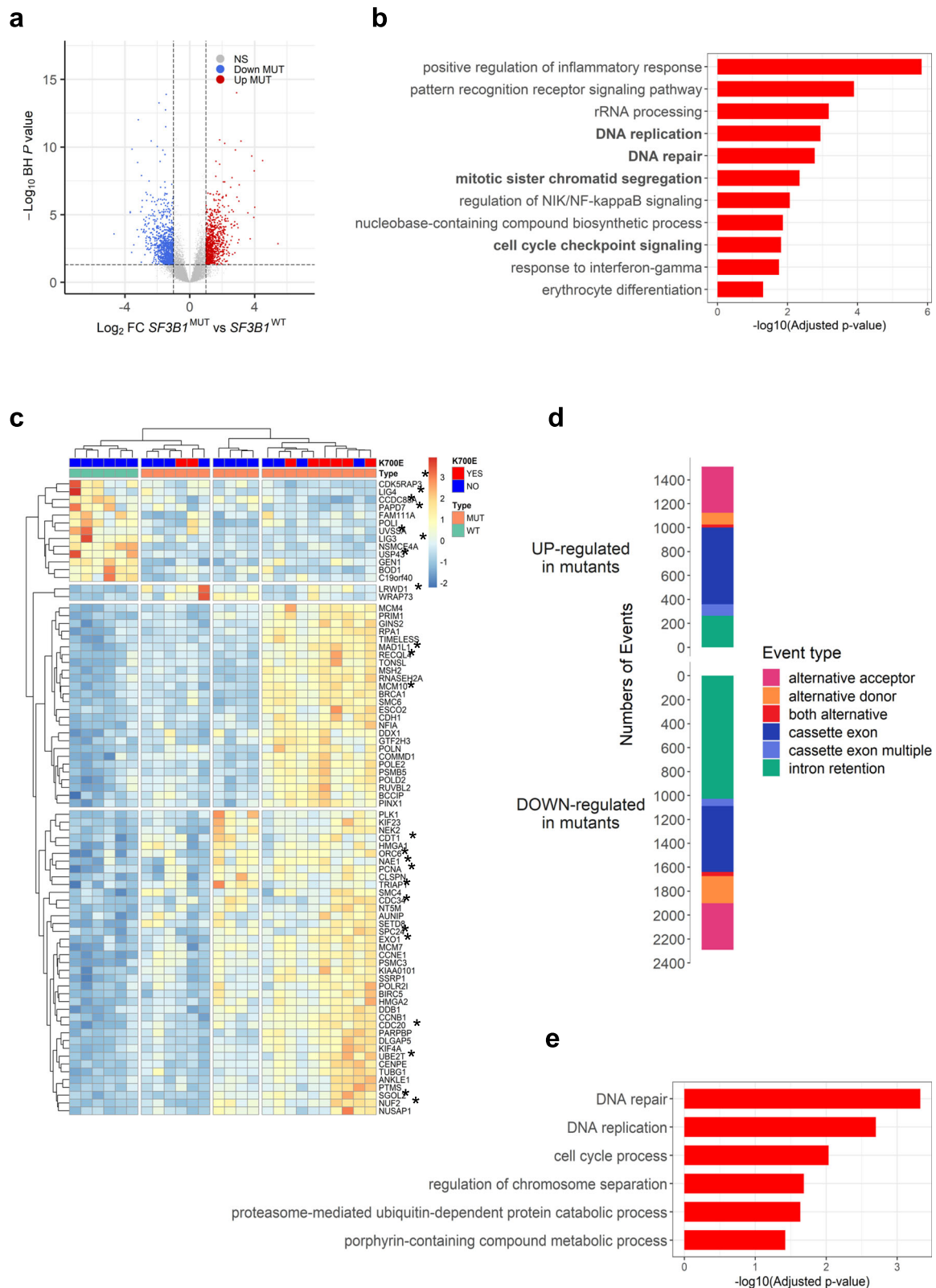
Unpaired t-test to compare quantitative values and Fisher's exact test to compare categorical variables.

SLD single lineage dysplasia, MLD multilineage dysplasia, RS ring sideroblasts, EB1 excess of blasts type 1 (5 to 9% bone marrow blasts), IPSS-R International Scoring System-Revised, no number, IQR interquartile range, NA not available.

uncommon in the *SRSF2*<sup>MUT</sup> or *U2AF1*<sup>MUT</sup>-MDS of a large cohort of 189 lower-risk patients (Supplementary Fig. 1b). A GO analysis of the 822 genes affected by IRR revealed their over-representation in DNA replication, DNA repair, cell cycle regulation, and mRNA splicing (Supplementary Fig. 1c). Combined DEG and DSE analyses showed that 296 DEGs targeted by 384 IRR events referred to DNA repair, DNA replication, cell cycle process and regulation of chromosome separation (Supplementary Data 1; Fig. 1e). Thus, IRR might contribute to gene expression changes of *SF3B1*<sup>MUT</sup>-BM MNC.

### *SF3B1*-mutated erythroid precursors demonstrate characteristic transcriptomic signatures

To decipher how transcriptional reprogramming driven by *SF3B1* mutation affected erythroid lineage, we expanded erythroblasts from CD34<sup>+</sup> hematopoietic stem and progenitor cells (HSPCs) collected from 14 *SF3B1*<sup>MUT</sup>, 10 without *SF* (*SF3B1*, *SRSF2*, *U2AF1*) mutation (*SF*<sup>WT</sup>) MDS and 6 healthy controls (Fig. 2a). May-Grünwald-Giemsa (MGG)-stained cytopspins showed the predominance of basophilic erythroblasts (basoE) at days 11–13 and polychromatophilic erythroblasts



(polyE) at days 14–15. Compared to healthy controls, the proportion of immature erythroblasts was significantly higher in MDS samples (Fig. 2b). Despite a significant increase of apoptotic cells, the rate of proliferation of *SF3B1*<sup>MUT</sup> erythroblasts was similar to control and *SF3B1*<sup>WT</sup> erythroblasts (Supplementary Fig. 2). In *SF3B1*<sup>MUT</sup>-cultures, the range of *SF3B1* variant allele frequency was 23–54% at d11–13 and 17–51% at d14–15 (Fig. 2c).

RNA-sequencing provided an equivalent mean number of million reads per sample in basoE (116 ± 18) and polyE (97 ± 19) (t-test;  $P = 0.543$ ). With a  $\log_2(\text{FC}) > |1|$  (BH-adj  $P$  value < 0.05), a total of 675 DEGs (514 up and 161 down) separated mutant and wild-type basoE. The number of DEGs was 765 (390 up, 375 down) in polyE with an overlap of 179 genes with basoE (Fig. 2d, Fig. 2e; Supplementary Data 2). GO enrichment analysis showed that up- and down-regulated

**Fig. 1 | Intron retention reduction correlates with transcriptomic changes of human *SF3B1*<sup>MUT</sup> bone marrow mononuclear cells.** RNA-sequencing data of 21 *SF3B1*<sup>MUT</sup> and 6 *SF3B1*<sup>WT</sup> (4 *SRSF2*<sup>MUT</sup> and 2 *SF*<sup>WT</sup>) bone marrow mononuclear cell samples were re-analyzed. **a** Volcano plot showing 1764 up or down-regulated genes in *SF3B1*<sup>MUT</sup> samples ( $\text{Log}_2(\text{FC}) < |1|$ ; Two-sided Wald test and Benjamini-Hochberg (BH)-adjusted *P* value < 0.05). **b** Gene Ontology (GO) enrichment analysis of the up and downregulated genes showing significantly enriched terms according to  $-\text{log}_{10}(\text{adjusted } P \text{ value})$ . Fisher's exact test corrected by false discovery rate (FDR) < 0.05. Terms of interest are in bold. **c** Heatmap representing the clustering of samples by the variations of expression of a subset of 80 genes belonging to GO

terms highlighted in **(b)**. Genes affected by 1 to 8 differential splicing events are marked with an asterisk. **d** Barplot representing the number and types of differential splicing events in *SF3B1*<sup>MUT</sup> in comparison to *SF3B1*<sup>WT</sup> with  $\Delta\text{PSI} > |0.10|$  using two-sided Wald test and BH-adjusted *P* value < 0.05. The bars over 0 indicate the events upregulated in mutant cases and the bars under 0 indicate the events downregulated in wild type cases. **e** GO over-representation analysis of 296 significantly deregulated genes affected by 383 intron retention reductions. Fisher's exact test corrected by FDR < 0.05. FC: fold-change. Source data are provided as a Source Data file.

genes in *SF3B1*<sup>MUT</sup>-basoE or -polyE were involved in several pathways such as DNA repair, regulation of MAP kinase cascade, ubiquitin-dependent protein catabolism, cellular response to DNA damage and oxygen-containing compound (Fig. 2f; Supplementary Fig. 3a). In basoE, a GeneSet Enrichment Analysis (GSEA) refined these results showing the deregulation of specific DNA repair pathways such as base excision repair, and a trend for nucleotide excision repair, but neither homologous recombination nor non-homologous end-joining or mismatch repair. Genes involved in DNA replication and G1/S phase checkpoint were also significantly deregulated (Fig. 2g; Supplementary Fig. 3b, c). The DSE profiles identified in basoE and polyE were similar to that of BM MNC (Fig. 2h). IRR represented 194/829 (23.4%) DSEs in *SF3B1*<sup>MUT</sup>-basoE and 383/1182 (32.4%) DSEs in *SF3B1*<sup>MUT</sup>-polyE, respectively, showing that IRR frequency increased with cell differentiation (Fig. 2i; Supplementary Data 2). Genes affected by IRR regardless of the stage of erythroid differentiation were involved in the response to DNA damage, mitotic cell cycle regulation, nucleocytoplasmic transport and in the positive regulation of histone deacetylation (Fig. 2j). As the nuclear retention or cytoplasmic degradation of IR-containing transcripts participate in the differentiation and specialization of normal erythroid cells<sup>26–28</sup>, we reasoned that IRR-transcripts detected in *SF3B1*<sup>MUT</sup>-erythroblasts may reshape the transcriptome contributing to defective maturation.

### ***SF3B1*-mutated erythroid precursors demonstrate characteristic proteomic signatures**

We concomitantly performed a proteomic analysis of proE, basoE and polyE with label-free quantification (LFQ). Principal component analysis showed that proE and basoE clustered together and separately from polyE. In each group, the second dimension discriminated *SF3B1*<sup>MUT</sup> from *SF3B1*<sup>WT</sup>-samples (Supplementary Fig. 4a). In subsequent analyses, proE and basoE were grouped. The mean number of identified proteins with LFQ values per sample was 4231 ( $\pm 263$ ), without significant difference between *SF3B1*<sup>MUT</sup> and *SF3B1*<sup>WT</sup>-samples, or between proE/basoE and polyE samples of each genetic group. A total of 443 and 290 differentially expressed proteins were detected between *SF3B1*<sup>MUT</sup> and *SF3B1*<sup>WT</sup>-samples in proE/basoE and in polyE, respectively (*t*-test,  $\text{Log}_2(\text{LFQ intensity}) > |0.20|$ , *P* value < 0.05) (Fig. 2k; Supplementary Fig. 4b, c).

In *SF3B1*<sup>MUT</sup>-proE/basoE proteome, mitochondrial ABCB7 (ATP-binding cassette B member 7), SUCLA2 (succinyl coA ligase), NNT (NAD(P) transhydrogenase), PPOX (protoporphyrinogen oxidase) were decreased, while SOD2 (superoxide dismutase) and ATP5A1 (ATP synthetase subunit alpha) were increased, further indicating mitochondrial dysfunction. Several key components of DNA replication pathway, such as DNA ligases LIG1 and LIG3 were decreased (Supplementary Fig. 4d; Supplementary Data 3). In *SF3B1*<sup>MUT</sup>-polyE, FEN1 (5'FLAP-endonuclease and gap endonuclease) and POLE (polymerase  $\epsilon$ ) expression was decreased while XRCC3 (X-ray repair cross-complementing 3), and ORC1 (Origin recognition complex subunit 1) expression was increased (Supplementary Data 3). Ingenuity Pathway Analysis showed that changes of *SF3B1*<sup>MUT</sup>-proE/basoE and polyE proteomes overlapped with changes of their transcriptomes (Fig. 2f, g) such as base excision repair, nucleotide excision repair,

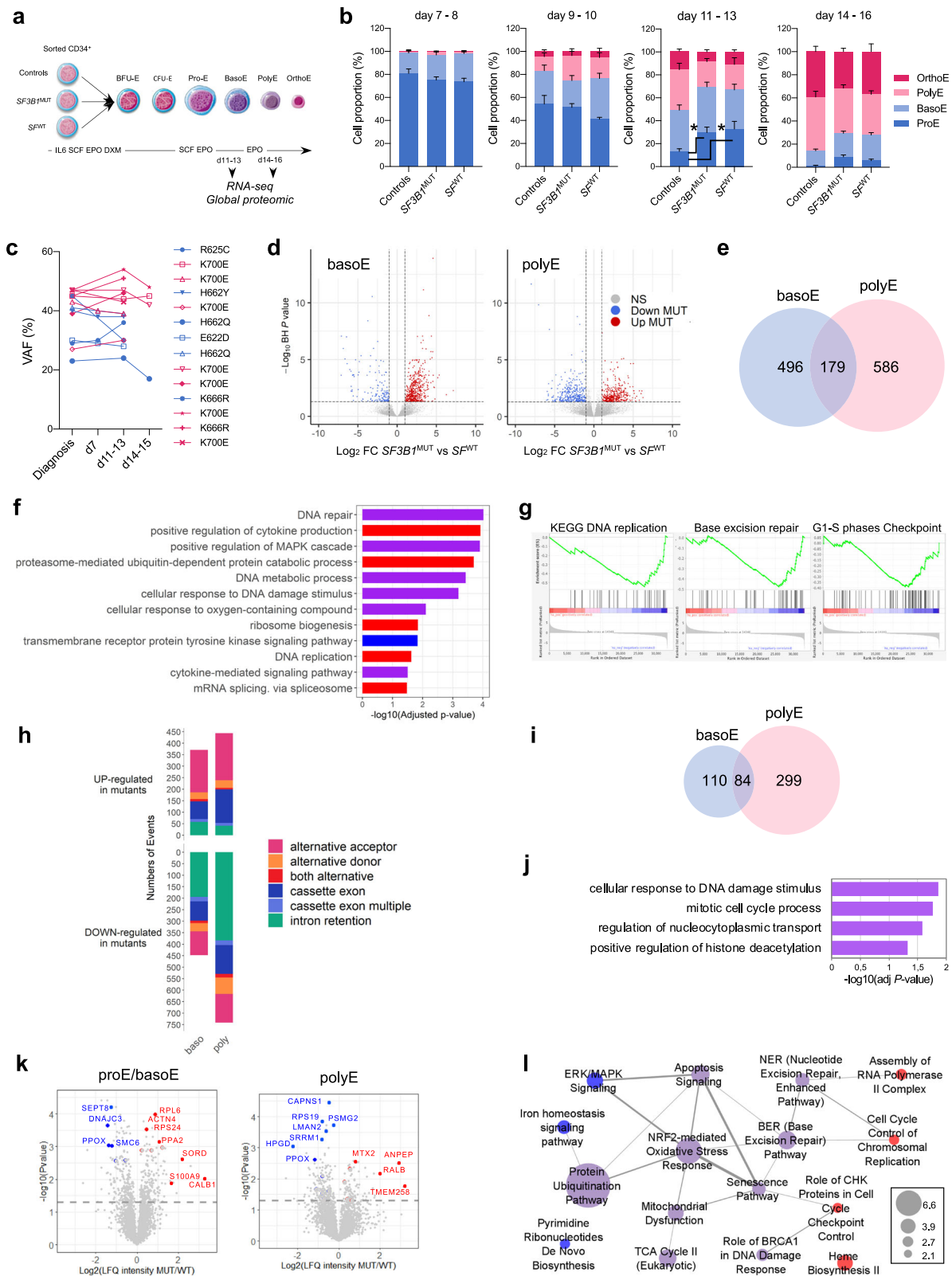
oxidative stress response, protein ubiquitination pathways. MAP kinase pathway, pyrimidine de novo biosynthesis and iron metabolism were specifically deregulated in proE/basoE, while heme biosynthesis, cell cycle checkpoint control and cell cycle control of replication were deregulated in polyE (Fig. 2i). Together, proteomic changes, notably those affecting components of DNA damage response kept the imprint of IRR-transcripts associated with *SF3B1* mutation in erythroid precursors (Fig. 2j).

### ***SF3B1* mutation induces a significant loss of R-loops in erythroid cells**

RNA splicing may attenuate the probability of forming R-loops by reducing the homology between nascent RNAs and their DNA templates and/or by recruiting splicing factor that antagonize RNA:DNA hybrid formation<sup>29–33</sup>. We examined the profiles of R-loops genome-wide in human primary basoE by performing DNA-RNA immunoprecipitation (DRIP) using S9.6 antibody followed by sequencing in 5 *SF3B1*<sup>MUT</sup>, 6 *SF3B1*<sup>WT</sup> including 3 with low variant allele frequency, 3 with high risk-mutations (1 *SRSF2* mutation, 1 *SRSF2* and bi-allelic *TET2* (*biTET2*) co-mutations, 1 *NRAS* and *biTET2* co-mutations) and 4 controls. DRIP specificity was assessed by a pre-treatment with RNase H1 (RNH1). Stringent calling identified true-positive peaks (BH-adj *P* value < 0.05) and we considered shared peaks between samples of each group. The number of shared peaks in *SF3B1*<sup>MUT</sup> erythroblasts was significantly lower than *SF3B1*<sup>WT</sup> or control cells (Fig. 3a). Visualization of R-loop profiles of a 50 kb region on chromosome 7 demonstrated the overall reduction of the peaks in *SF3B1*<sup>MUT</sup> sample compared to *SF3B1*<sup>WT</sup> and controls (Fig. 3b). We then compared the localization of the peaks to gene features between *SF3B1*<sup>MUT</sup> and *SF3B1*<sup>WT</sup> cells. In *SF3B1*<sup>MUT</sup>, the proportion of R-loops decreased at 5'UTR, promoter-transcription start site (TSS) and gene body (Fig. 3c). R-loops remaining in *SF3B1*<sup>MUT</sup> cells, were more frequently detected in intergenic regions and at 3'UTR of the genes (Fig. 3d). As an example, the *SUZ2* gene exhibited a reduced R-loop near its promoter in a *SF3B1*<sup>MUT</sup>-sample (Fig. 3e). By contrast, at mitochondrial DNA loci, rDNA repeats, or at some loci like *CPNE7* gene, peak intensity was found elevated in *SF3B1*<sup>MUT</sup>-samples showing that their R-loop profiles were selectively changed (Supplementary Fig. 5a).

R-loops assemble dynamically at TSS and TTS where they contribute to the regulation of transcription<sup>34–36</sup>. Thus, we compared the mRNA level of genes overlapping shared R-loops at TSS, gene body, TTS and 3'UTR in *SF3B1*<sup>WT</sup> and *SF3B1*<sup>MUT</sup>-samples. The expression of genes in which R-loops were detected at TTS or 3'UTR in *SF3B1*<sup>WT</sup> samples was similar in *SF3B1*<sup>WT</sup> and *SF3B1*<sup>MUT</sup>-cells while the expression of genes in which R-loops were detected at gene body in *SF3B1*<sup>WT</sup>-samples increased in *SF3B1*<sup>MUT</sup>-samples. By contrast, the expression of genes in which R-loops were detected at TSS in *SF3B1*<sup>WT</sup> but not in *SF3B1*<sup>MUT</sup> samples dramatically decreased in *SF3B1*<sup>MUT</sup>-samples (Fig. 3f).

To look for differential peaks between the groups, the count of restriction fragments overlapping with significant peaks was normalized using DESeq2. We confirmed a global decrease in the number of R-loops in *SF3B1*<sup>MUT</sup> versus controls or *SF3B1*<sup>WT</sup> basoE. In these analyses, the number of differential peaks (with  $\text{log}_2(\text{FC}) > |1|$  and BH-adj *P* value < 0.05), was 4589 between *SF3B1*<sup>MUT</sup> and controls with only 52 up



in *SF3B1*<sup>MUT</sup> samples, and 19,394 between *SF3B1*<sup>MUT</sup> and *SF3B1*<sup>WT</sup> samples, of which only 33 were up-regulated in *SF3B1*<sup>MUT</sup> cells (Fig. 3g; Supplementary Data 4). In comparison to controls, 53% and 20% of the peaks lost in *SF3B1*<sup>MUT</sup> cells were located in gene bodies (mostly centered on intronic sequences), and promoter-TSS, respectively (Supplementary Fig. 5b). Similarly, in comparison to *SF3B1*<sup>WT</sup> cells, 63% and 12% of the peaks lost in *SF3B1*<sup>MUT</sup> cells were located in gene bodies and

promoter-TSS (Fig. 3h). In this latter comparison, we identified 7,039 unique genes affected by losses of R-loops either in their UTRs, TSS, TTS or gene bodies, which overlapped 345/498 (69%) of the recurrent IRRs shared by *SF3B1*<sup>MUT</sup> samples (Fig. 3i). For example, at *RAD9A* and *IQGAP3* loci (Fig. 3j) or at *COASY* locus (Supplementary Fig. 5c), all occupied by an IRR, R-loops were significantly reduced in *SF3B1*<sup>MUT</sup> cells. To validate these findings, we selected two genes (*ABCC5* and

**Fig. 2 | Transcriptomic and proteomic features of *SF3B1*<sup>MUT</sup> bone marrow-derived basophilic (basoE) and polychromatophilic (polyE) erythroblasts.** Erythroid precursors were expanded in culture from *SF3B1*<sup>MUT</sup>, *SF*<sup>WT</sup> MDS and controls samples. **a** Schematic representation of the protocol. **b** Erythroid differentiation evaluated on May-Grünwald Giemsa-stained cytospins. Histograms representing the proportion of erythroid precursors in up to 7 controls, 11 *SF3B1*<sup>MUT</sup> and 7 *SF*<sup>WT</sup> independent samples at days (d)7-8, 9-10, 11-13 and 14-16. Results are expressed as means ± standard error of the mean. 2-way ANOVA for multiple comparisons. Controls versus *SF3B1*<sup>MUT</sup>,  $P = 0.017$ ; controls versus *SF*<sup>WT</sup>,  $P = 0.012$ . **c** Variant allele frequencies of *SF3B1* mutation in erythroblasts at d7, d11-13 and d14-15 of 14 independent *SF3B1*<sup>MUT</sup> samples. **d** Volcano plot representing up- and down-regulated transcripts in *SF3B1*<sup>MUT</sup> basoE and polyE compared to *SF*<sup>WT</sup> ones. Two-sided Wald-test and BH-adjusted  $P$  value < 0.05. **e** Venn diagram representing the numbers of differentially expressed genes between *SF3B1*<sup>MUT</sup> and *SF*<sup>WT</sup> samples at basoE and polyE stages. **f** Gene Ontology (GO) enrichment analysis of up- and down-regulated genes in *SF3B1*<sup>MUT</sup> versus *SF*<sup>WT</sup> erythroblasts. Fisher's exact test

corrected by false discovery rate (FDR) < 0.05. Specific terms to basoE or polyE as blue or red bars, respectively, shared terms as violet bars. **g** Gene set Enrichment Analysis (GSEA) showing terms deregulated in *SF3B1*<sup>MUT</sup> basoE. **h** Barplots representing numbers and types of differential splicing events in *SF3B1*<sup>MUT</sup> versus *SF3B1*<sup>WT</sup> basoE and polyE with  $\Delta\text{PSI} > |0.10|$  using two-sided Wald-test and BH-adjusted  $P$  value < 0.05. Bars over 0 indicate events upregulated and bars under 0 indicate events downregulated in *SF*<sup>WT</sup> erythroblasts. **i** Venn diagram of intron retention reductions (IRR) in *SF3B1*<sup>MUT</sup> basoE and polyE. **j** GO terms overrepresented among genes with IRR in *SF3B1*<sup>MUT</sup> basoE and polyE. Fisher's exact-test corrected by FDR < 0.05. **k** Volcano plots representing differentially expressed proteins in *SF3B1*<sup>MUT</sup> versus *SF*<sup>WT</sup> samples at proE/basoE and polyE stages (Wald-test, BH-adjusted  $P$  value < 0.05). **l** Cytoscape representation of Ingenuity Pathway Analysis showing deregulated pathways in *SF3B1*<sup>MUT</sup> versus *SF3B1*<sup>WT</sup> samples ( $P$  values < 0.05 by Student t-test) either basoE-specific (blue dots), polyE-specific (red dots) or shared (violet dots). Scale: dot size proportional to  $-\log_{10}$  (adjusted- $P$  value). Source data are provided as a Source Data file.

*TCIRG1*) with IRR and two genes (*IREB2* and *TMX2*) without IRR in *SF3B1*<sup>MUT</sup>-erythroblasts. We performed DRIP-qPCR in basoE generated from 4 *SF3B1*<sup>MUT</sup> MDS, 3 *SF3B1*<sup>WT</sup> MDS including one MDS with *U2AF1* mutation, and 4 healthy donors. Positive controls for R-loops detection were *RPL13A* and *TFPT* genes and for each sample, the specificity of the signal was assessed by RNHI pre-treatment (Supplementary Fig. 6a, b). R-loops detected at *ABCC5* and *TCIRG1* loci in healthy donor and *SF3B1*<sup>WT</sup> cells were not or hardly detected in *SF3B1*<sup>MUT</sup>-samples. By contrast, at *IREB2* and *TMX2* loci, the enrichment signal was faint, suggesting the absence of R-loop, whatever the sample (Fig. 3k).

Altogether, these results validate a link between *SF3B1* mutation, decreased R-loop formation, intron retention reduction and deregulated gene expression.

### SF3B1 mutation promotes a DNA replication stress in human erythroblasts

Since the accumulation of R-loops is reported to slow down the DNA replication fork velocity and cell proliferation<sup>37</sup>, loss of R-loops in *SF3B1*<sup>MUT</sup> erythroblasts may reduce obstacles to fork progression, promote DNA replication and cell proliferation. To explore this, we expanded erythroblasts from CD34<sup>+</sup> HSPCs of 5 *SF3B1*<sup>MUT</sup>, 3 *SRSF2*<sup>MUT</sup> or *U2AF1*<sup>MUT</sup>, 3 *SF*<sup>WT</sup>-MDS and 4 healthy donors. Flow cytometry analysis showed significant increase in the BrdU intensity of S-phase cells (Fig. 4a, b) and percentage of S-phase cells (Fig. 4c) in *SF3B1*<sup>MUT</sup>-basoE.

To further monitor DNA replication, we performed DNA combing in primary erythroblasts. We labelled basoE from 4 *SF3B1*<sup>MUT</sup>, 2 *SRSF2*<sup>MUT</sup> and 1 *U2AF1*<sup>MUT</sup> (*SF*<sup>MUT</sup>), 2 *SF*<sup>WT</sup> MDS and 2 healthy control samples with 5-iodo-2'-deoxyuridine (IdU) and then with 5-chloro-2'-deoxyuridine (CldU) for 30 min each (Fig. 4d). Incorporation of thymidine analogs allowed measurement of DNA fibers length and symmetry. We observed a significant increase of replication fork speed in *SF3B1*<sup>MUT</sup>-erythroblasts ( $0.75 \pm 0.35$  kb/min) compared to *SF*<sup>MUT</sup>- ( $0.60 \pm 0.25$  kb/min), *SF*<sup>WT</sup>- ( $0.53 \pm 0.21$  kb/min) and healthy donor ( $0.64 \pm 0.25$  kb/min) erythroblasts while fork symmetry of *SF3B1*<sup>MUT</sup>-cells measured by IdU/CldU ratio remained similar to that in controls. Compared to *SF3B1*<sup>MUT</sup>-cells, *SF*<sup>MUT</sup>-erythroblasts exhibited a slower and asymmetric fork progression (Fig. 4e, f). Altogether, an accelerated replication fork speed defines the DNA replication stress of *SF3B1*<sup>MUT</sup>-erythroblasts, whereas, it is associated with fork stalling in *SRSF2*<sup>MUT</sup> or *U2AF1*<sup>MUT</sup> cells<sup>19,21</sup>.

To address whether *SF3B1*<sup>MUT</sup>-erythroblasts endure a DNA damage, we investigated the expression by immunofluorescence of phospho(p)-RPA32 serine 33 that is recruited on single-stranded (ss) DNA during DNA replication and phosphorylated by ATR, and of p-RPA32 serine 4/8 that is phosphorylated by DNA-PK to regulate replication stress checkpoint activation<sup>38,39</sup>. We also used pan DNA-damage markers  $\gamma\text{H}_2\text{AX}$  or 53BP1 in human basoE from *SF3B1*<sup>MUT</sup>, *SF*<sup>MUT</sup> (*SRSF2*<sup>MUT</sup> or *U2AF1*<sup>MUT</sup>), *SF*<sup>WT</sup> MDS and controls, at d11 of the culture

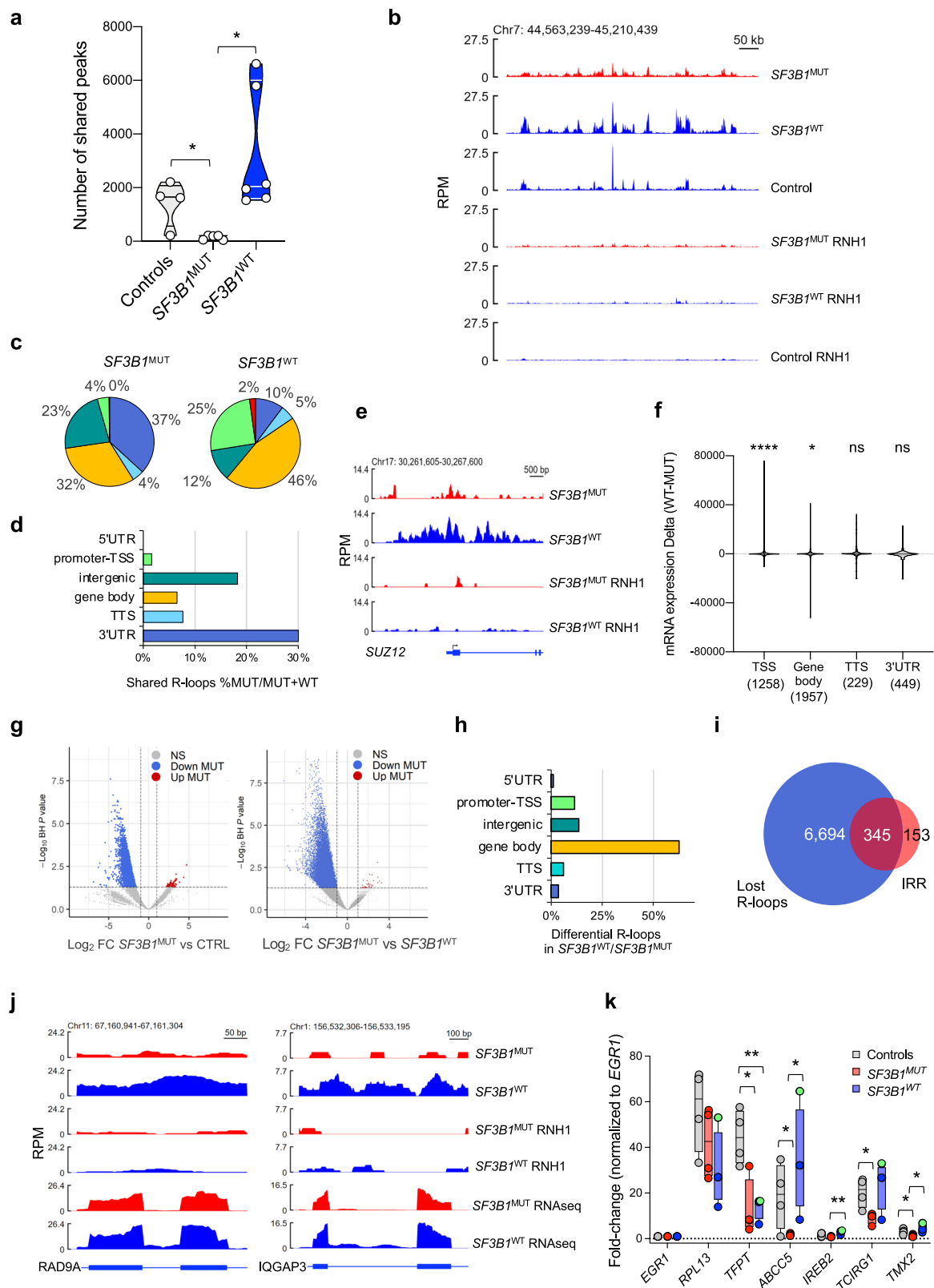
(Fig. 4g-k) or at different timepoints d9, d11, d13 and d15 (Supplementary Fig. 7a-c). p-RPA s33, p-RPA s4/8,  $\gamma\text{H}_2\text{AX}$  and 53BP1 foci were detected in control erythroblasts treated with hydroxyurea (HU), *SRSF2*<sup>MUT</sup> and *U2AF1*<sup>MUT</sup>-erythroblasts. By contrast,  $\gamma\text{H}_2\text{AX}$  and 53BP1 foci were undetectable in *SF3B1*<sup>MUT</sup> erythroblasts, while these cells were positive for p-RPA32s33 and p-RPAs4/s8 indicating ssDNA exposure. Of note, p-RPA32 s33 foci were significantly less abundant in *SF3B1*<sup>MUT</sup> than in *SRSF2*<sup>MUT</sup> cells.

Altogether, these results show that accelerated fork velocity was observed when R-loops were lost in *SF3B1*<sup>MUT</sup>-erythroblasts. Exposure of ssDNA without evidence for DNA damage marked by  $\gamma\text{H}_2\text{AX}$ /53BP1 indicates that *SF3B1*<sup>MUT</sup>-erythroblasts endured a milder replication stress than *SRSF2*<sup>MUT</sup> and *U2AF1*<sup>MUT</sup>-erythroblasts.

### *Sf3b1*<sup>K700E/+</sup> in murine erythroblasts reproduces DNA replication stress

To validate these findings in another model, we used the murine proerythroblastic cell lines G1E-ER4 CRISPR-Cas9 *Sf3b1*<sup>K700E/+</sup> and its isogenic G1E-ER4 *Sf3b1*<sup>+/+</sup>, which could differentiate into basoE upon induction with estradiol of GATA1, with no excess of apoptosis (Supplementary Fig. 8a-f)<sup>11</sup>. RNA sequencing of the murine *Sf3b1*<sup>K700E/+</sup> and *Sf3b1*<sup>+/+</sup> erythroblasts identified 1226 (719 up and 507 down) and 1434 (574 up and 860 down) DEGs before (t0) and after (t24) induction of GATA1, with  $\log_2$  (FC) > |1| and BH-adj  $P$  value < 0.05, respectively (Supplementary Data 5; Supplementary Fig. 9a, b). We detected 1116 and 1301 genes affected by DSEs, mainly IRR, in *Sf3b1*<sup>K700E/+</sup> proE and basoE, respectively (Supplementary Data 5; Fig. 5a). Despite the substantial species specificity of RNA splicing, the deregulated pathways associated with DSE in murine cells seemed similar to those identified in human cells (Fig. 5b, upper panel; Fig. 2f). Notably, DNA repair, cellular response to DNA damage and nucleic acid metabolic process GO terms gathered genes presenting IRR in *Sf3b1*<sup>K700E/+</sup>-proE ( $n = 272$ ) or -basoE ( $n = 355$ ) (Fig. 5b, bottom panel). Such deregulated pathways were conserved at protein level with deregulated expression of Lig1, Lig3, Pnkp, Parp1 in murine and human cells while DNA damage checkpoint proteins like Atm, Gmnn, Tp53bp1 were specifically deregulated in murine cells highlighting some differences between the two models (Fig. 5c).

To address the functional consequences of these dysregulations, we compared the proliferation of the G1E-ER4 clones. Before GATA1 induction, the proliferation of *Sf3b1*<sup>K700E/+</sup> proerythroblasts was significantly higher than that of *Sf3b1*<sup>+/+</sup> cells and remained higher upon induction ( $P < 0.0001$ , Fig. 5d). The differentiation to basoE was significantly lower in *Sf3b1*<sup>K700E/+</sup> cells (Fig. 5e, f). In accordance with this, *Sf3b1*<sup>K700E/+</sup> cells showed a higher BrdU incorporation and a higher G1/S fraction than *Sf3b1*<sup>+/+</sup> cells, before and after induction (Fig. 5g). Concomitantly to GATA1 induction, when a mild replication stress was imposed by inducing a cell cycle arrest in early S-phase with



aphidicolin, *Sf3b1*<sup>K700E/+</sup> compared to *Sf3b1*<sup>+/+</sup> cells exhibited a significantly higher G1/S phase fraction (Fig. 5h, right panel). Inhibition of dNTP biosynthesis by 0.2 mM hydroxyurea (HU) for 16 h, normalized BrdU incorporation in *Sf3b1*<sup>K700E/+</sup> cells after induction (Fig. 5i). In metabolomic analysis, the quantities of dATP, dCTP and dTTP before induction were equivalent in *Sf3b1*<sup>K700E/+</sup> to *Sf3b1*<sup>+/+</sup> cells, showing that DNA replication stress was not related to nucleotide pool depletion

(Fig. 5j). After induction of differentiation with estradiol, dNTP quantities increased in *Sf3b1*<sup>+/+</sup> cells, consistent with a slowdown of DNA synthesis. In *Sf3b1*<sup>K700E/+</sup> cells, lower dNTP quantities argued for persistent DNA synthesis (Fig. 5j; Supplementary Fig. 9c).

Upon induction of differentiation, more *Sf3b1*<sup>K700E/+</sup> cells were positive to p-Rpa32s4/s8 labelling with a higher sensitivity to HU treatment than *Sf3b1*<sup>+/+</sup>-cells (Fig. 5k). In estradiol-treated cells,



**Fig. 3 | Loss of R-loops overlaps intron retention reductions in *SF3B1*<sup>MUT</sup> human primary erythroblasts.** DRIP-sequencing. **a** Violin plots showing the shared peak numbers in 4 controls, 5 *SF3B1*<sup>MUT</sup> and 6 *SF3B1*<sup>WT</sup> (4 *SF*<sup>WT</sup>, 2 *SRSF2*<sup>MUT</sup>). Two-sided unpaired t-test. Controls versus *SF3B1*<sup>MUT</sup>,  $P = 0.011$ ; *SF3B1*<sup>WT</sup> versus *SF3B1*<sup>MUT</sup>,  $P = 0.015$ . **b** DRIP-seq profiles  $\pm$  RNaseH1 (RNH1) of a 50-kb region on chr7 showing the distribution of R-loops in reads per million. **c** Pie charts representing localizations of shared R-loops at gene features. **d** Proportion of shared peaks at gene features in *SF3B1*<sup>MUT</sup> samples relative to *SF3B1*<sup>MUT</sup> + *SF3B1*<sup>WT</sup> samples. **e** DRIP-seq profiles showing R-loops near *SUZ12* promoter. **f** Comparison of the expression of genes with overlapping R-loops at TSS, gene body, TTS and 3'UTR in *SF3B1*<sup>WT</sup> samples and without overlapping R-loops in *SF3B1*<sup>MUT</sup> samples. Violin plots represent the difference of mean expression intensity between *SF3B1*<sup>WT</sup> and *SF3B1*<sup>MUT</sup> samples (d11). Central lines represent the means. Gene numbers in each category are indicated. One sample Wilcoxon signed rank test is used for comparison of actual mean to theoretical mean. TSS,  $P < 0.0001$ ; gene body,  $P = 0.031$ ; TTS,  $P = 0.179$ ; 3'UTR,  $P = 0.118$ . **g** Volcano plot representing differential restriction

fragments overlapping with peaks between *SF3B1*<sup>MUT</sup> and control samples (left panel) and *SF3B1*<sup>MUT</sup> and *SF3B1*<sup>WT</sup> samples (right panel) with  $\log_2(\text{FC}) > |1|$  using two-sided Wald-test and a BH-adjusted  $P$  value  $< 0.05$ . **h** Distribution to gene features of differential R-loops in *SF3B1*<sup>WT</sup> samples and lost in *SF3B1*<sup>MUT</sup> samples. **i** Venn diagram showing overlap between genes that lost one R-loop and genes with intron retention reduction (IRR) in *SF3B1*<sup>MUT</sup> erythroblasts. **j** DRIP-seq and RNA-seq overlays at *RAD9A* and *IQGAP3* loci showing R-loop losses and IRR events in *SF3B1*<sup>MUT</sup> erythroblasts. Gene structures using GENCODE GRCh37. **k** DRIP-qPCR analysis of 4 controls, 3 *SF3B1*<sup>WT</sup> including 1 *U2AF1*<sup>MUT</sup> designated as green dot and 4 *SF3B1*<sup>MUT</sup> samples. Enrichment signals (normalized to input) at specific loci were normalized to *EGRI* (no R-loop). *RPL13A* and *TFPT* as positive controls. In box plots, central lines represent medians, bounds represent lower and upper quartiles and whiskers correspond to min-max values. Two-sided unpaired t-test for  $P$  values (see Suppl. information). **b, e, j** RPM: reads per million. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\*\* $P < 0.0001$ ; ns: not significant. Source data are provided as a Source Data file.

Western blot confirmed the engagement of Rpa, but did not show phosphorylation of Chk1 suggesting that the Atr-Chk1 pathway was not activated (Fig. 5l). Finally, the delayed differentiation of *Sf3b1*<sup>K700E/+</sup> murine erythroblasts can be partially rescued by lowering their high rate of DNA synthesis with HU at the expense of cell viability (Supplementary Fig. 9d, e).

### Targeting of R-loops improves the differentiation of human *SF3B1*<sup>MUT</sup> erythroblasts

To establish a link between the loss of R-loops and the phenotypic characteristics of fork velocity and replication stress in human primary *SF3B1*<sup>MUT</sup> erythroblasts, we thought to modulate the level of R-loops in the cell. Previous works have established that the THO complex which contributes to prevent R-loop accumulation interacts with SIN3A-histone deacetylase complex. Furthermore, inhibiting histone deacetylase activity by depleting SIN3A or treating the cells with trichostatin A stabilizes R-loops<sup>40</sup>. We used pan-HDAC inhibitor Suberoylanilide hydroxamic acid (SAHA)/vorinostat (further denoted HDACi) in DRIP-seq experiment. Human erythroblasts from 3 *SF3B1*<sup>MUT</sup>, 3 *SF3B1*<sup>WT</sup> (1 *SRSF2*, 1 *SRSF2*/bi*TET2* or 1 *NRAS*/bi*TET2* mutations), and 4 controls samples were pre-treated with HDACi 0.5  $\mu\text{M}$  for 20 h, at day11 of culture. The numbers of shared peaks increased in 3/4 controls, even not significantly. HDACi treatment restored the level of R-loops of 2/3 *SF3B1*<sup>MUT</sup> erythroblast samples up to normal (Fig. 6a; Supplementary Data 6). Unexpectedly, the numbers of R-loops counted in *SRSF2* or *TET2*/*NRAS* mutated samples collapsed almost entirely ( $P < 0.001$ ). In *SF3B1*<sup>MUT</sup> samples, the number of shared R-loops was quantitatively important in the gene bodies. The augmentation of R-loops also affected intergenic regions, 3'UTR and TTS more than 5'UTR or promoter-TSS (Fig. 6b). We visualized the changes in R-loop profiles at specific loci. As shown in Fig. 6c and Supplementary Fig. 10, HDACi treatment produced large R-loops near the promoter of *BCL2L1*, *PTPN11*, *ARPC3* and *NCOA4* genes specifically in *SF3B1*<sup>MUT</sup> cells. By contrast, HDACi did not change the profile of R-loops in *SF3B1*<sup>MUT</sup> cells at *HK1* locus (Supplementary Fig. 10). To verify whether, by rescuing R-loops, gene expression may change, we performed RT-qPCR at these 4 loci. Upon treatment with HDACi, the expression of *BCL2L1* increased significantly in *SF3B1*<sup>MUT</sup>. While the expression of *NCOA4* and *PTPN11* also tended to increase, *HK1* did not (Fig. 6d). These data suggest the relationship between R-loops and gene expression in these cells.

Then we wondered whether the effect of HDACi could be at least partly due to a modification of IRR profiles. To explore this hypothesis, we selected 5 genes with known IRR (*PPOX*, *PPM1A*, *COASY*, *S100A4*, *BCL2L1*) and performed RT-PCR to visualize the IRR-transcripts and the spliced isoforms in 4 *SF3B1*<sup>MUT</sup>, 6 *SF3B1*<sup>WT</sup> and 2 control erythroblast samples. HDACi did not modify the pattern of transcripts in controls and *SF3B1*<sup>WT</sup> samples. The abundance of IRR-transcripts in *SF3B1*<sup>MUT</sup> remained similar in the presence or absence of HDACi suggesting that

the restoration of R-loops observed under this treatment did not depend on intron retention (Fig. 6e). We confirmed these results for *BCL2L1* and *COASY* transcripts by fluorescent PCR fragment analysis (Supplementary Fig. 11a, b).

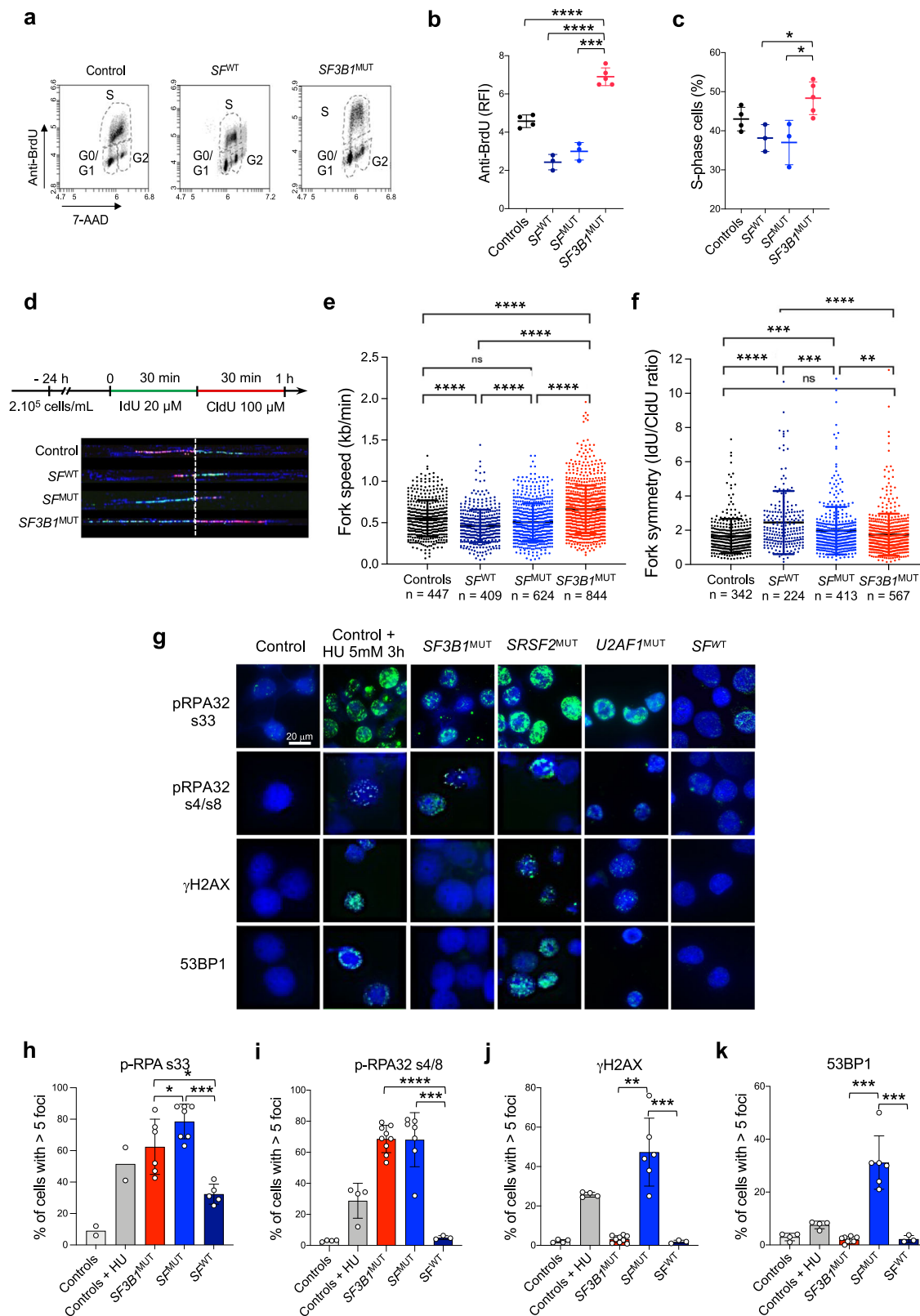
To evaluate the impact of R-loop restoration on fork progression, we performed DNA combing of 3 samples of erythroblasts treated or not with 0.2  $\mu\text{M}$  HDACi (1 control, 2 *SF3B1*<sup>MUT</sup> including one with a frameshift mutation in the histone acetyltransferase *EP300*). The fork velocity in *SF3B1*<sup>MUT</sup> erythroblasts was high and decreased significantly after HDACi treatment. By contrast, neither the *SF3B1*<sup>MUT</sup>/*EP300*<sup>MUT</sup> sample nor the control showed variations of fork velocity after HDACi (Fig. 7a). To assess DNA synthesis in a larger number of patients, we performed BrdU assays. The percentage of S-phase cells decreased significantly in *SF3B1*<sup>MUT</sup> cells. However, the BrdU intensity higher in *SF3B1*<sup>MUT</sup> samples did not change (Fig. 7b). Then, we verified the impact of HDACi (0.5  $\mu\text{M}$  20 h) on the frequency of p-RPA32 s33 and  $\gamma$ H2AX foci. No increase of positive cells was observed in control, *SF3B1*<sup>MUT</sup> or *SF3B1*<sup>WT</sup> erythroblast samples suggesting that HDACi at the concentration used did not provide DNA damage to these cells (Fig. 7c–e).

Finally, we studied the consequences of HDACi treatment on erythroid cell differentiation. At progenitor level, HDACi did not change the number or size of the BFU-E type colonies formed in methylcellulose (Fig. 7f). Looking at mutations in single colonies we did not observe any clonal selection during the 14 days of semi-solid culture (Supplementary Fig. 11c). By contrast, HDACi drastically forced the maturation of erythroblasts at late stage as shown by a significant increase of mature erythroblasts (Fig. 7g, h). This effect was also confirmed by the increase of GPA<sup>+</sup>CD49<sup>low</sup> cell proportion corresponding to orthochromatic erythroblasts (Fig. 7i). Interestingly we did not observe any inhibitory effect of HDACi at 0.5  $\mu\text{M}$  on the overall rate of expansion of *SF3B1*<sup>MUT</sup> erythroid precursors compared to *SF3B1*<sup>WT</sup> or control erythroblasts (Supplementary Fig. 11d), or increase of cell death in *SF3B1*<sup>MUT</sup>, *SF3B1*<sup>WT</sup>, and control cell cultures (Fig. 7j). Altogether, these results showed that HDACi by producing R-loops, slowed DNA replication and facilitated erythroid cell differentiation without altering cell proliferation.

## Discussion

The present study shows that *SF3B1* mutations causing an increased proliferation of immature erythroblasts with a reduced capacity to terminal differentiation trigger a replication stress with ssDNA exposure in erythroid cells. Accelerated DNA replication fork velocity is observed when R-loops are lost. HDAC inhibition restores R-loops and decreases fork speed, which correlates with erythroid differentiation improvement. These features distinguish *SF3B1* mutation from other splicing factor mutations.

Alternative splicing of pre-mRNA contributes to physiological hematopoiesis<sup>26–28,41</sup>. Intron retentions increase along terminal steps of erythroid differentiation<sup>28</sup>. Intron retention modulates gene



expression by generating transcripts that are either retained in the nucleus or degraded in the cytoplasm by the NMD<sup>41,42</sup>. Here, we show that *SF3B1*<sup>MUT</sup> promotes a reduction of intron retention in erythroblasts and that the number of retained introns lost in *SF3B1*<sup>MUT</sup>-erythroid precursors increases between basoE and polyE. By reversing a physiological process, *SF3B1* mutation changes gene expression profile and reshapes the proteome, affecting several pathways such as

DNA replication, DNA repair mainly base excision repair and nucleotide biosynthesis. *SF3B1*<sup>MUT</sup>-erythroblasts retain proliferative capacities, which contributes to enrichment of the bone marrow in immature erythroblasts and to defective production of mature erythroblasts, defining ineffective erythropoiesis.

To synthesize DNA properly, the replication machinery must overcome several obstacles, including R-loops and transcription

**Fig. 4 | DNA replication stress in *SF3B1*-mutated human primary erythroblasts.** Erythroblasts were derived in culture from 9 *SF3B1*<sup>MUT</sup>, 11 *SF3B1*<sup>WT</sup>, 6 controls. **a–c.** Cell cycle analysis by BrdU incorporation. **a** Representative flow cytometry scatter plot of control, *SF*<sup>WT</sup> and *SF3B1*<sup>MUT</sup> samples. **b, c** Scatter plots representing ratios of fluorescence intensity (RFI) anti-BrdU antibody/control Ig (**b**) and percentages of S-phase cells (**c**) in 4 controls, 3 *SF*<sup>WT</sup>, 3 *SF*<sup>MUT</sup> and 5 *SF3B1*<sup>MUT</sup> samples. Results are expressed as means ± SD. Two-sided unpaired t-test. **b** Controls vs *SF3B1*<sup>MUT</sup>,  $P = 0.003$ ; *SF*<sup>WT</sup> vs *SF3B1*<sup>MUT</sup>,  $P < 0.0001$ ; *SF*<sup>MUT</sup> vs *SF3B1*<sup>MUT</sup>,  $P < 0.0001$ . **c** *SF*<sup>WT</sup> vs *SF3B1*<sup>MUT</sup>,  $P = 0.012$ ; *SF*<sup>MUT</sup> vs *SF3B1*<sup>MUT</sup>,  $P = 0.017$ . **d–g** DNA combing performed in proE/basoE from 4 *SF3B1*<sup>MUT</sup>, 3 *SF*<sup>MUT</sup>, 2 *SF*<sup>WT</sup> MDS and 2 controls. **d** Upper panel: Schematic representation of IdU/CldU pulse labelling. Bottom panel: Representative microphotographs of DNA fibers. **e** Scatter plot showing fork speed (kb/min) expressed in means ± SD. Numbers of fibers analyzed are indicated. Two-sided Mann-Whitney test. Controls vs *SF3B1*<sup>MUT</sup>,  $P < 0.0001$ ; *SF*<sup>WT</sup> vs *SF3B1*<sup>MUT</sup>,  $P < 0.0001$ ;

*SF*<sup>MUT</sup> vs *SF3B1*<sup>MUT</sup>,  $P < 0.0001$ ; *SF*<sup>WT</sup> vs *SF*<sup>MUT</sup>,  $P = 0.0009$ ; *SF*<sup>WT</sup> vs controls,  $P < 0.0001$ ; *SF*<sup>MUT</sup> vs controls,  $P = 0.002$ . **f** Scatter plot showing fork symmetry as ratios of IdU/CldU length expressed in means ± SD. Two-sided Mann-Whitney test. Controls vs *SF3B1*<sup>MUT</sup>,  $P = 0.302$ ; *SF*<sup>WT</sup> vs *SF3B1*<sup>MUT</sup>,  $P < 0.0001$ ; *SF*<sup>MUT</sup> vs *SF3B1*<sup>MUT</sup>,  $P = 0.005$ ; *SF*<sup>WT</sup> vs *SF*<sup>MUT</sup>,  $P = 0.0003$ ; *SF*<sup>WT</sup> vs controls,  $P < 0.0001$ ; *SF*<sup>MUT</sup> vs controls,  $P = 0.0004$ . **g–k** Immunofluorescence experiments in MDS or control erythroblasts treated or not with 5 mM hydroxyurea (HU). **g** Representative images of phospho(p)-RPA32s33, p-RPA32s4/s8,  $\gamma$ H2AX and 53BP1 at d11. Nuclei were labeled with DAPI. Magnification 100X (scale: 20  $\mu$ m). **h–k** Quantification of positive cells with >5 intranuclear foci. **h** p-RPA32s33 (2 controls, 6 *SF3B1*<sup>MUT</sup>, 7 *SF*<sup>MUT</sup>, 5 *SF*<sup>WT</sup>). **i** p-RPA32s4/s8. **j**  $\gamma$ H2AX. **k** 53BP1 (4 controls, 9 *SF3B1*<sup>MUT</sup>, 7 *SF*<sup>MUT</sup>, 3 *SF*<sup>WT</sup>). Results are expressed as mean percentages of positive cells ± SD. Two-sided unpaired t-tests; \*  $P < 0.05$ ; \*\*  $P < 0.01$ ; \*\*\*  $P < 0.001$ ; \*\*\*\*  $P < 0.0001$ ; ns: not significant. Source data are provided as a Source Data file.

complexes<sup>37</sup>. Most of the R-loops located at promoter-TSS, in gene bodies and intergenic regions in *SF3B1*<sup>WT</sup>-erythroblasts were lost in *SF3B1*<sup>MUT</sup>-erythroblasts, in contrast with the augmented R-loops detected in *SRSF2*<sup>MUT</sup> erythroblasts. Previous studies in cell lines or in primary CD34<sup>+</sup> *SF*<sup>MUT</sup> progenitors using single-cell imaging with S9.6 antibody have shown that not only *SRSF2*<sup>MUT</sup> or *U2AF1*<sup>MUT</sup>, but also *SF3B1*<sup>MUT</sup> cells may produce undesirable R-loops<sup>19–21</sup>. However, S9.6 foci may indicate RNA:DNA hybrids and also double-stranded RNA which are more abundant, making difficult the quantification of R-loops even when RNaseH1 is overexpressed in cell lines to assess the specificity of the signals<sup>19,43,44</sup>. Here we used DRIP-seq to obtain a genome-wide landscape of R-loops. This technique allowed us to identify short type I R-loops located at GC-skewed promoter-proximal regions, or in intergenic regions that may contain active enhancers<sup>21</sup>, which are preferentially detected by R-ChIP, and also, large type II R-loops (spanning over 300 bp to 1 kb) distributed along the body of transcribed genes<sup>45,46</sup>. The decreased number of R-loops at promoters-TSS and gene bodies in *SF3B1*<sup>MUT</sup> cells suggest that both large and short R-loops are lost. As the GC skew characteristic of R-loops at promoter progressively diminishes after the first exon/intron junction, the mechanism of R-loop formation may be different in gene bodies<sup>45</sup>. Intron retention, by increasing homology between nascent RNA and its DNA template can initiate the co-transcriptional formation of large R-loops spreading over the gene coding sequence<sup>22,29,47,48</sup>. Conversely, the binding of splicing machinery making intron excision hinders R-loop accumulation<sup>30,33,49</sup>. Increased intron excision occurring when *SF3B1* is mutated may suppress R-loops along the gene bodies. We report here that most of IRR overlap with lost R-loops. However, because R-loops were lost also in promoter-TSS, intergenic regions and TTS, IRR is not the unique mechanism for R-loop loss in these cells.

The presence of unscheduled R-loops at promoter-TSS as a consequence of RNA polymerase II pausing usually correlates with high gene expression<sup>45</sup>. R-loop-positive regions overlap with DNase I-hypersensitive regions, indicative of open chromatin<sup>21,45</sup>. Our integrative analysis of R-loops and gene expression associated the loss of R-loops with low transcript level in *SF3B1*<sup>MUT</sup> erythroblasts. Importantly, we identified *BCL2L1*, a GATA1 and STAT5 target gene, which expression increased, without splicing changes, when R-loops formed near its promoter under HDACi treatment. This is consistent with the interference of R-loops with transcription. In addition, since SF3B1 protein together with U2AF1 and SRSF2 were detected in the vicinity of R-loops at promoters<sup>50</sup>, the SF3B1 mutant protein, could notably modify the kinetics of transcription elongation as generally reported for splicing factors<sup>51</sup>.

Replication stress that appears at the early stages of malignant cell transformation, has been linked to events that impair DNA synthesis integrity such as reduced or increased origin firing, nucleotides or replication factor depletion and replication-transcription conflicts<sup>52,53</sup>. Replication stress can produce stalled forks, under-replicated DNA, or supra-acceleration of forks<sup>54</sup>. As opposed to *SRSF2*<sup>MUT</sup> or *U2AF1*<sup>MUT</sup> cells in which fork progression is slowed down, we detected an accelerated

fork speed when R-loops were lost in *SF3B1*<sup>MUT</sup>-erythroblast. Alternative causes of increased fork velocity were associated with overexpression of oncogenes, downregulation of mRNA biogenesis, or PARP1 inhibition<sup>55–58</sup>. Moreover, activation of oncogenic HRAS in pre-senescent cells was shown to accelerate forks by inducing overexpression of topoisomerase 1 (TOP1), which is known to resolve unwanted R-loops<sup>59</sup>. TOP1 was heavily expressed in *SF3B1*<sup>MUT</sup> and *SF3B1*<sup>WT</sup> erythroblasts, while senataxin and THO complex proteins, other R-loop modulators were specifically upregulated in *SF3B1*<sup>MUT</sup> erythroblasts and could contribute to fork velocity<sup>60</sup>. Furthermore, if the transcription rate decreases when R-loops are resolved, not only R-loop loss but also the reduction of transcription-replication conflicts could prevent replication fork stalling<sup>61,62</sup>.

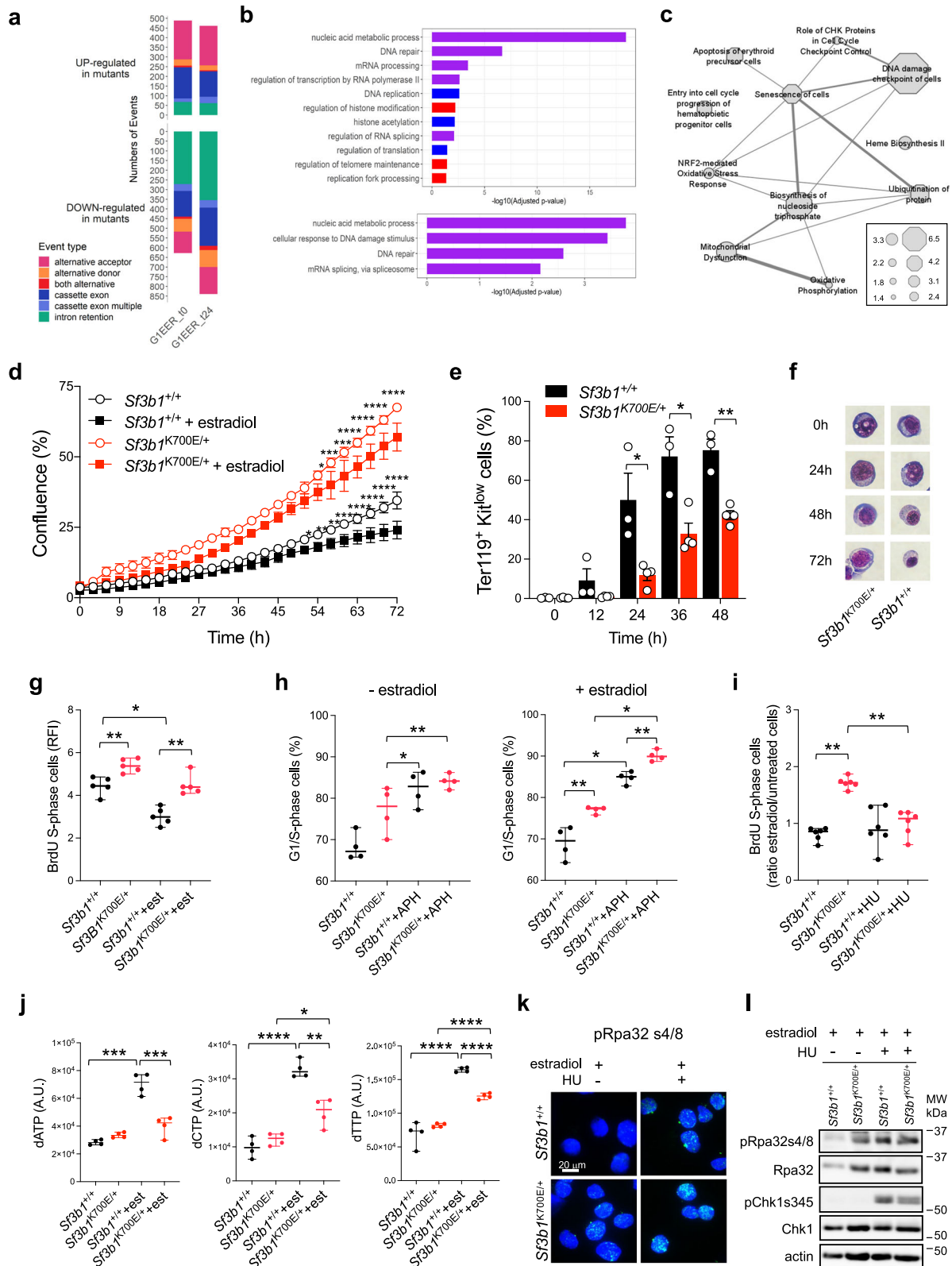
The detection of p-RPA32 foci without DNA damage marks  $\gamma$ H2AX or 53BP1 suggested replicative ssDNA gaps. ATR continuously monitors the recruitment of RPA32 on ssDNA within the replisomes and its phosphorylation on serine 33, independently of CHK1<sup>63</sup>. *SF3B1*<sup>MUT</sup>-erythroblasts endure a mild replication stress without engagement of the CHK1 pathway (Fig. 5I). In cancer cells, the tolerance to replication stress is supported by the overexpression of the upstream components of ATR-CHK1 pathway, Claspin and Timeless, independently of ATR signaling<sup>64</sup>. As *SF3B1*<sup>MUT</sup>-cells overexpressed Claspin and Timeless genes and Timeless protein, we cannot exclude their role in the mechanism of replication stress tolerance. Alternatively, RPA could bind the displaced ssDNA of R-loops and recruit RNaseH1 to facilitate its resolution<sup>65</sup>. Further studies are needed to test these hypotheses.

Screening for synthetic lethal approaches using *SF3B1*<sup>MUT</sup>-cell lines have revealed potential vulnerabilities by targeting of ATR, PARP1 or NMD<sup>19,65–67</sup>. An alternative, more conservative approach to improve ineffective erythropoiesis could be slightly restoring R-loops. The formation and degradation of R-loops involve multiple actors that include splicing factors, topoisomerases, DNA/RNA helicases, DNA repair molecules (BRCA1, BRCA2 or FANCD2/A/M), RNaseH1 or H2 which degrade R-loops directly, or SAMHD1 (alpha motif and HD-domain containing protein 1) that promotes ssDNA degradation at stalled forks<sup>68,69</sup>. Chromatin accessibility also interferes with R-loop generation<sup>62</sup>. In accordance, R-loops were shown to increase in HDACi-treated or SIN3A-depleted cells<sup>40</sup>. We show here that R-loops reappeared in *SF3B1*<sup>MUT</sup> cells treated with HDACi, which correlated with significant increase of *BCL2L1* expression and improvement of erythroid differentiation without evidence of DNA damage. Since clinical trials have shown a limited hematotoxicity of vorinostat alone compared to its combination with azacitidine<sup>70,71</sup>, our results provide a rationale for testing if low doses of vorinostat could improve erythropoiesis of patients with *SF3B1*<sup>MUT</sup> MDS.

## Methods

### Patients

A total of 143 subjects including 70 MDS patients with *SF3B1* mutation, 49 lower-risk (LR)-MDS patients without *SF3B1* mutation (17 *SRSF2*<sup>MUT</sup>,



7 *U2AF1*<sup>MUT</sup>, 25 triple-negative as *SF*<sup>WT</sup>) and 24 age-matched healthy subjects as controls was enrolled between 2015 and 2023. Bone marrow (BM) aspirates were collected after each patient gave his informed consent for biological investigations according to the recommendations of institutional review board (IRB) and ethics committee (IRB numbers: IdFV 212-A01395-38 EudraCT 2012-002990-7338; OncoCCH 2015-08-11-DC). BM aspirates or femoral head samples were

collected from healthy subjects. Patient characteristics including age, gender, WHO, hemogram, BM blast, erythroblast and ring sideroblast percentages, karyotype, IPSS-R are indicated in (Table 1).

**Genomic studies**

BM mononuclear cells (MNC) were purified on Ficoll gradient and were processed for DNA extraction using the DNA/RNA Kit (Qiagen, Hilden,

**Fig. 5 | DNA replication stress in murine G1E-ER4 Crispr-Cas9 *Sf3b1*<sup>K700E/+</sup> proerythroblasts.** **a** Barplots representing numbers and types of differential splicing events in *Sf3b1*<sup>K700E/+</sup> (clone 5.13) versus *Sf3b1*<sup>+/+</sup> (clone 9.82) cells at t0 (proE) and t24 (basoE) after induction of differentiation with estradiol ( $\Delta\text{PSI} > |0.10|$ ) using two-sided Wald test and BH-adjusted *P* value < 0.05). **b** Gene Ontology (GO) over-representation analysis. Upper panel: Pathways involving differentially spliced genes in *Sf3b1*<sup>K700E/+</sup> cells shared at t0 and t24 (violet bars), specific to t0 (blue bars) or t24 (red bars). Bottom panel: Pathways involving genes with IRR in *Sf3b1*<sup>K700E/+</sup> cells. Shared GO terms at t0 and t24. Fisher's exact test corrected by false discovery rate (FDR) < 0.05. **c** Ingenuity Pathway Analysis of differential proteins at t0 (Student *t*-test, *P* values < 0.05. Canonical Pathways (hexagons), Diseases and Functions pathways (circles). **d** Live cell imaging. Mean percentages ( $\pm$  SD) of confluence (*n* = 3). 2-way ANOVA test for multiple comparisons. **e** Differentiation of *Sf3b1*<sup>K700E/+</sup> (*n* = 4) and *Sf3b1*<sup>+/+</sup> cells (*n* = 3) by flow cytometry. Mean percentages of Ter119<sup>+</sup>Kit<sup>low</sup> cells  $\pm$  SEM. Unpaired *t*-test for multiple comparisons. t24h: *q* = 0.033; t36h: *q* = 0.027; t48h: *q* = 0.007. **f** May-Grünwald-Giemsa-stained cytopins. **g** BrdU

incorporation in S-phase  $\pm$  estradiol 24 h. Medians  $\pm$  95% confidence intervals (CI) of RFI anti-BrdU antibody/control Ig (5 independent experiments). *Sf3b1*<sup>+/+</sup> vs *Sf3b1*<sup>K700E/+</sup>, *P* = 0.002; *Sf3b1*<sup>+/+</sup>+estradiol vs *Sf3b1*<sup>K700E/+</sup>+estradiol, *P* = 0.001; *Sf3b1*<sup>+/+</sup> vs *Sf3b1*<sup>K700E/+</sup>+estradiol, *P* = 0.011. **h** Cell cycle analysis by BrdU incorporation  $\pm$  aphidicolin (APH). Median percentages ( $\pm$  95% CI) of G1/S-phase cells (4 independent experiments). Left: *Sf3b1*<sup>+/+</sup> vs *Sf3b1*<sup>K700E/+</sup>, *P* = 0.035; *Sf3b1*<sup>+/+</sup> vs *Sf3b1*<sup>+/+</sup>+APH, *P* = 0.002. Right: *Sf3b1*<sup>+/+</sup> vs *Sf3b1*<sup>K700E/+</sup>, *P* = 0.007; *Sf3b1*<sup>+/+</sup>+APH vs *Sf3b1*<sup>K700E/+</sup>+APH, *P* = 0.002; *Sf3b1*<sup>+/+</sup> vs *Sf3b1*<sup>+/+</sup>+APH, *P* = 0.029; *Sf3b1*<sup>K700E/+</sup> vs *Sf3b1*<sup>K700E/+</sup>+APH, *P* = 0.029. **i** BrdU incorporation in S-phase  $\pm$  hydroxyurea (HU). Medians  $\pm$  95% CI of RFI estradiol-treated/untreated cells (5 independent experiments). **j** dNTP relative quantities. Medians  $\pm$  95%CI (4 independent experiments). **g-j** Two-sided unpaired *t*-test. **k** Immunofluorescence of pRpa32s4/8 ( $\pm$  estradiol 24h, HU 16 h). **l** Western blot of pRpa32s4/s8, Rpa32, pChk1s345 and Chk1. Actin as loading control. **k, l** Representative of 3 independent experiments. \*\*\*\**P* < 0.0001, \*\*\**P* < 0.001, \*\**P* < 0.01, \**P* < 0.05; ns not significant. Source data are provided as a Source Data file.

Germany). Mutations in *SF3B1* were screened by Sanger sequencing or next generation sequencing of a panel of 37 genes (*ASXL1*, *ATM*, *BCOR*, *BCORL1*, *BRAF*, *CBL*, *CEBPA*, *CUX1*, *DDX41*, *DNMT3A*, *EP300*, *ETV6*, *EZH2*, *FLT3*, *GATA2*, *IDH1*, *IDH2*, *JAK2*, *KIT*, *KRAS*, *NRAS*, *MPL*, *NPM1*, *PHF6*, *PTPN11*, *RAD21*, *RIT1*, *RUNX1*, *SETBP1*, *SF3B1*, *SRSF2*, *STAG2*, *TET2*, *TP53*, *U2AF1*, *WT1*, *ZRSR2*).

### Primary cells and cell lines cultures

BM CD34<sup>+</sup>-derived erythroblasts were expanded from 92 MDS (49 *SF3B1*<sup>MUT</sup> and 43 *SF3B1*<sup>WT</sup> including 24 *SF*<sup>WT</sup>, 13 *SRSF2*<sup>MUT</sup> and 6 *U2AF1*<sup>MUT</sup>) and 24 healthy donors as controls. For erythroid cell expansion, CD34<sup>+</sup> cells were isolated from the MNC fraction of BM samples using the MidiMacs system (Miltenyi Biotec, Bergisch Gladbach, Germany). CD34<sup>+</sup> cells, which purity was higher than 80% were then cultured at  $0.8 \times 10^6$  per mL for 4 days in Iscove's modification of Dulbecco medium (IMDM; Thermo Fisher Scientific, Waltham, MA) containing 15% BIT9500 (Miltenyi Biotec), 100 U/mL penicillin, 100  $\mu$ g/mL streptomycin, 2 mM L-glutamine (all from Thermo Fisher Scientific), 2 U/mL recombinant human erythropoietin (rHu Epo; Roche, Basel, Switzerland), 100 ng/mL stem cell factor (SCF; Miltenyi Biotec), 10 ng/mL interleukin 6 (IL6; Miltenyi Biotec) and  $2.10^{-7}$  M dexamethasone (Merck, Darmstadt, Germany). Cells were diluted every day in the same medium until day 4. From day 4, IL6 was removed. From days 10 to 16 cells were switched to rHu Epo (2 U/mL) to obtain terminal erythroid differentiation.

The murine G1E-ER4 cell line expressing a GATA1-estrogen receptor fusion gene<sup>72</sup> was cultured in IMDM, containing 20% fetal calf serum (FCS; GE Healthcare, Chicago, IL), 100 U/mL penicillin, 100  $\mu$ g/mL streptomycin, 2 mM L-glutamine, 2 U/mL Epo, SCF in Chinese Hamster Ovary cell conditioned medium, monothioglycerol and 0.5  $\mu$ g/mL puromycin (Merck, Darmstadt, Germany) to select cells expressing the GATA1-ER fusion gene. This cell line was used to edit mutant *Sf3b1*<sup>K700E</sup> and isogenic *Sf3b1*<sup>WT</sup> using CRISPR-Cas9 strategy<sup>11</sup>.

For some experiments, cells were arrested in G1/S phase with 0.6  $\mu$ g/mL aphidicolin (APH, Merck, Darmstadt, Germany, cat no. 38966-21-1) or with 0.2 mM hydroxyurea (HU, Merck, cat no. H8627), or treated with 0.1 to 1  $\mu$ M histone deacetylase inhibitor SAHA/vorinostat (HDACi, Merck, cat no. #SML0061) for 20 h.

### RNA-sequencing

RNA-seq data from a first cohort of BM MNC from 27 lower-risk (LR)-MDS (21*SF3B1*<sup>MUT</sup> and 6 *SF3B1*<sup>WT</sup> (4 *SRSF2*<sup>MUT</sup> and 2 *SF*<sup>WT</sup>) previously published<sup>2</sup> were re-analyzed. RNA-sequencing of two additional cohorts was performed: one cohort of BM MNC from 185 LR-MDS (74 *SF3B1*<sup>MUT</sup>, 30 *SRSF2*<sup>MUT</sup>, 11 *U2AF1*<sup>MUT</sup>, 70 triple-negative samples) and one cohort of basoE and/or polyE obtained from 13 MDS (8 *SF3B1*<sup>MUT</sup> and 5 *SF3B1*<sup>WT</sup>). RNA integrity (RNA integrity number  $\geq$  7.0) was checked on the Agilent Fragment Analyzer (Agilent, Santa Clara, CA)

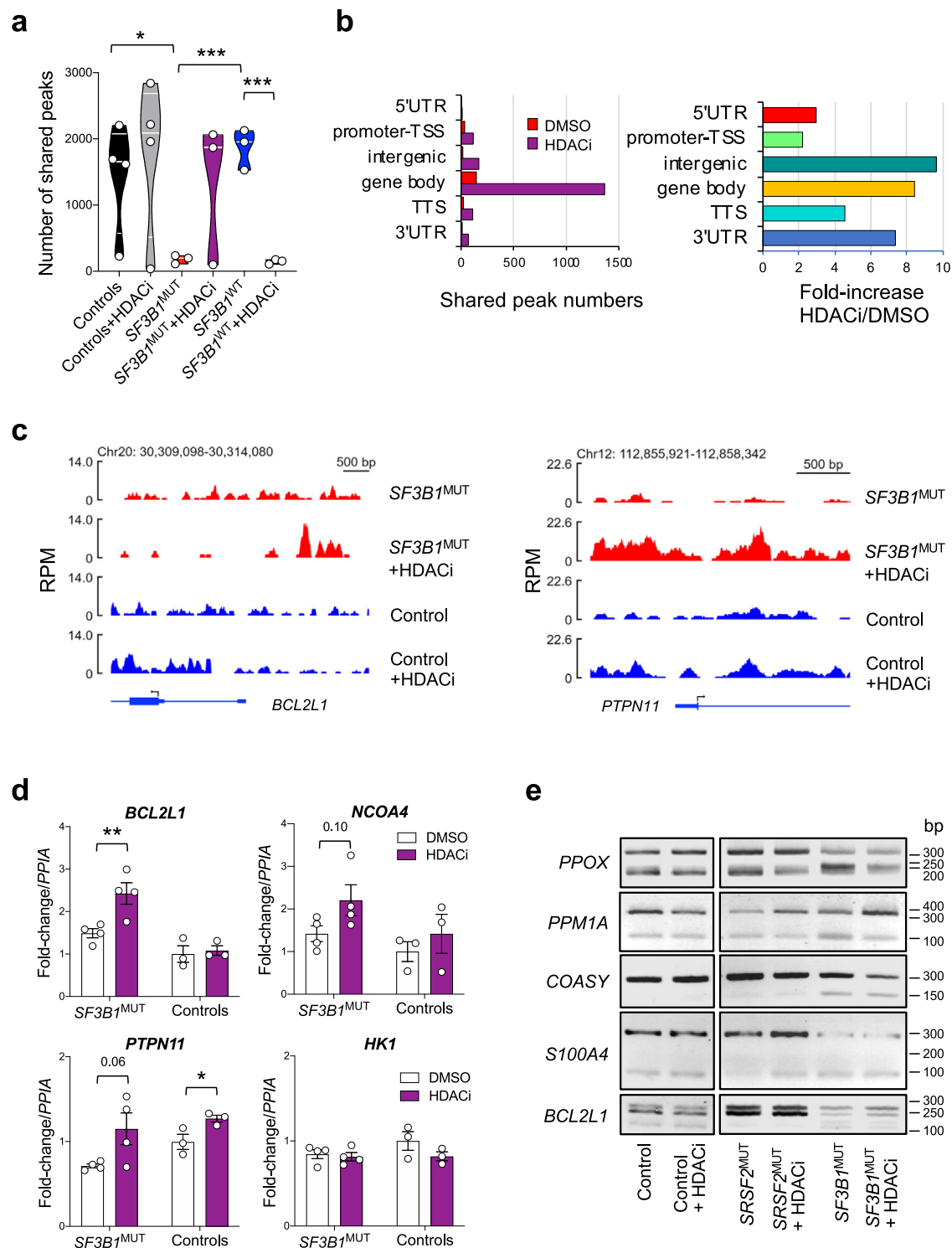
and quantities were determined using Qubit (Invitrogen, Waltham, CA). 50-100 ng of total RNA sample was used for poly-A mRNA selection using oligo(dT) beads and subjected to thermal fragmentation. For BM MNC, MuLV Reverse Transcriptase (Invitrogen) was used for cDNA synthesis. Libraries were constructed using the TruSeq Stranded mRNA Sample Preparation Kit (Illumina, San Diego, CA) and sequenced on an Illumina HiSeq 2500 platform using a 100-bp paired-end sequencing strategy. For RNA-sequencing of erythroblasts, fragmented mRNA samples were subjected to cDNA synthesis, converted into double stranded DNA using SureSelect Automated Strand Specific RNA Library Preparation Kit. Libraries were bar-coded, and subjected to 100-bp paired-end sequencing on Novaseq-6000 sequencer (Illumina, San Diego, CA). For RNA-sequencing of murine G1E-ER4 erythroblasts, libraries were constructed using the TruSeq Stranded mRNA Sample Preparation Kit (Illumina) and sequenced on an Illumina NextSeq500 platform (Illumina) using a 75-bp paired-end sequencing strategy.

### Bioinformatic analysis of RNA-sequencing

FASTQ files were mapped using STAR (v2.7.9) to align the reads against the human reference genome GRCh37 (UCSC version hg19) downloaded from the GENCODE project website. Reference genome and annotations are available on Gencode ([https://www.encodegenes.org/human/release\\_19.html](https://www.encodegenes.org/human/release_19.html))<sup>73</sup>. Read count normalizations and groups comparisons were performed using DESeq2 (v1.30.1), with the Wald test for significance testing. Genes with low counts were filtered, if at least half the samples have less than 20 normalized reads. Outliers removal with Cook's distance was left at default. For differential expression study, the results obtained after DESeq2 comparison were selected for further analysis and filtered at Benjamini-Hochberg (BH)-adjusted *P* value < 0.05 and  $\log_2(\text{FC}) > |1|$ <sup>74</sup>. To identify differentially expressed splicing events, we used KisSplice (v2.6.2), KisSplice2r-efgenome (v2.0.7) and KissDE (v1.15.3)<sup>75,76</sup>, and filtered results at BH-adjusted *P* value < 0.05 and  $\Delta\text{PSI} > |0.10|$ . Biological processes associated with genes that were differentially expressed or spliced were determined using the Gene Ontology enrichment or over-representation analyses and Gene Set Enrichment analysis with reference to specific gene sets (Supplementary Data 7).

### Sample preparation and mass spectrometry analysis

Sample preparation was done using the FASP procedure<sup>77</sup>. Briefly, cells were solubilized in 100  $\mu$ L of Tris/HCl 100 mM pH8.5 buffer containing 2% sodium dodecyl sulfate (SDS). Total protein amounts were quantified using BiCinchoninic acid Assay (BCA; Thermo Fischer Scientific). Then, 10 mM TCEP and 40 mM chloroacetamide were added and the samples were boiled for 5 min. 50  $\mu$ g of proteins were sampled and treated with urea to remove SDS. After urea removal, proteins were digested overnight with 1  $\mu$ g of sequencing-grade modified trypsin

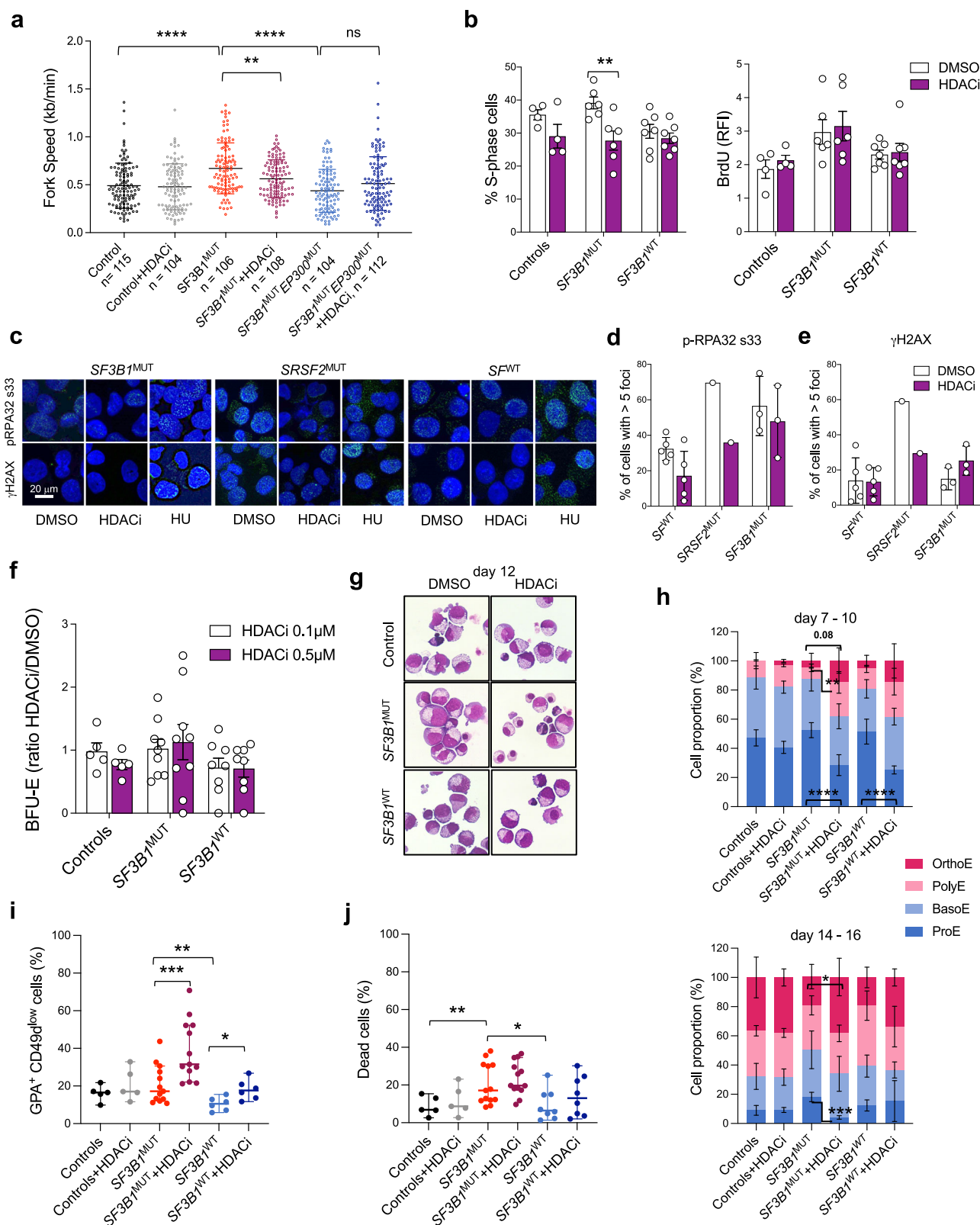


**Fig. 6 | Targeting of R-loops by HDACi may change gene expression without modifying the pattern of spliced isoforms. a** Violin plots with medians and quartiles representing R-loops as shared peak numbers by DRIP-seq of 3 *SF3B1*<sup>MUT</sup>, 3 *SF3B1*<sup>WT</sup> (2 *SRSF2*<sup>MUT</sup>, 1 *TET2*/*NRAS*<sup>MUT</sup>) and 4 control erythroblasts  $\pm$  HDACi at 0.5  $\mu$ M for 20 h. Two-sided unpaired t-test. **b** R-loop annotation to gene features in *SF3B1*<sup>MUT</sup> samples. Left panel: Shared peak numbers. Right panel: Fold-increase of peak number between HDACi and DMSO conditions. **c** R-loop profiles

near *BCL2L1* and *PTPN11* promoter. RPM: reads per million. **d** Quantification of *BCL2L1*, *NCOA4*, *PTPN11* and *HK1* transcripts by RT-qPCR in 4 *SF3B1*<sup>MUT</sup> and 3 control samples. Mean quantities normalized to *PPIA*  $\pm$  SD. Two-sided unpaired t-test. **e** Expression of *PPOX*, *PPM1A*, *COASY*, *S100A4*, and *BCL2L1* transcript isoforms by RT-PCR representative of 4 *SF3B1*<sup>MUT</sup>, 6 *SF3B1*<sup>WT</sup> and 2 control samples. \*\*\*\*  $P < 0.0001$ , \*\*\*  $P < 0.001$ , \*\*  $P < 0.01$ , \*  $P < 0.05$ . Source data are provided as a Source Data file.

(Promega, Madison, WI) in 50 mM Tris/HCl pH8.5 buffer. Peptides were recovered by filtration, desalted on C18 reverse phase StageTips and dried. They were then separated in 5 fractions by strong cationic exchange StageTips and analyzed using an Orbitrap Fusion mass

spectrometer (Thermo Fisher Scientific). Peptides from each fraction were separated on a C18 reverse phase column (2  $\mu$ m particle size, 100 Å pore size, 75  $\mu$ m inner diameter, 25 cm length) with a 170 min gradient starting from 99% of solvent A containing 0.1%



formic acid in milliQ-H<sub>2</sub>O and ending in 55% of solvent B containing 80% acetonitrile and 0.085% formic acid in milliQ-H<sub>2</sub>O. The mass spectrometer acquired data throughout the elution process. The MS1 scans spanned from 350 to 1500 Th with 1.10<sup>6</sup> Automatic Gain Control (AGC) target, 60 ms maximum ion injection time (MIIT) and resolution of 60 000. MS Spectra were recorded in profile mode. High

energy Collision Dissociation (HCD) fragmentations were performed from the most abundant ions in top speed mode for 3 seconds with a dynamic exclusion time of 30 s. Precursor selection window was set at 1.6Th. HCD Normalized Collision Energy was set at 30% and MS/MS scan resolution was set at 30000 with AGC target 1.10<sup>5</sup> within 60 ms MIIT.

**Fig. 7 | HDACi improves the differentiation of human *SF3B1*<sup>MUT</sup> erythroblasts.** **a** Scatter plot of fork speed measurement (kb/min) by DNA combing in 2 *SF3B1*<sup>MUT</sup> (1 *SF3B1/DNMT3A*, 1 *SF3B1/TET2/EP300*) and 1 control ± HDACi 0.2 μM for 20 h expressed in means ± SD. Two-sided Mann-Whitney test. **b** Cell cycle analysis by BrdU incorporation in 6 *SF3B1*<sup>MUT</sup>, 7 *SF3B1*<sup>WT</sup> and 4 controls. Left panel: Mean percentages ± SD of S-phase cells. Right panel: Mean BrdU RFIs ± SD in S-phase. Two-sided paired t-test. *SF3B1*<sup>MUT</sup> versus *SF3B1*<sup>MUT</sup> + HDACi, *P* = 0.002. **c–e** Immunofluorescence images of p-RPA32s33 and γH2AX representative of 3 *SF3B1*<sup>MUT</sup>, 6 *SF3B1*<sup>WT</sup> MDS (1 *SRSF2*<sup>MUT</sup>, 3 *SF*<sup>WT</sup>, 2 w/o mutation) at day 11. Nuclei labelling with DAPI. Magnification X100 (scale: 20 μm). **d, e** Mean percentages ±

SD of positive cells with > 5 intranuclear foci. **f**. Burst forming unit-erythroid (BFU-E) colony assays in 9 *SF3B1*<sup>MUT</sup>, 8 *SF3B1*<sup>WT</sup> and 5 controls. Mean ratios between HDACi and DMSO conditions ± SD. **g** May-Grünwald-Giemsa-stained cytopins (d12). **h** Proportions of erythroid precursors in 7 *SF3B1*<sup>MUT</sup>, 3 *SF3B1*<sup>WT</sup> and 4 controls at d7-10 and d14-16. Means ± SEM and 2-way ANOVA multiple comparisons for *q* values. **i** Scatter plots showing differentiation by flow cytometry expressed as mean percentages ± SD of GPA<sup>+</sup>CD49d<sup>low</sup> cells. **j** Scatter plots showing mean percentages ± SD of dead cells (FSC/SSC) at d12-14. **i, j** Two-sided paired t-test for *P* values. \*\*\*\* *P* < 0.0001, \*\*\* *P* < 0.001, \*\* *P* < 0.01, \* *P* < 0.05. Source data are provided as a Source Data file.

### Proteomic data analysis

The mass spectrometry data were analyzed using Maxquant version 2.1.1.0<sup>78</sup>. The database used was a concatenation of Human sequences from the Uniprot-Swissprot database (Uniprot, release 2022-05) and the list of contaminant sequences from Maxquant. Cystein carbamidomethylation was set as constant modification and acetylation of protein N-terminus and oxidation of methionine were set as variable modifications. Second peptide search and the “match between runs” (MBR) options were allowed. False discovery rate (FDR) was kept below 1% on both peptides and proteins. Label-free protein quantification (LFQ) was done using both unique and razor peptides with at least 2 peptides ratios required for LFQ.

Statistical analysis was done using “R”. Among identified proteins, those with at least 70% of values in at least one condition were selected. Then, proteins with t-test *P* value < 0.05 were defined as significantly differentially expressed, and proteins with 100% of valid values in one condition and 0% of valid values in the other condition were defined as “Appeared” or “Disappeared”. Log<sub>2</sub>(LFQ intensity) matrix were filtered before imputation. For heatmaps, Z-score were calculated after imputation.

Functional analyses were generated through Ingenuity Pathway Analysis (QIAGEN Inc., <https://www.qiagenbioinformatics.com/products/ingenuitypathway-analysis>) version 76765 844 for each list of differential proteins. Significantly over-represented biological terms (Canonical pathways or Diseases and functions) were identified with a right-tailed Fisher’s Exact test that calculates an overlap *P* value determining the probability that each term associated with our lists of differential proteins was due to chance alone. The z-score is a statistical measure of correlation between relationship direction and experimental protein expression. Its calculation assessed the activation (positive z-score) or repression (negative one) of each term. To be considered significant the z-score has to be greater than 2 in absolute value. Overlap of selected pathways and functions was designed using Cytoscape version 3.9.1.

### DRIP-qPCR and DRIP-sequencing

DRIP-qPCR or DRIP-seq was performed on human CD34-derived proE/basoE from 6 MDS with *SF3B1* mutation, 6 MDS without *SF3B1* mutation and 4 healthy controls<sup>79</sup>. Briefly, 8 × 10<sup>6</sup> cells were harvested and processed for genomic DNA extraction by SDS/ Proteinase K treatment at 37 °C followed by phenol-chloroform extraction and ethanol precipitation. 8 μg of genomic DNA was fragmented using HindIII, EcoRI, BsrGI, XbaI, and SspI. Then DNA pre-treated or not with 16 U/ml RNase HI (New Engl Biolabs, Ipswich, MA) overnight was immunoprecipitated using S9.6 antibody at 40 μg/ml (Kerafast, Shirley, MA or ATCC mouse hybridoma) overnight at 4 °C (Supplementary Methods 1). Quality control of immunoprecipitation was performed by qPCR, at R-loop-positive loci (*RPL13A*, *CALM3*, *TFPT*) and R-loop-negative loci (*EGRI*, *SNRPI*). DRIP-qPCR was also performed at specific loci (*ABCC5*, *IREB2*, *TCIRG1*, *TMX2*) (Supplementary Method 2).

Percentage of input expected between 1-15%

and fold enrichment (expected between 20-300)

$$\text{Fold enrichment} = \frac{[2^{(\text{Ct input (positive locus corrected)})} - \text{Ct DRIPed (DNA positive locus)}]}{[2^{(\text{Ct input negative locus (corrected)})} - \text{Ct DRIPed (negative locus)}]}$$

were calculated to assess the immunoprecipitation efficiency and specificity, respectively. For DRIP-seq, good quality DRIPed DNA samples (with and without RNase HI-treatment) were sonicated to get an average length of 200-300 bp. Then samples were subjected to end-repair, dATP tailing, adaptor ligation and library indexing. Lastly, the libraries were cleaned up using AMPure beads, and amplified. After checking library quality on Agilent Bioanalyzer using Agilent High Sensitivity DNA 1000 kit, libraries were sequenced on Illumina NovaSeq instruments.

### DRIP-seq bioinformatic analysis

100 bp paired-end reads were trimmed and filtered using fastp (v0.23.4) to remove low quality reads, low complexity sequences, as well as polyG tails (corresponding to no signal in the Illumina two-color systems, in NovaSeq data). Reads were mapped to the human reference genome (GENCODE, GRCh37; [https://www.encodegenes.org/human/release\\_19.html](https://www.encodegenes.org/human/release_19.html)) with Bowtie2 (v2.3.5.1), using parameters – no-discordant and – no-mixed. Resulting SAM files were piped through Samblaster (v0.1.24) to remove duplicate reads, then through Samtools (v1.10) to generate sorted (by coordinates) BAM files, with a cut-off for MAPQ score of 10. Peaks were called for each replicate with the MACS algorithm (MACS3) in broad mode, with the input DNA as control as well as the same sample treated with RNase HI, at *q* value < 0.1. Resulting peaks were analysed by groups of biological replicates with MSPC (v5.5.0)<sup>80</sup>. Firstly, by categorizing them as either background, weak, or stringent (with both cut-off on *P* value at 1e-4 and 1e-8). Weak peaks were rescued if stringent in other biological replicates. Peaks were then confirmed or discarded based on the combined stringency test supported by enough replicates and if their combined stringency, using Fisher’s combined probability test, satisfies the threshold of 1e-8. Confirmed peaks are qualified as true positive if they pass the Benjamini-Hochberg (BH) multiple testing correction at level 0.05.

To identify differentially expressed R-loops, we intersected the peaks from MACS3 calling with the 5.5 million of restriction fragments generated before S9.6 immunoprecipitation, using the resulting matrix as a reference frame for featureCounts (v2.0.0). Normalization and differential R-loop expression analysis between wild-type, mutant and control samples was performed with DESeq2, using BH-adjusted *P* value < 0.05 and log<sub>2</sub> (FC) > |1|. Bedtools (v2.27.1) intersect was used for overlap analyses.

### May-Grünwald Giemsa staining and flow cytometry

Erythroid differentiation of human primary erythroblasts and murine GIE-ER4 clones was followed by May-Grünwald Giemsa staining of cytopins. Cell viability was assessed by the scatter profile (FSC/SSC) using flow cytometry. For erythroid differentiation, 5 × 10<sup>4</sup> cells are washed in 1X phosphate buffered saline (PBS) supplemented with 2% FCS and incubated for 20 min at 4 °C with fluorescent antibodies to

%input = 100 × 2<sup>(Ct input corrected – Ct DRIPed DNA)</sup> where Ct (cycle threshold) input (corrected) = (Ct input – log<sub>2</sub>(10))



GPA (CD235a), CD49d and CD71 for human primary cells (Beckman Coulter, Brea, CA) or Kit (CD117) and Ter-119 for murine cells (BD Biosciences, Franklin Lane, NJ). Analysis was performed on LSRFortessa apparatus (BD Biosciences) with Kaluza software (Beckman Coulter) (Supplementary Methods 3).

### Cell cycle analysis and live cell imaging

Cell cycle was analysed by double labelling using a fluorescent anti-BrdU antibody and 7-aminoactinomycin D (7-AAD, BD Biosciences). Cells were incubated with 10  $\mu$ M BrdU for 30 min at 37 °C (BD Biosciences). Then cells were pelleted, fixed in 500  $\mu$ L 70% ethanol for 20 min at room temperature and washed in 1X PBS supplemented with 0.5% bovine serum albumin (BSA). Cell pellet is resuspended in 2 M chlorhydric acid for 20 min at room temperature and washed in 1X PBS. Acid was neutralized by a 0.1 M borate solution pH 8.5 for 2 min (Borax Na<sub>2</sub>B<sub>4</sub>O<sub>7</sub>, Merck). Cells were washed 3 times and transferred to 96-well plates for incubation with 10  $\mu$ L anti-BrdU antibody or isotype control for 20 min at room temperature either FITC-anti-BrdU (BD Biosciences, clone 3D4) or APC-anti-BrdU (BD Biosciences, clone 3D4) at a final concentration of 0.5  $\mu$ g/mL. Cells were washed 3 times and incubated with 50  $\mu$ L of a 10  $\mu$ g/mL 7-AAD and 50  $\mu$ g/mL RNase A (Macherey-Nagel, Hoerdt, France) solution for 30 min at room temperature. Finally, cells were transferred into Eppendorf vials containing 100  $\mu$ L 1X PBS and analysed on BD Accuri C6 flow cytometer using CFlow Plus software (BD Biosciences) (Supplementary Method 4). The proliferation capacities of GIE-ER4 cells treated or not with  $\beta$ -estradiol at 10<sup>-7</sup> M (Merck) were measured using IncuCyte live-cell imaging system (Essen Instruments, Ann Arbor, MI). 48-well plates were coated with 0.01% Poly-L-lysine sterile-filtered solution (Merck). Cells were seeded at 30,000 cells/well and monitored over time with 9 images per well every 3 hours for 72 h. Results were expressed as cell confluence (in % of occupied space).

### DNA fiber combing

DNA replication was analyzed by pulse labelling with fluorescent thymidine analogs of human erythroblasts. Cells were first labelled with 20  $\mu$ M iododesoxyuridine (IdU) for 30 min at 37 °C and then with 100  $\mu$ M chlorodesoxyuridine (CldU) for 30 min at 37 °C. DNA replication was blocked by the addition of an excess of thymidine (300  $\mu$ M) on ice and cells were washed in 1X PBS and counted. DNA combing was performed after DNA extraction in 1% low melting agarose plugs (0.3  $\times$  10<sup>6</sup> cells in 90  $\mu$ L per plug). Briefly, low melting agarose was maintained at 45 °C. Cells were resuspended at 6.66  $\times$  10<sup>6</sup> in one volume of 1X PBS, mixed in the same volume of 2% agarose and distributed in plug mold. Then proteins in agarose plugs were digested with 1 mg/mL proteinase K in 0.25 M EDTA pH 8/ 1% SDS at 42 °C for 48 h. Finally, plugs were washed 3 times in 10 mM Tris-HCl pH 8/1 mM EDTA (TE), and agarose was eliminated by digestion using  $\beta$ -agarase for 48 h. For DNA fiber stretching, extracted DNA was resuspended in 0.25 M 2-(N-morpholino)-ethanesulfonic acid pH 5.5 in Eppendorf vials for 30 min at 65 °C, 30 min at room temperature and 2 weeks at 4 °C. Before combing, vials were placed at room temperature for 30 min. DNA was then placed in FiberComb reservoir (Genomic Vision, Bagneux, France) and stretched on silane treated coverslips. Coverslips were stuck on glass slides and incubated for 2 h at 60 °C and stored at -20 °C. For hybridization, DNA was denatured in a 1N NaOH solution, rinsed in cold 1X PBS and dehydrated in 70%, 85% and 100% ethanol. Aspecific sites were blocked for 30 min at 37 °C. DNA was first incubated with primary anti-IdU (BD Biosciences, cat no. 347580, dilution: 1/50) mouse antibody and anti-CldU (Abcam, cat no. Ab 6326, dilution: 1/50) rat antibody, and then with fluorescent secondary antibodies. Finally, DNA was labelled with a primary anti-ssDNA antibody (mouse), and with a first fluorescent secondary anti-mouse antibody (goat), and with a second fluorescent secondary anti-goat antibody (donkey). Slides were mounted in VectaShield medium (Vector Laboratories, Burlingame, CA). All antibodies

and reagents are described in Supplementary Methods 5. Fiber length and symmetry of a minimum of 200 fibers per group were measured. Fork symmetry was expressed as the IdU/CldU ratio. The speed of the replication fork was calculated by the ratio  $(d_i + d_{ci}) / (t_i + t_{ci})$ , where  $d_i$  and  $t_i$  represent respectively the measured distance (in kb) and labelling time (in min) for IdU incorporation, and  $d_{ci}$  and  $t_{ci}$  denote the corresponding parameters for CldU incorporation.

### Immunofluorescence experiments

Immunofluorescence was used to detect phospho-RPA32 (p-RPA32, serine 33 and serine 4/8), phospho-H2AX ( $\gamma$ -H2AX, serine 139) and 53BP1 nuclear foci. Cells treated with 5 mM hydroxyurea for 3 h were used as positive controls (Merck). Briefly, cytopins were prepared with 10<sup>5</sup> cells in 1X PBS, fixed in 1X PBS/2% paraformaldehyde (Santa Cruz Biotechnologies, Santa Cruz, CA) for 20 min at room temperature and washed. Cells were permeabilized in 1X PBS/0.5% Triton X-100 for 10 min and washed in cold 1X PBS. Saturation was performed using 1X PBS/3% BSA (Euromedex, Souffelweyersheim, France) for 30 min and cells were incubated with primary antibodies for 1 h at 37 °C (Supplementary Methods 6). After washing, cells were incubated with secondary antibody for 30 min at 37 °C in the dark. DNA was counterstained with 4',6-diamidino-2-phenylindole (DAPI) for 5 min and cytopins were rinsed in 1X PBS and mounted with Fluoromount-G (Clinisciences, Nanterre, France). Analysis was conducted using an inverted DMI600 microscope at 100X magnification (Leica, Wetzlar, Germany). Images were analyzed using the ImageJ software (NIH, Bethesda, MD).

### Western blot analysis

Cell lysates were solubilized for 5 minutes at 95 °C in Laemmli buffer (65 mM Tris [pH 6.8], 20% glycerol, 5%  $\beta$ -mercaptoethanol, 0.01% bromophenol blue, and 2% sodium dodecyl sulfate [SDS]) Proteins were separated by SDS-polyacrylamide gel electrophoresis and transferred to a nitrocellulose membrane (VWR, Radnor, PA). Membranes were blocked with 5% of dry milk in TBS-T buffer (10 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.15% Tween 20) for 1 h and incubated in specific antibody overnight at 4 °C (Supplementary Methods 7). Membranes were washed in TBS-T buffer and incubated for 1 h at room temperature with secondary horseradish peroxidase (HRP)-linked antibody (horse anti-mouse HRP-linked antibody 7076 S or goat anti-rabbit HRP-linked antibody 7074 V, Cell Signaling, Danvers, MA). Enzyme activity was visualized by an ECL-based detection system (VWR, Radnor, PA). Blot imaging was performed on the Fujifilm LAS-3000 Imager (Fujifilm, Tokyo, Japan) and images were analysed using the Multi Gauge software (Fujifilm).

### Targeted LC-MS metabolomics analyses

3.10<sup>5</sup> *Sf3b1*<sup>K700E/+</sup> or *Sf3b1*<sup>+/+</sup> GIE-ER4 cells were collected after 0 h or 24 h of  $\beta$ -estradiol treatment and with or without 0.2 mM HU for 16 h (Merck). For metabolomics analysis, extraction was performed in 30  $\mu$ L of 50% methanol, 30% acetonitrile (ACN) and 20% water. After centrifugation at 16,000 g for 15 min at 4 °C, supernatants were collected and stored at -80 °C until analysis. LC/MS analyses were conducted on a QExactive Plus Orbitrap mass spectrometer equipped with an Ion Max source and a HESI II probe coupled to a Dionex UltiMate 3000 uHPLC system (Thermo). Samples (5  $\mu$ L) were injected onto a ZIC-pHILIC column with a guard column (Millipore) for LC separation in a gradient of buffer A (20 mM ammonium carbonate, 0.1% ammonium hydroxide pH 9.2), and buffer B (ACN) with a flow rate of 0.200  $\mu$ L  $\cdot$  min<sup>-1</sup> as follows: 0–20 min, linear gradient from 80% to 20% of buffer B; 20–20.5 min, linear gradient from 20% to 80% of buffer B; 20.5–28 min, 80% buffer B. The mass spectrometer was operated in full scan, polarity switching mode with the spray voltage set to 2.5 kV and the heated capillary held at 320 °C. The sheath gas flow was set to 20 units, the auxiliary gas flow to 5 units and the sweep gas flow to 0

units. The metabolites were detected across a mass range of 75–1000 m/z at a resolution of 35,000 (at 200 m/z) with the automatic gain control target at  $10^6$  and the maximum injection time at 250 ms. Lock masses were used to ensure mass accuracy below 5 ppm. Data were acquired with Thermo Xcalibur software (Thermo Fisher Scientific). The peak areas of metabolites were determined using Thermo TraceFinder software (Thermo Fisher Scientific), identified by the exact mass of each singly charged ion and by the known retention time on the HPLC column. Each metabolite was quantified as the area under the curve and results were expressed as arbitrary unit (A.U).

### RT-PCR and PCR on colonies

For RT-PCR, RNA was extracted using RNeasy Mini Kit (Qiagen) and retrotranscribed with the Maxima First Strand cDNA synthesis kit (Thermo Fisher Scientific). cDNA was amplified using Phire Hot Start II DNA polymerase (Thermo Fisher Scientific). Amplicons were analysed by electrophoresis on a 2% agarose gel. For fluorescent RT-PCR, the PCR was performed using the same primers except that the forward primers were labelled with 6-carboxyfluorescein (6-FAM) at 5' end. For capillary electrophoresis, 1  $\mu$ L of diluted PCR product was added to 0.2  $\mu$ L of GeneScan 500 ROX dye standard and 18  $\mu$ L of RNase-free water. After denaturation for 5 min at 95 °C, fragments were separated using the 3730xl DNA analyzer and analyses was performed using GeneMapper Software 5 (Thermo Fisher Scientific). For qPCR, the primers were used with the SYBRGreen Master Mix (Meridian Bioscience) in a LightCycler480 (Roche). Expression levels were normalized to actin (*ACT*) and cyclophilin A (*PPIA*) expression using geometric averaging, and analyzed using  $\Delta\Delta$ Ct method. Primers for RT-PCR, fluorescent RT-PCR and qPCR are listed in Supplementary Methods 8.

For the detection of mutants by PCR, colonies were picked and transferred in 50  $\mu$ L of lysis buffer (KCl 50 mM, Tris HCl pH8 10 mM, 0.4%NP-40, 0.4% Tween-20, 0.2 mg/mL proteinase K). After 1h incubation at 55 °C, DNA was treated for 10 min at 95 °C. Samples (5  $\mu$ L) was used for 40 cycles of PCR cycles (30 s at 95 °C, 30 s at 60 °C, 1 min at 72 °C) using specific primers (Supplementary Data 2). Amplicons were analysed by gel electrophoresis.

### Statistical analysis

For quantitative variables, values were expressed as median and interquartile range (IQR) or means and standard error of the mean (SEM) and compared using the Student t-test or non parametric Mann-Whitney or Kruskal-Wallis tests. Chi-squared or Fisher exact tests were used to compare categorical variables. For transcript quantification, the Mann-Whitney test was used to assign a statistical significance for each group comparison. *P* values < 0.05 were considered significant (JMP version 10.0.2, SAS Institute Inc, Cary, NC).

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The RNA-sequencing and DRIP-sequencing data generated in this study have been deposited in the NCBI's Gene Expression Omnibus (GEO) database. Accession codes are provided below. The processed RNA-sequencing and DRIP-sequencing data are available in the Supplementary information as Supplementary Data indicated for each of them. RNA-seq BM MNC cohort of 27 cases (Supplementary Data 1): [GSE220525](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE220525), RNA-seq of human basophilic erythroblasts and polychromatophilic erythroblasts (Supplementary Data 2): [GSE220523](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE220523). DRIP-seq of human basophilic erythroblasts (Supplementary Data 4): [GSE220271](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE220271). RNA-seq of mouse GIEER erythroblasts (Supplementary Data 5): [GSE220516](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE220516). RNA-seq data of the BM MNC cohort of 185 MDS patients (Supplementary Fig. 1b) are available at [GSE220518](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE220518) under

restricted access since these data are considered sensitive personal data according to the European Union General Data Protection Regulation (GDPR) and thus cannot be shared with third-parties without prior approval. Access can only be granted for research purposes. An application must be sent to [michaela.fontenay@inserm.fr](mailto:michaela.fontenay@inserm.fr). The proteomic data are available on ProteomeXchange Consortium via the PRIDE partner repository: Human erythroblast proteome (Supplementary Data 3) and mouse erythroblast proteome (Supplementary Data 5): [PXD038700](https://www.ebi.ac.uk/PRIDE/archive/PXD038700). Source data are provided with this paper.

### References

- Yoshida, K. et al. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* **478**, 64–69 (2011).
- Khoury, 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 (2022).
- Arber, D. A. et al. International Consensus Classification of Myeloid Neoplasms and Acute Leukemias: integrating morphologic, clinical, and genomic data. *Blood* **140**, 1200–1228 (2022).
- Hellström-Lindberg, E. Efficacy of erythropoietin in the myelodysplastic syndromes: a meta-analysis of 205 patients from 17 studies. *Br. J. Haematol.* **89**, 67–71 (1995).
- Park, S. et al. Outcome of lower-risk patients with myelodysplastic syndromes without 5q deletion after failure of erythropoiesis-stimulating agents. *J. Clin. Oncol.* **35**, 1591–1597 (2017).
- Fenaux, P. et al. Luspatercept in patients with lower-risk myelodysplastic syndromes. *N. Engl. J. Med.* **382**, 140–151 (2020).
- Zeidan, A. M. et al. Longer-term benefit of luspatercept in transfusion-dependent lower-risk myelodysplastic syndromes with ring sideroblasts. *Blood* **140**, 2170–2174 (2022).
- Platzbecker, U. et al. Long-term utilization and benefit of luspatercept in transfusion-dependent, erythropoiesis-stimulating agent-refractory or -intolerant patients with lower-risk myelodysplastic syndromes with ring sideroblasts. *Leukemia* **37**, 2314–2318 (2023).
- Darman, R. B. et al. Cancer-associated SF3B1 hotspot mutations induce cryptic 3' splice site selection through use of a different branch point. *Cell Rep.* **13**, 1033–1045 (2015).
- Alsafadi, S. et al. Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. *Nat. Commun.* **7**, 10615 (2016).
- Bondu, S. et al. A variant erythroferrone disrupts iron homeostasis in SF3B1-mutated myelodysplastic syndrome. *Sci. Transl. Med.* **11**, eaav5467 (2019).
- Nikpour, M. et al. The transporter ABCB7 is a mediator of the phenotype of acquired refractory anemia with ring sideroblasts. *Leukemia* **27**, 889–896 (2013).
- Clough, C. A. et al. Coordinated missplicing of TMEM14C and ABCB7 causes ring sideroblast formation in SF3B1-mutant myelodysplastic syndrome. *Blood* **139**, 2038–2049 (2022).
- Dalton, W. B. et al. Hotspot SF3B1 mutations induce metabolic reprogramming and vulnerability to serine deprivation. *J. Clin. Invest.* **129**, 4708–4723 (2019).
- Kim, E. et al. SRSF2 mutations contribute to myelodysplasia by mutant-specific effects on exon recognition. *Cancer Cell* **27**, 617–630 (2015).
- Yip, B. H. et al. The U2AF1S34F mutation induces lineage-specific splicing alterations in myelodysplastic syndromes. *J. Clin. Invest.* **127**, 2206–2221 (2017).
- Shiozawa, Y. et al. Aberrant splicing and defective mRNA production induced by somatic spliceosome mutations in myelodysplasia. *Nat. Commun.* **9**, 3649 (2018).
- Walter, M. J. et al. Clonal architecture of secondary acute myeloid leukemia. *N. Engl. J. Med.* **366**, 1090–1098 (2012).

19. Flach, J. et al. Replication stress signaling is a therapeutic target in myelodysplastic syndromes with splicing factor mutations. *Haematologica* **106**, 2906–2917 (2021).
20. Singh, S. et al. SF3B1 mutations induce R-loop accumulation and DNA damage in MDS and leukemia cells with therapeutic implications. *Leukemia* **34**, 2525–2530 (2020).
21. Chen, L. et al. The augmented R-Loop is a unifying mechanism for myelodysplastic syndromes induced by high-risk splicing factor mutations. *Mol. Cell* **69**, 412–425.e6 (2018).
22. Paulsen, R. D. et al. A genome-wide siRNA screen reveals diverse cellular processes and pathways that mediate genome stability. *Mol. Cell* **35**, 228–239 (2009).
23. Savage, K. I. et al. Identification of a BRCA1-mRNA splicing complex required for efficient DNA repair and maintenance of genomic stability. *Mol. Cell* **54**, 445–459 (2014).
24. Papaemmanuil, E. et al. Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *N. Engl. J. Med.* **365**, 1384–1395 (2011).
25. Malcovati, L. et al. SF3B1 mutation identifies a distinct subset of myelodysplastic syndrome with ring sideroblasts. *Blood* **126**, 233–241 (2015).
26. Edwards, C. R. et al. A dynamic intron retention program in the mammalian megakaryocyte and erythrocyte lineages. *Blood* **127**, e24–e34 (2016).
27. Pimentel, H. et al. A dynamic intron retention program enriched in RNA processing genes regulates gene expression during terminal erythropoiesis. *Nucleic Acids Res.* **44**, 838–851 (2016).
28. Pimentel, H. et al. A dynamic alternative splicing program regulates gene expression during terminal erythropoiesis. *Nucleic Acids Res.* **42**, 4031–4042 (2014).
29. Li, X. & Manley, J. L. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* **122**, 365–378 (2005).
30. Bonnet, A. et al. Introns protect eukaryotic genomes from transcription-associated genetic instability. *Mol. Cell* **67**, 608–621.e6 (2017).
31. Jangi, M. et al. SMN deficiency in severe models of spinal muscular atrophy causes widespread intron retention and DNA damage. *Proc. Natl. Acad. Sci. USA* **114**, E2347–E2356 (2017).
32. Okamoto, Y. et al. FANCD2 protects genome stability by recruiting RNA processing enzymes to resolve R-loops during mild replication stress. *FEBS J.* **286**, 139–150 (2019).
33. Goulielmaki, E. et al. The splicing factor XAB2 interacts with ERCC1-XPF and XPG for R-loop processing. *Nat. Commun.* **12**, 3153 (2021).
34. Ginno, P. A., Lim, Y. W., Lott, P. L., Korf, I. & Chédin, F. GC skew at the 5' and 3' ends of human genes links R-loop formation to epigenetic regulation and transcription termination. *Genome Res.* **23**, 1590–1600 (2013).
35. Skourti-Stathaki, K., Proudfoot, N. J. & Gromak, N. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol. Cell* **42**, 794–805 (2011).
36. Promonet, A. et al. Topoisomerase 1 prevents replication stress at R-loop-enriched transcription termination sites. *Nat. Commun.* **11**, 3940 (2020).
37. Gan, W. et al. R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes Dev.* **25**, 2041–2056 (2011).
38. Toledo, L. I. et al. ATR prohibits replication catastrophe by preventing global exhaustion of RPA. *Cell* **155**, 1088–1103 (2013).
39. Ashley, A. K. et al. DNA-PK phosphorylation of RPA32 Ser4/Ser8 regulates replication stress checkpoint activation, fork restart, homologous recombination and mitotic catastrophe. *DNA Repair (Amst)* **21**, 131–139 (2014).
40. Salas-Armenteros, I. et al. Human THO-Sin3A interaction reveals new mechanisms to prevent R-loops that cause genome instability. *EMBO J.* **36**, 3532–3547 (2017).
41. Wong, J. J.-L. et al. Orchestrated intron retention regulates normal granulocyte differentiation. *Cell* **154**, 583–595 (2013).
42. Braunschweig, U. et al. Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res.* **24**, 1774–1786 (2014).
43. Hartono, S. R. et al. The Affinity of the S9.6 Antibody for Double-Stranded RNAs impacts the accurate mapping of r-loops in fission yeast. *J. Mol. Biol.* **430**, 272–284 (2018).
44. Smolka, J. A., Sanz, L. A., Hartono, S. R. & Chédin, F. Recognition of RNA by the S9.6 antibody creates pervasive artifacts when imaging RNA:DNA hybrids. *J. Cell Biol.* **220**, e202004079 (2021).
45. Sanz, L. A. et al. Prevalent, dynamic, and conserved R-loop structures associate with specific epigenomic signatures in mammals. *Mol. Cell* **63**, 167–178 (2016).
46. Castillo-Guzman, D. & Chédin, F. Defining R-loop classes and their contributions to genome instability. *DNA Repair (Amst)* **106**, 103182 (2021).
47. Huertas, P. & Aguilera, A. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol. Cell* **12**, 711–721 (2003).
48. Sollier, J. et al. Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Mol. Cell* **56**, 777–785 (2014).
49. Teloni, F. et al. Efficient Pre-mRNA cleavage prevents replication-stress-associated genome instability. *Mol. Cell* **73**, 670–683.e12 (2019).
50. Mosler, T. et al. R-loop proximity proteomics identifies a role of DDX41 in transcription-associated genomic instability. *Nat. Commun.* **12**, 7314 (2021).
51. Boddur, P. C. et al. Transcription elongation defects link oncogenic splicing factor mutations to targetable alterations in chromatin landscape. *Mol. Cell.* <https://doi.org/10.1016/j.molcel.2024.02.032> (2024).
52. Bester, A. C. et al. Nucleotide deficiency promotes genomic instability in early stages of cancer development. *Cell* **145**, 435–446 (2011).
53. Kotsantis, P. et al. Increased global transcription activity as a mechanism of replication stress in cancer. *Nat. Commun.* **7**, 13087 (2016).
54. Di Micco, R. et al. Oncogene-induced senescence is a DNA damage response triggered by DNA hyper-replication. *Nature* **444**, 638–642 (2006).
55. Rimmelé, P. et al. Spi-1/PU.1 oncogene accelerates DNA replication fork elongation and promotes genetic instability in the absence of DNA breakage. *Cancer Res.* **70**, 6757–6766 (2010).
56. Bhatia, V. et al. BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature* **511**, 362–365 (2014).
57. Sugimura, K., Takebayashi, S.-I., Taguchi, H., Takeda, S. & Okumura, K. PARP-1 ensures regulation of replication fork progression by homologous recombination on damaged DNA. *J. Cell Biol.* **183**, 1203–1212 (2008).
58. Maya-Mendoza, A. et al. High speed of fork progression induces DNA replication stress and genomic instability. *Nature* **559**, 279–284 (2018).
59. Sarni, D. et al. Topoisomerase 1-dependent R-loop deficiency drives accelerated replication and genomic instability. *Cell Rep.* **40**, 111397 (2022).

60. Hamperl, S. & Cimprich, K. A. The contribution of co-transcriptional RNA:DNA hybrid structures to DNA damage and genome instability. *DNA Repair (Amst)* **19**, 84–94 (2014).
61. Merrih, H., Machón, C., Grainger, W. H., Grossman, A. D. & Soutanas, P. Co-directional replication-transcription conflicts lead to replication restart. *Nature* **470**, 554–557 (2011).
62. García-Muse, T. & Aguilera, A. R loops: from physiological to pathological roles. *Cell* **179**, 604–618 (2019).
63. Yin, Y. et al. A basal-level activity of ATR links replication fork surveillance and stress response. *Mol. Cell* **81**, 4243–4257.e6 (2021).
64. Bianco, J. N. et al. Overexpression of Claspin and Timeless protects cancer cells from replication stress in a checkpoint-independent manner. *Nat. Commun.* **10**, 910 (2019).
65. Nguyen, H. D. et al. Functions of replication protein A as a sensor of R loops and a regulator of RNaseH1. *Mol. Cell* **65**, 832–847.e4 (2017).
66. Cheruiyot, A. et al. Nonsense-mediated RNA decay is a unique vulnerability of cancer cells harboring SF3B1 or U2AF1 mutations. *Cancer Res.* **81**, 4499–4513 (2021).
67. Lappin, K. M. et al. Cancer-associated SF3B1 mutations confer a BRCA-like cellular phenotype and synthetic lethality to PARP Inhibitors. *Cancer Res.* **82**, 819–830 (2022).
68. Park, K. et al. Aicardi-Goutières syndrome-associated gene SAMHD1 preserves genome integrity by preventing R-loop formation at transcription-replication conflict regions. *PLoS Genet.* **17**, e1009523 (2021).
69. Wang, H. et al. Disruption of dNTP homeostasis by ribonucleotide reductase hyperactivation overcomes AML differentiation blockade. *Blood* **139**, 3752–3770 (2022).
70. Sekeres, M. A. et al. Randomized Phase II Study of Azacitidine Alone or in Combination With Lenalidomide or With Vorinostat in Higher-Risk Myelodysplastic Syndromes and Chronic Myelomonocytic Leukemia: North American Intergroup Study SWOG S1117. *J. Clin. Oncol.* **35**, 2745–2753 (2017).
71. Prebet, T. et al. Addition of suberoylanilide hydroxamic acid (Vorinostat) to azacitidine for patients with higher risk myelodysplastic syndromes and azacitidine failure: a phase II add-on study from the Groupe Francophone des Myelodysplasies. *Br. J. Haematol.* **180**, 735–737 (2018).
72. Weiss, M. J., Yu, C. & Orkin, S. H. Erythroid-cell-specific properties of transcription factor GATA-1 revealed by phenotypic rescue of a gene-targeted cell line. *Mol. Cell. Biol.* **17**, 1642–1651 (1997).
73. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
74. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
75. Sacomoto, G. A. T. et al. KISSPLICE: de-novo calling alternative splicing events from RNA-seq data. *BMC Bioinformatics* **13**, S5 (2012). **Suppl 6**.
76. Benoit-Pilven, C. et al. Complementarity of assembly-first and mapping-first approaches for alternative splicing annotation and differential analysis from RNAseq data. *Sci. Rep.* **8**, 4307 (2018).
77. Gautier, E.-F. et al. Comprehensive proteomic analysis of human erythropoiesis. *Cell Rep.* **16**, 1470–1484 (2016).
78. Cox, J. et al. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell Proteomics* **13**, 2513–2526 (2014).
79. Sanz, L. A. & Chédin, F. High-resolution, strand-specific R-loop mapping via S9.6-based DNA-RNA immunoprecipitation and high-throughput sequencing. *Nat. Protoc.* **14**, 1734–1755 (2019).
80. Jalili, V., Matteucci, M., Masseroli, M. & Morelli, M. J. Using combined evidence from replicates to evaluate ChIP-seq peaks. *Bioinformatics* **31**, 2761–2769 (2015).

## Acknowledgements

The authors dedicate this work to the memory of their colleague Dr Angelos Constantinou whose discoveries inspired this work. The authors acknowledge Pr Seishi Ogawa (Kyoto University, Japan) for very helpful discussion and comments on the manuscript, Dr Evelyne Lauret for support and discussion, Institut Cochin, Paris; Dr Sarah Lambert (Institut Curie, Orsay), Dr Jean-Charles Cadoret (Institut Jacques Monod, Paris), Dr Valeria Naim (Institut Gustave Roussy, Villejuif), Pr Raphaël Itzykson (Institut de Recherche Saint-Louis, Paris) for discussion. The authors thank Dr Carole Almire, Laboratory of Hematology, Cochin Hospital, Dr Emilie-Fleur Gautier, Platform Proteom'IC, Institut Cochin, Dr Franck Letourneur, Platform Genom'IC, Institut Cochin, Ms Stella Hartono, Department of Molecular and Cellular Biology and Genome Center, University of California, Davis, Dr. Ivan Nemazanyy, Platform for Metabolic Analyses (INSERM US24/CNRS UAR 3633, Necker, Paris, France) for their expertise, Ms Katherine Wetmore, Ms Ania Alik, Mrs Angélique Marcon, Mrs Marlène Dejean, Ms Camille Knops, Ms Laïla Zarouli, Ms Alice Rousseau (Laboratory of Hematology, Cochin Hospital, Paris) for technical help. **Grants:** This work was funded by the Institut National du Cancer INCa PLBio 2015-129 (M.F., M-H. S., A. C.), the Fondation pour la Recherche Médicale Equipe labellisée FRM202003010191 (M.F.), the Laboratoire d'Excellence GR-Ex and the MDS-RIGHT European Union's Horizon 2020 research and innovation programme under grant 634789 (P.F., M.F.).

## Author contributions

T.R., S.Bo., A.L. and R.M.M. contributed equally to this study. M.F., M-H.S., A.C., F.C., B.M., B.P. designed the study. D.R., C.L., B.F., A.L., N.D., M.W., R.M. developed the methods. C.L., T.R., S.Bo., R.M.M., B.F., A.L-P, A.L., S.Ba., M.DC., D.C-G., F.L., A.T., C.F. performed the experiments. C.F., M.T., L.A., E.C., P.F., S.R., T.C., S.P., L.W., D.B., N.C., O.K. acquired and managed patient clinical and biological data. M.L., M.LG., N.D. provided facilities. M.W., S.A., R.M. provided tools. D.R., I.B., A.H. performed computational analyses. D.R., C.L., A.L., B.M., S.Bo., B.F., F.C., D.C-G., M.F. analyzed the data and performed biostatistical analyses. M.F., D.R. wrote the manuscript. All authors reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-46547-7>.

**Correspondence** and requests for materials should be addressed to Michaela Fontenay.

**Peer review information** *Nature Communications* thanks Stephanie Halene and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://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 4.0 International License, which permits use, sharing, adaptation, 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 changes were made. 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/4.0/>.

© The Author(s) 2024

---

<sup>1</sup>Université Paris Cité, Centre National de la Recherche Scientifique, Institut National de la Santé et de la Recherche Médicale, Institut Cochin, Paris, France. <sup>2</sup>Equipe labellisée par la Fondation pour la Recherche Médicale, Paris, France. <sup>3</sup>Laboratoire d'excellence du Globule Rouge GR-Ex, Université Paris Cité, Paris, France. <sup>4</sup>Assistance Publique-Hôpitaux de Paris.Centre-Université Paris Cité, Hôpital Cochin, Laboratory of Hematology, Paris, France. <sup>5</sup>Université Paris Cité, CNRS, Institut Jacques Monod, Paris, France. <sup>6</sup>Department of Molecular and Cellular Biology and Genome Center, University of California, Davis, CA, USA. <sup>7</sup>Platform Proteom'IC, Université Paris Cité, Institut Cochin, Paris, France. <sup>8</sup>Institut Curie, PSL Research University, Sorbonne University, INSERM U830, DNA repair and uveal melanoma, Equipe labellisée par la Ligue Nationale contre le Cancer, Paris, France. <sup>9</sup>Assistance Publique-Hôpitaux de Paris.Centre-Université Paris Cité, Hôpital Cochin, Clinical Department of Hematology, Paris, France. <sup>10</sup>Department of Hematology, Centre Hospitalier Universitaire, Université de Grenoble Alpes, Grenoble, France. <sup>11</sup>Laboratory of Hematology, Université Côte d'Azur, Centre Hospitalier Universitaire, Nice, France. <sup>12</sup>Clinical Department of Hematology, Université Côte d'Azur, Centre Hospitalier Universitaire, Nice, France. <sup>13</sup>Assistance Publique-Hôpitaux de Paris.Nord-Université Paris Cité, Saint-Louis Hospital, Laboratory of Hematology, Paris, France. <sup>14</sup>Assistance Publique-Hôpitaux de Paris.Nord-Université Paris Cité, Saint-Louis Hospital, Service Hématologie Séniors, Paris, France. <sup>15</sup>Institut Curie, Paris Sciences Lettres Research University, Sorbonne University, INSERM U934, UMR3215 Paris, France. <sup>16</sup>Institut Gustave Roussy, INSERM 1287, Université Paris Saclay, Villejuif, France. <sup>17</sup>Institut de Génétique Humaine, Centre National de la Recherche Scientifique, Université de Montpellier, Montpellier, France. <sup>18</sup>These authors contributed equally: David Rombaut, Carine Lefèvre. <sup>19</sup>Deceased: Angelos Constantinou. ✉e-mail: [michaela.fontenay@inserm.fr](mailto:michaela.fontenay@inserm.fr)