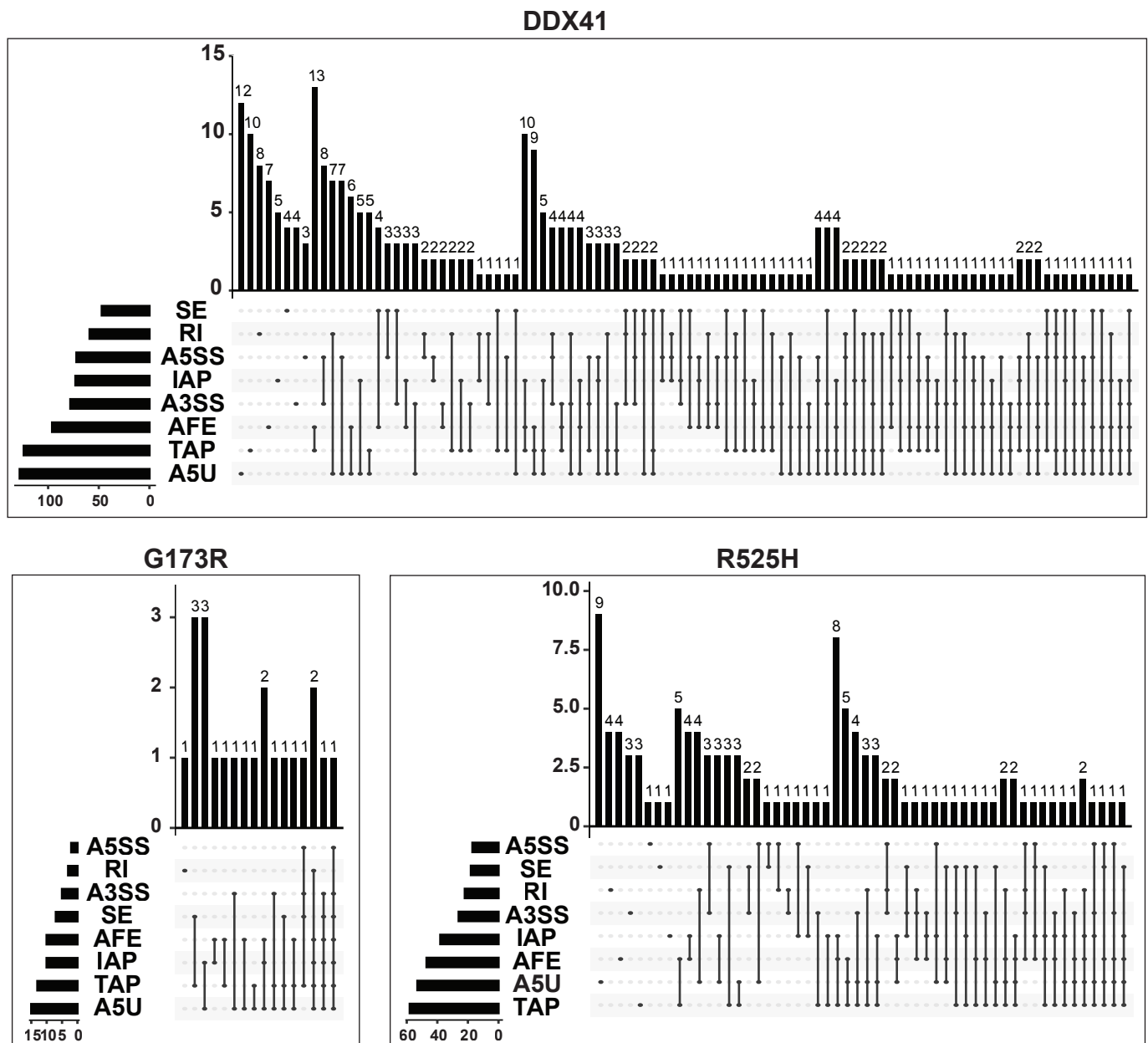


Supplementary information

Oncogenic DEAD-box ATPase DDX41 establishes transcript ensembles via CLK3-dependent and -independent mechanisms

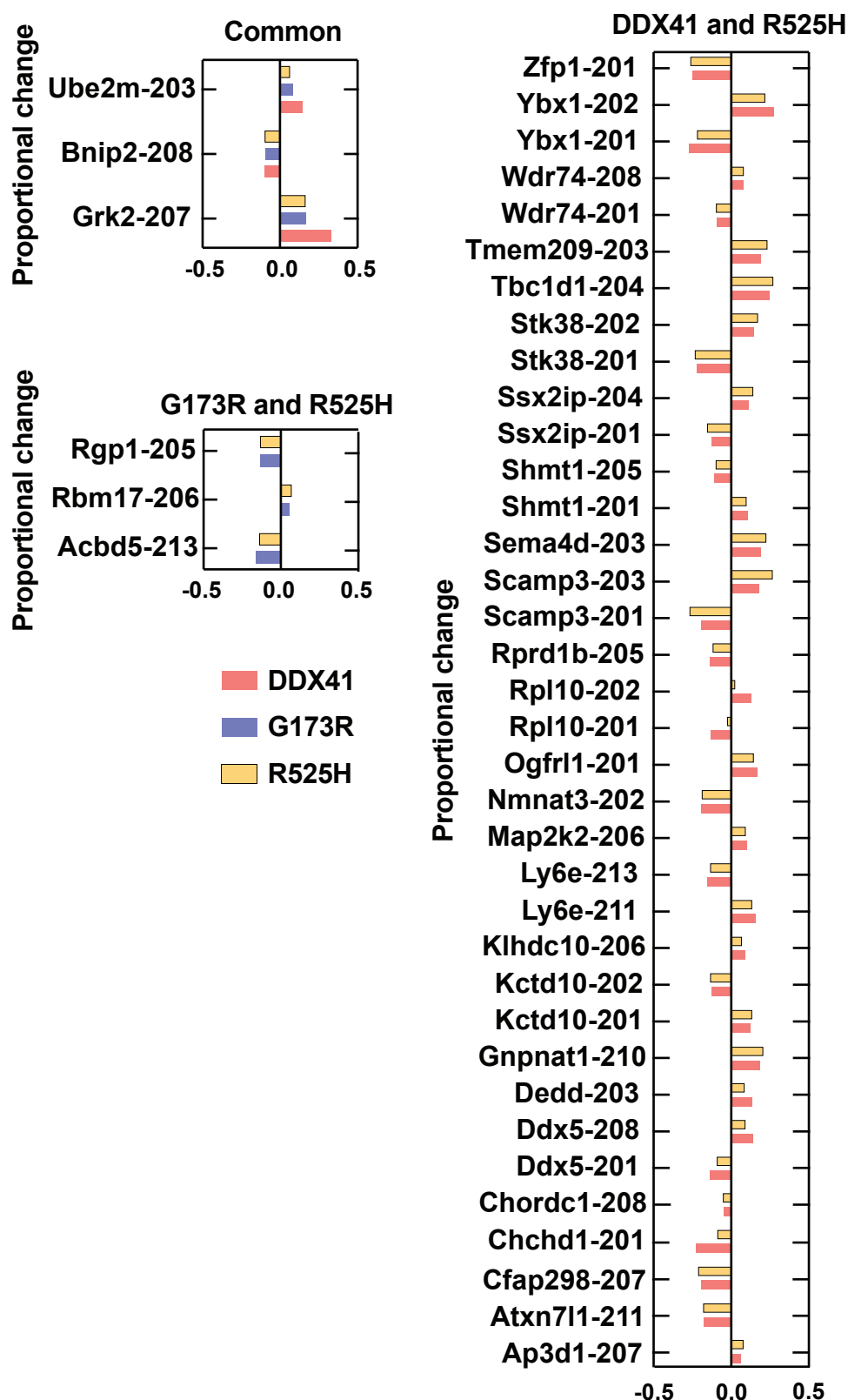
Jeong-Ah Kim, Siqi Shen, Christina M. Jurotich, Saritha S. D'Souza, Johnathan J. Caldon,
Igor I. Slukvin, Jane E. Churpek, Sunduz Keles, and Emery H. Bresnick



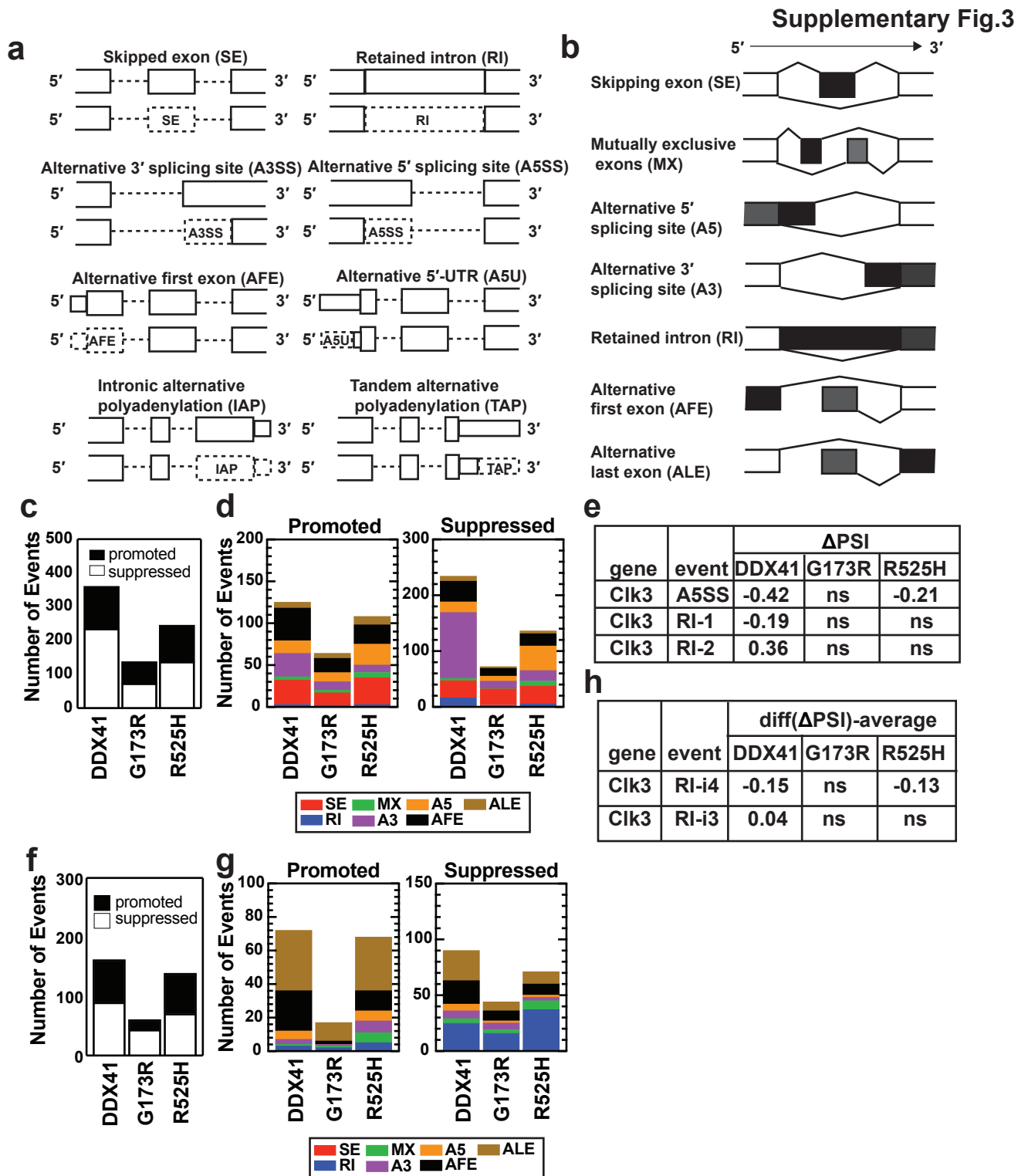
Supplementary Fig.1 The combination of alternative transcript regulation by DDX41, G173R, and R525H variants in the genetic rescue system with hi-*Ddx41*^{+/-} cells.

Related to Fig.1 and 2.

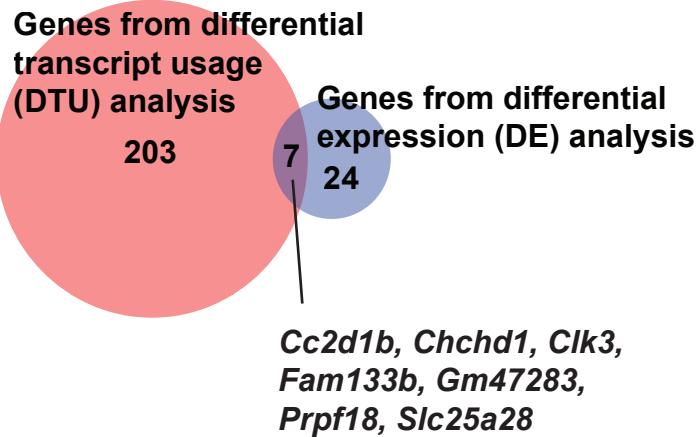
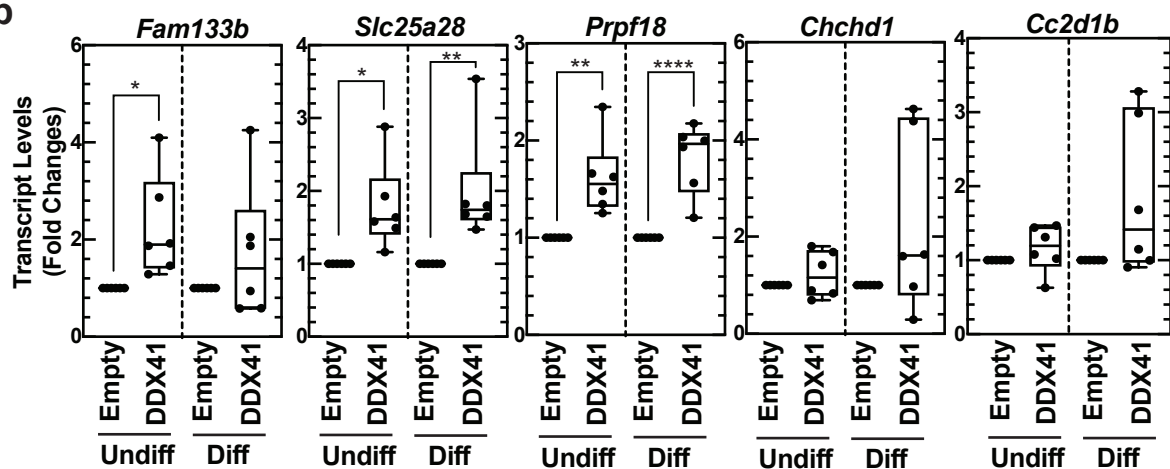
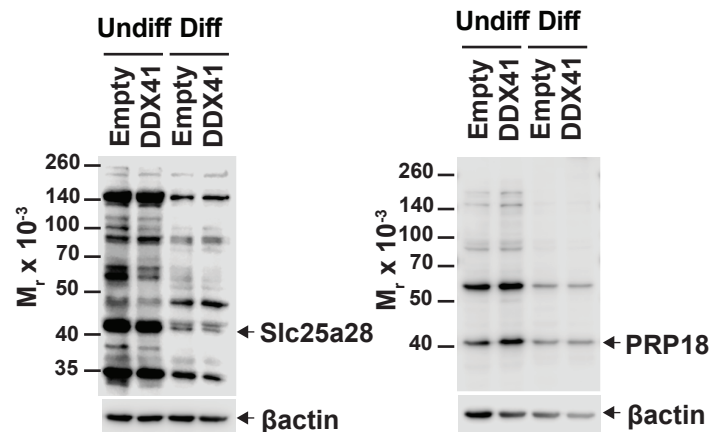
Alternative 3' splicing site; A3SS, alternative 5' splicing site; A5SS, intron retention; RI,skipped exon; SE, alternative 5'UTR; A5U, alternative first exon; AFE, alternative intronic polyadenylation; IAP, alternative tandem polyadenylation; TAP.



Supplementary Fig.2 Comparison of proportional change of transcripts regulated by DDX41 vs. variants. Proportional change of DDX41-regulated transcript cohorts (commonly regulated by: G173R/R525H, DDX41/R525H, and DDX41/G173R/R525H) in Fig.2j.

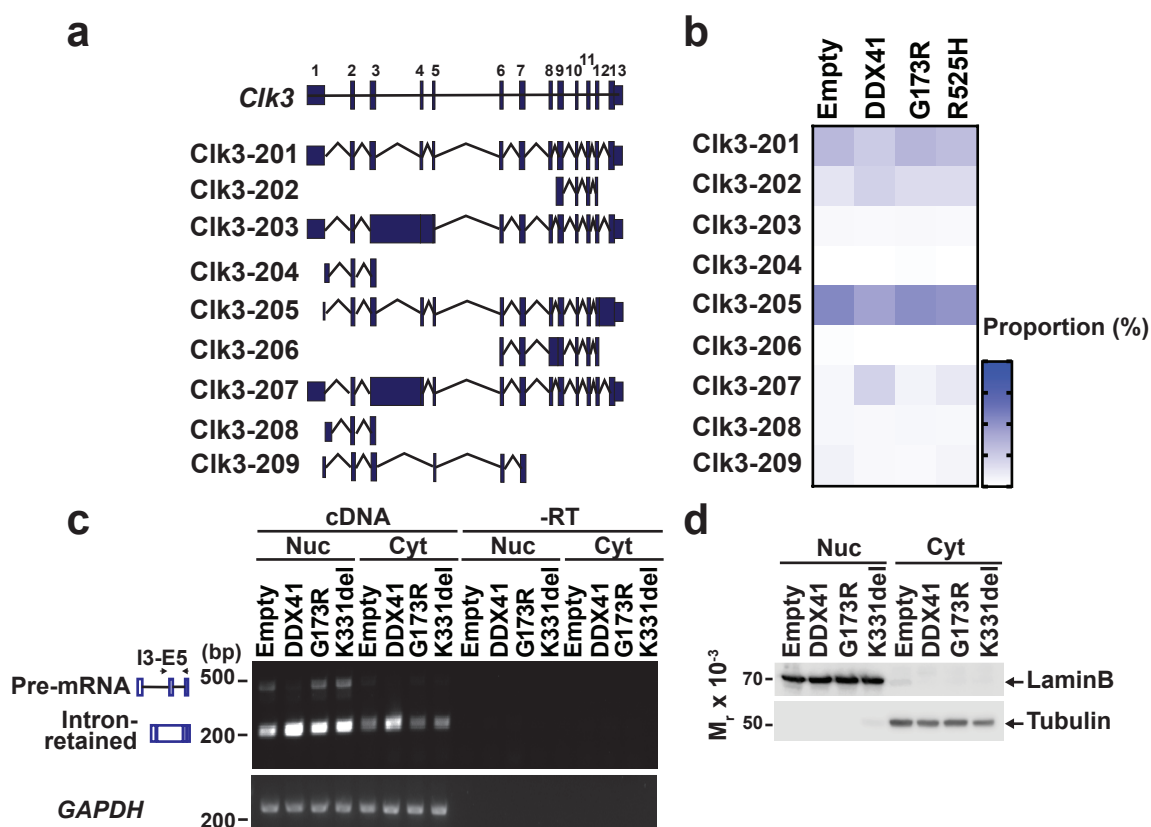


Supplementary Fig.3 Alternative splicing analysis using SUPPA and MISO. Related to Fig.2 and discussion. **a** and **b** Types of alternative splicing events in SURF (**a**) and SUPPA (**b**). The SUPPA alternative splicing event annotation was also used in MISO analysis. Analysis of DDX41- or variants-mediated alternative splicing events by SUPPA (**c-e**) and MISO (**f-h**). Total number of events (**c** and **f**) and number of each events (**d** and **g**) promoted or repressed by DDX41 and variants were analyzed (cutoffs; $p < 0.05$, $|\Delta\text{PSI}| > 10\%$ for SUPPA and Bayes factor > 10 , $|\Delta\text{PSI}| > 10\%$ for MISO). **e** and **h** Significantly altered events ($p < 0.05$) and its ΔPSI on *Clk3* transcripts by DDX41 and variants.

a**b****c**

Supplementary Fig.4 Expression of DDX41-mediated differentially regulated gene and transcripts. Related to Fig.3.

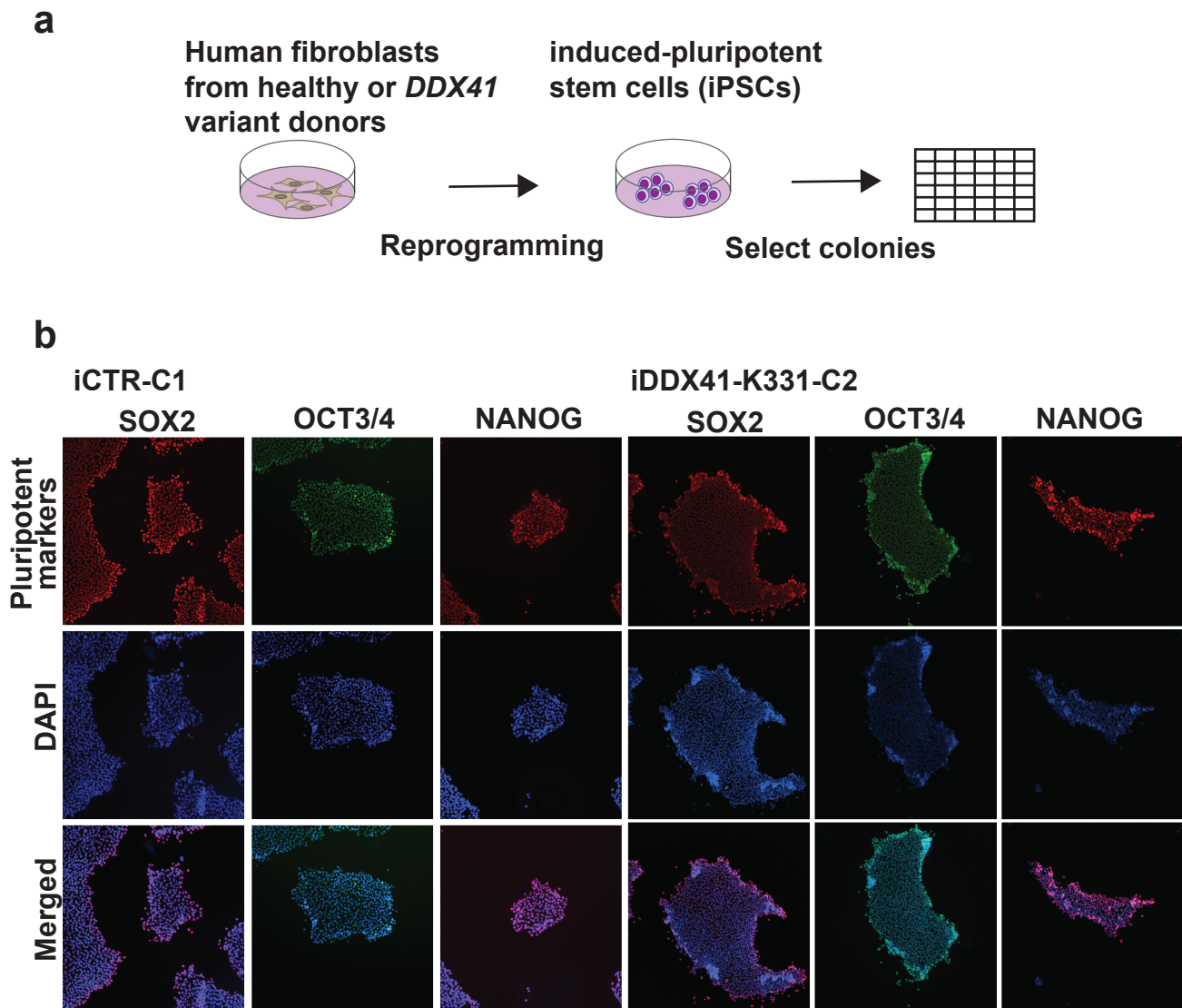
a Comparison of genes from differential transcript usage (DTU) analysis and differential expressed gene (DEG) analysis of DDX41-regulated transcripts in hi-*Ddx41*^{+/-} cells. **b** mRNA levels of protein coding genes regulated by DDX41 rescue in hi-*Ddx41*^{+/-} cells using RT-qPCR. **c** SLC25A28 and PRP18 level by DDX41 rescue in hi-*Ddx41*^{+/-} cells using anti-SLC25A28 (invitrogen, PA5-107261) and anti-PRP18 (Novus Biologicals, NBP1-92297). Exact p-values were supplied in Source Data file.



Supplementary Fig.5 DDX41 regulates *Clk3* differential transcript usage in *Ddx41*^{+/-} myeloid progenitor cells.

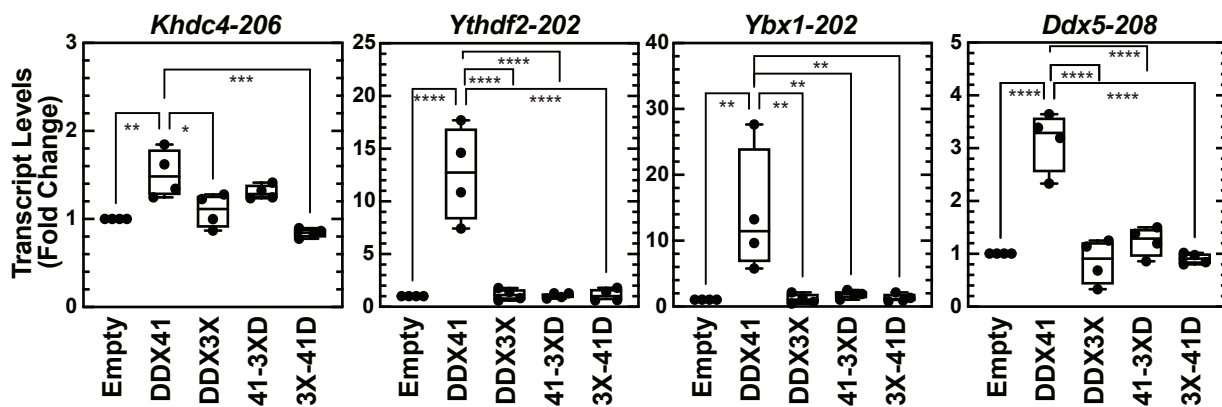
Related to Fig.3.

a Genomic structure of *Clk3* transcripts (ENSMUSG00000032316 in GRCm39). **b** Heatmap with proportional ratio to total transcripts of *Clk3* transcriptional isoforms. **c** Detection of intron-retained *Clk3* transcript with cDNA and -RT control from nuclear and cytoplasmic fractions in the rescue assay with DDX41 and variants by RT-qPCR. **d** Western blotting of LaminB (abcam, ab229025) and Tubulin (Cell Signaling Technology, 3873T) to assess purity of nuclear and cytoplasmic cell fractions.



Supplementary Fig.6 Generation of induced-pluripotent stem cells (iPSCs).
Related to Fig.3.

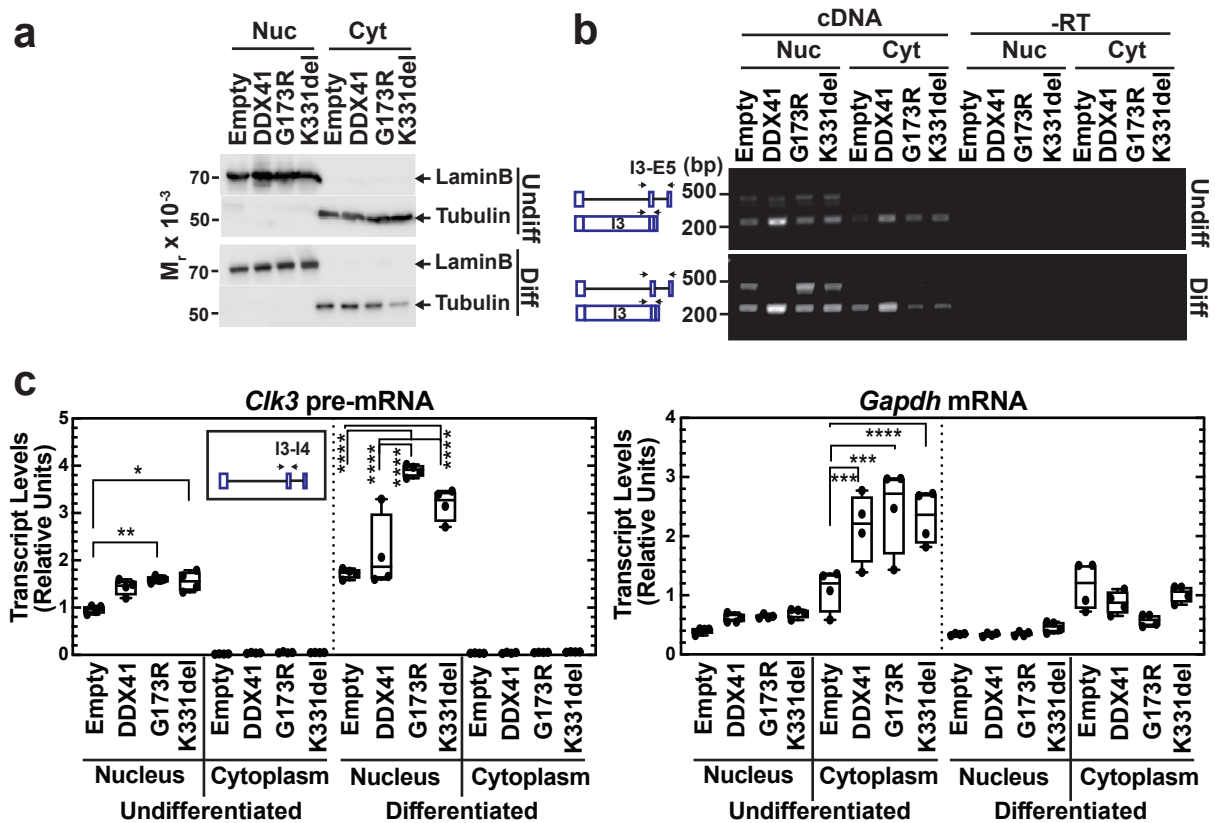
a iPSCs generation. **b** Immunofluorescence using pluripotent markers, SOX2, OCT3/4, and NANOG. iPSC colonies (iCTR-C1; *DDX41*, iDDX41-K331-C2; *DDX41* K331del) were incubated with primary antibodies, including anti-human OCT3/4-clone C10 (SantaCruz), anti-human Nanog-clone D73G4 (Cell signaling), anti-human SOX2-clone D6D9 (Cell signaling). After PBS wash, colonies were stained with secondary antibodies, including donkey anti-rabbit alexa fluor 488 (Life Technologies) or donkey anti-mouse alexa fluor 568 (Life Technologies).



Supplementary Fig.7 Quantification of transcripts in rescue assay with DDX41, DDX3X, 41-3XD (chimera substituted with DDX3X DEAD-box domain), or 3X-41D (chimera substituted with DDX41 DEAD-box domain) by RT-qPCR (n=4).

Related to Fig.5.

Result was presented as box-and-whisker plots with bounds from the 25th to the 75th percentiles and the median line, and whiskers ranging from minimum to maximum values. Points represent each replicate. Statistics were calculated using one-way ANOVA, followed by Šídák test; *, $P \leq 0.01$; **, $P \leq 0.01$; ***, $P \leq 0.001$; ****, $P \leq 0.0001$. n= biological replicates. Exact p-values were supplied in Source Data file.

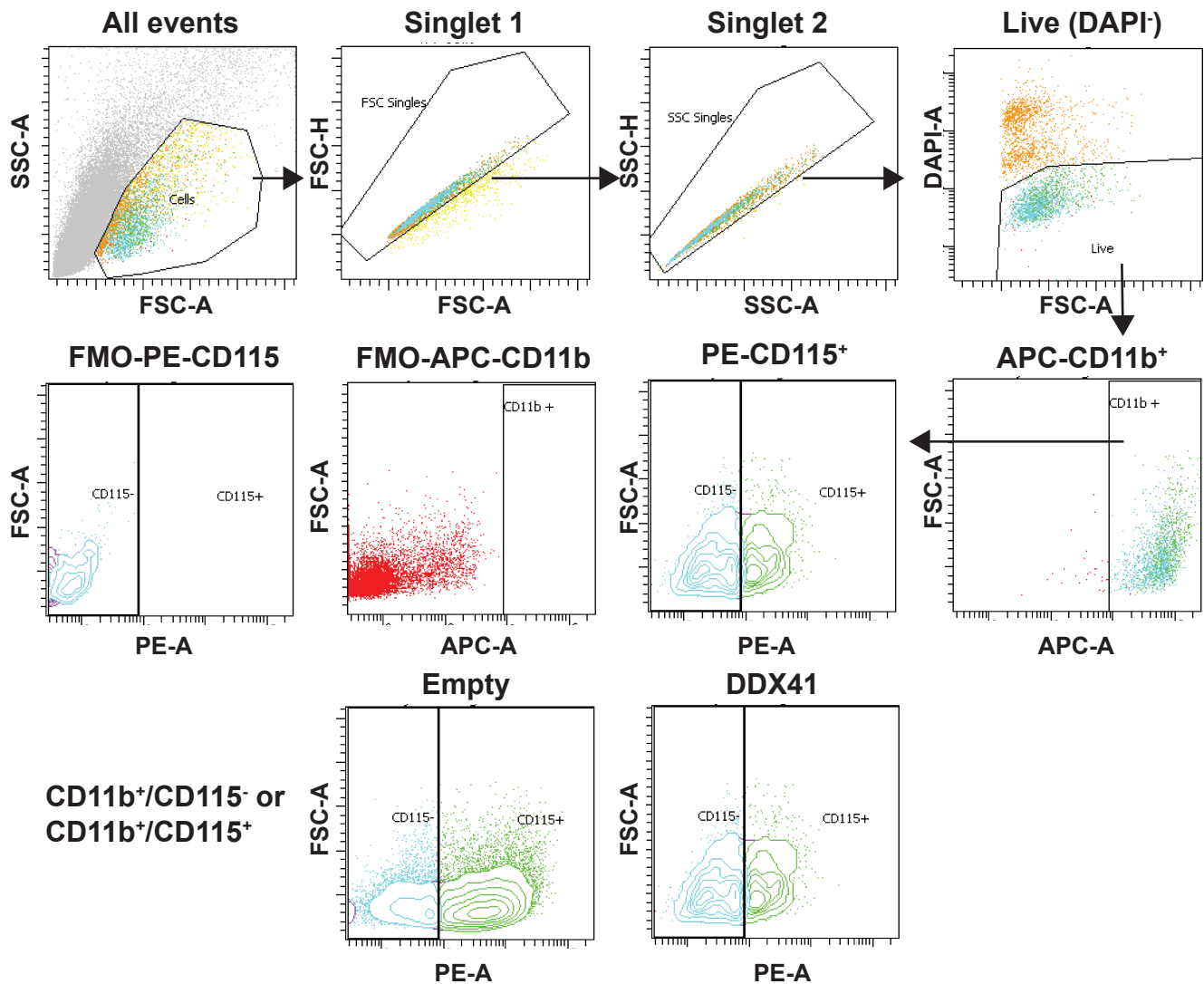


Supplementary Fig.8 Assessment of nuclear and cytoplasmic of *C/kl3* transcripts and *Gapdh* mRNAs in undifferentiated cells and upon differentiation

Related to Fig.6.

a Western blotting analysis of LaminB and Tubulin in nuclear and cytoplasmic fractions. **b,c** hi-*Ddx41*^{+/-} cells were infected with GFP or GFP-DDX41 and variants expressing retrovirus. GFP-positive cells were sorted after three days and cultured with or without G418 and β -estradiol for three days. Nuclear and cytoplasmic transcripts were analyzed by RT-PCR using indicated primers. (**b**) or RT-qPCR (**c**). qPCR primers of *C/kl3* pre-mRNA are indicated in the integrated box in the graph. Quantification of RT-qPCR were presented as box-and-whisker plots with bounds from the 25th to the 75th percentiles and the median line, and whiskers ranging from minimum to maximum values. Points represent each replicate. Statistics were calculated using one-way ANOVA, followed by Šídák test; *, $P \leq 0.01$; **, $P \leq 0.01$; ***, $P \leq 0.001$; ****, $P \leq 0.0001$. Exact p-values were supplied in Source Data file.

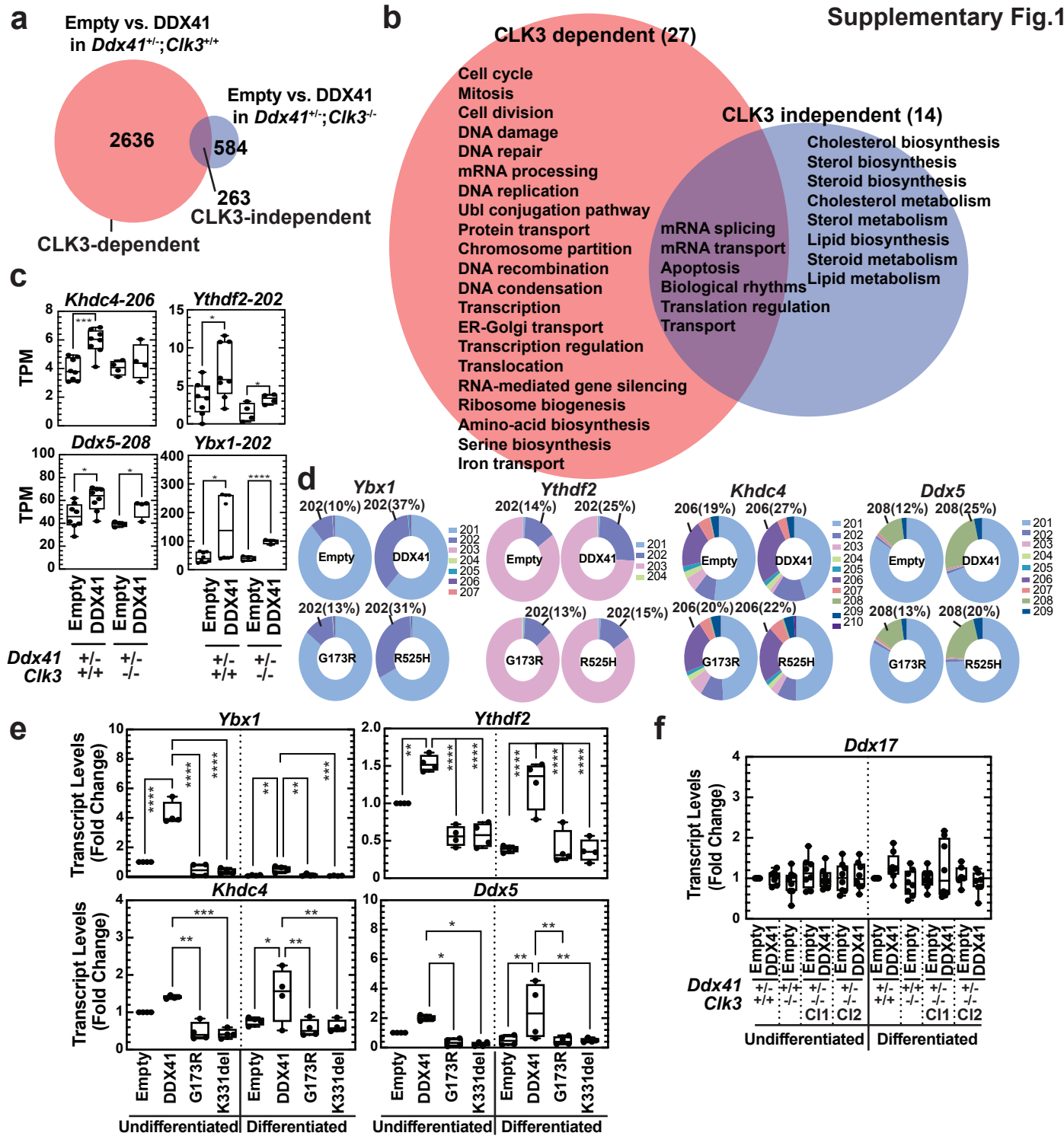
Supplementary Fig.9



Supplementary Fig.9 Cell sorting strategy for differentiated hi-*Ddx41*^{+/-} cells using flow cytometry.

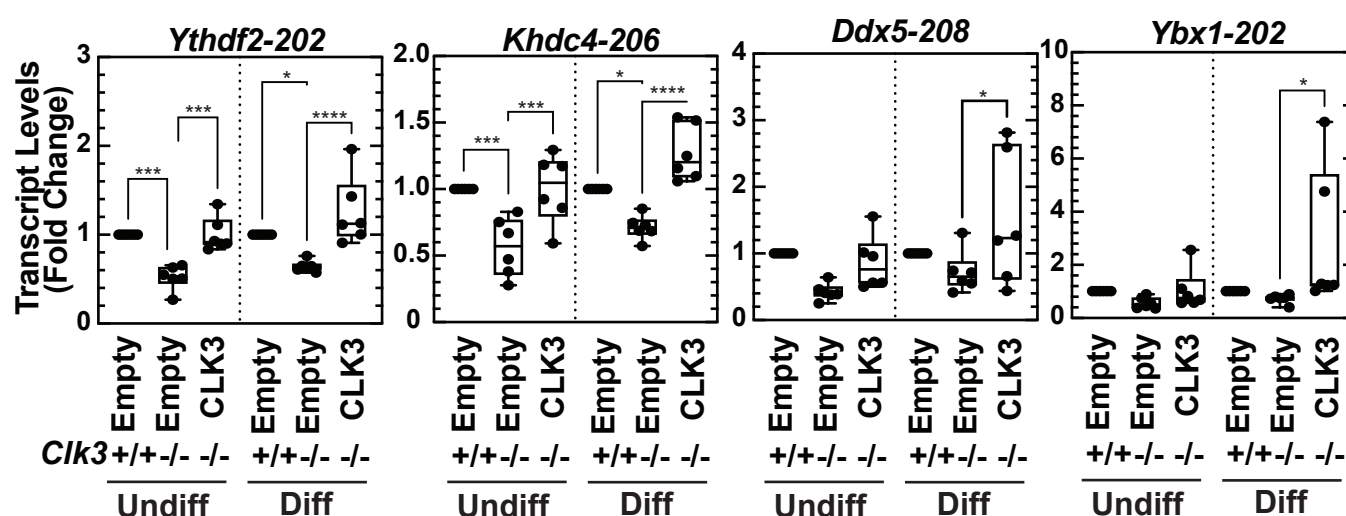
Related to Fig.6.

hi-*Ddx41*^{+/-} cells were infected with GFP or GFP-DDX41 expressing retrovirus. GFP⁺ cells were sorted after three days and cultured with or without G418 and β -estradiol for three days. Live CD11b⁺/CD115⁻ (granulocytic) and CD11b⁺/CD115⁺ (monocytic) cells were sorted by the above gating strategy. Differentiated cell populations were analyzed using FlowJo v10.8.0 software (BD Biosciences).



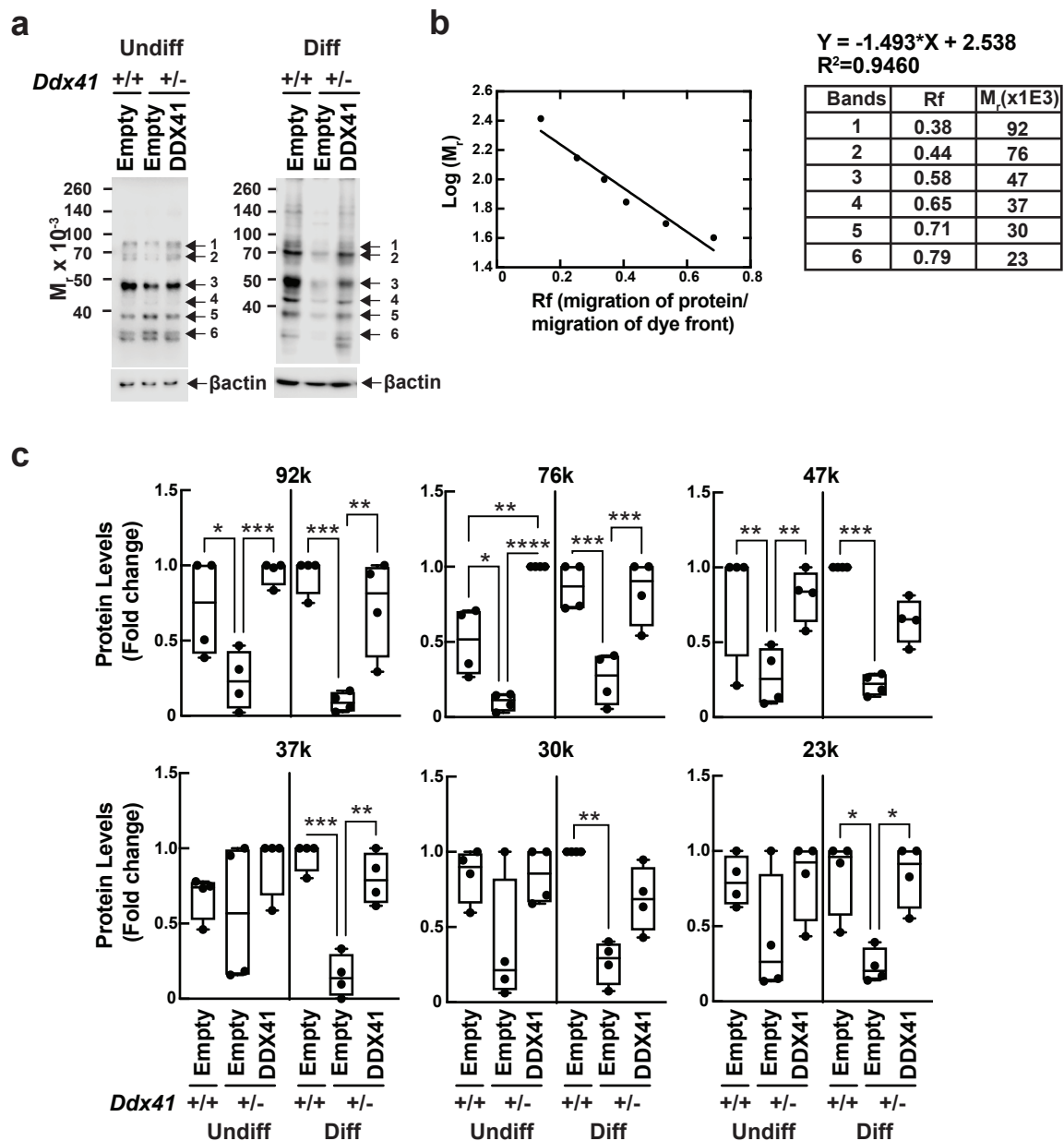
Supplementary Fig.10 CLK3 dependency in DDX41-mediated RNA transcript isoforms in myeloid progenitor cells. Related to Fig.7

a Comparison of differentially expressed transcripts by DDX41 expression in *Clk3^{+/-};Ddx41^{+/-}* and *Clk3^{-/-};Ddx41^{+/-}* cells. 2899 transcripts were altered by DDX41 in *Clk3^{+/-};Ddx41^{+/-}*, among them, 263 (10%, CLK3-independent) transcripts were also regulated in *Clk3^{-/-};Ddx41^{+/-}* cells. **b** Biological process analysis of transcripts using DAVID. DDX41-regulated transcripts related mRNA splicing (CLK3-dependent; $p=1.70E^{-02}$, CLK3-independent; $p=1.60E^{-06}$) and mRNA transport (CLK3-dependent; $p=4.90E^{-02}$, CLK3-independent; $p=3.20E^{-02}$). **c** TPM values of *Khdc4-206*, *Ythdf2-202*, *Ddx5-208*, and *Ybx1-202* transcripts in *Clk3^{+/-};Ddx41^{+/-}* ($n=8$) and *Clk3^{-/-};Ddx41^{+/-}* ($n=4$) with or without DDX41. **d** Pie charts depicting the proportion of transcripts in the total transcript pool of *Ybx1*, *Ythdf2*, *Khdc4*, and *Ddx5* using TPM values. **e** RT-qPCR quantification of transcripts in genetic rescue assay with DDX41, G173R, and K331del in undifferentiated and differentiated cells ($n=4$). **f** RT-qPCR quantification of *Ddx17* regulated by DDX41 in *Clk3^{+/-};Ddx41^{+/-}* and *Clk3^{-/-};Ddx41^{+/-}* ($n=4$). Statistics were calculated using one-way ANOVA, followed by Šidák test; *, $P \leq 0.01$; **, $P \leq 0.01$; ***, $P \leq 0.001$; ****, $P \leq 0.0001$. n = biological replicates. Exact p-values were supplied in Source Data file.



Supplementary Fig.11 Quantification of transcripts regulated by CLK3 expression *Clk3*^{-/-} hi-cells upon undifferentiated and differentiated states using RT-qPCR (n=6). Related to Fig.7.

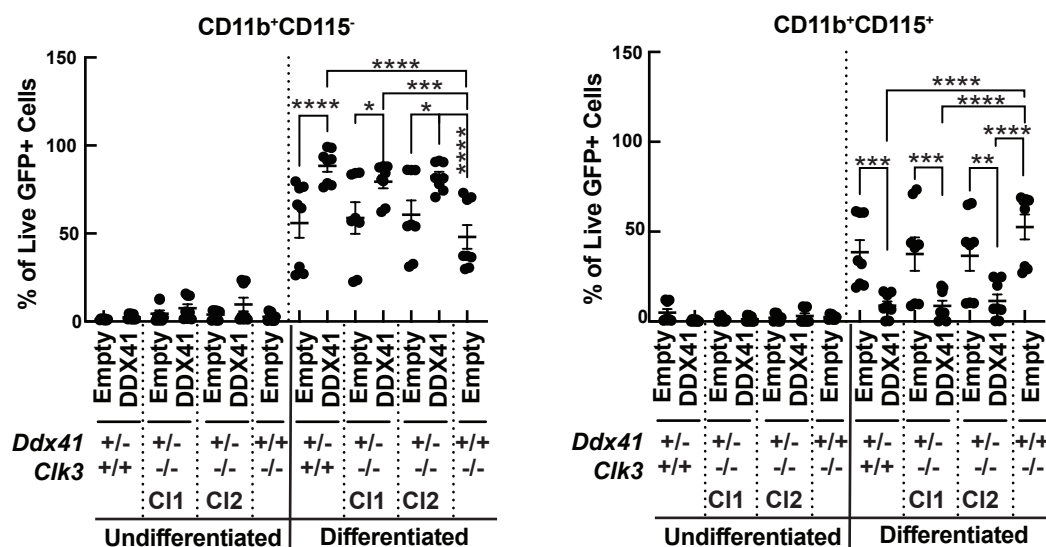
Result was presented as box-and-whisker plots with bounds from the 25th to the 75th percentiles and the median line, and whiskers ranging from minimum to maximum values. Points represent each replicate. Statistics were calculated using one-way ANOVA, followed by Šídák test; *, $P \leq 0.01$; **, $P \leq 0.01$; ***, $P \leq 0.001$; ****, $P \leq 0.0001$. n= biological replicates. Exact p-values were supplied in Source Data file.



Supplementary Fig.12 Quantification of phosphorylation of SR proteins by DDX41 expression using retroviral infection in *Ddx41*^{+/-} upon undifferentiated and differentiated states (n=4).

Related to Fig.7.

a PhosphoSR proteins were detected by western blotting using anti-phosphoSR (1H4). **b** Molecular weights of 6 obvious bands were calculated by a semi-log plot based on the migration of protein/migration of dye front (Rf). **c** Quantified protein levels were presented as box-and-whisker plots with bounds from the 25th to the 75th percentiles and the median line, and whiskers ranging from minimum to maximum values. Points represent each replicate. Statistics were calculated using one-way ANOVA, followed by Šídák test; *, $P \leq 0.01$; **, $P \leq 0.01$; ***, $P \leq 0.001$; ****, $P \leq 0.0001$. n= biological replicates. Exact p-values were supplied in Source Data file.



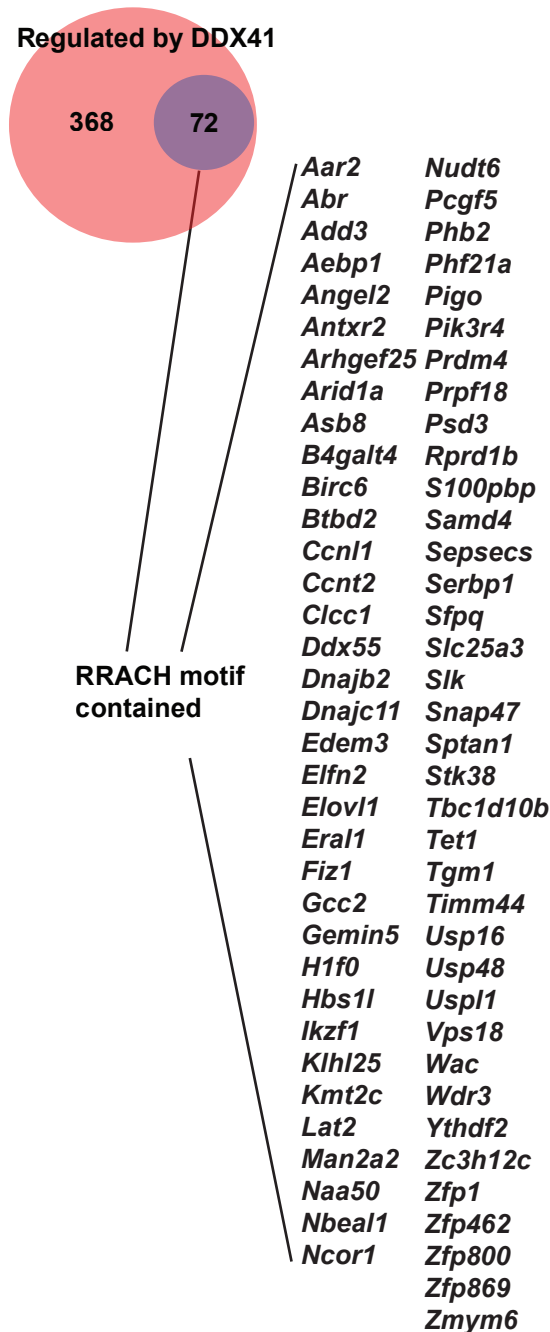
Supplementary Fig.13 Flow cytometry analysis of granulocytic and monocytic cells. Related to Fig.7.

hi-*Ddx41*^{+/-}, two *Ddx41*^{+/-};*Clk3*^{-/-} clones (clone1; C11 and clone2; C12), and *Clk3*^{-/-} cells were infected with GFP or GFP-DDX41-Flag expressing retrovirus. GFP⁺ cells were sorted and cultured with or without G418 and β -estradiol for three days. Cells were analyzed with CD11b and CD115 surface markers.

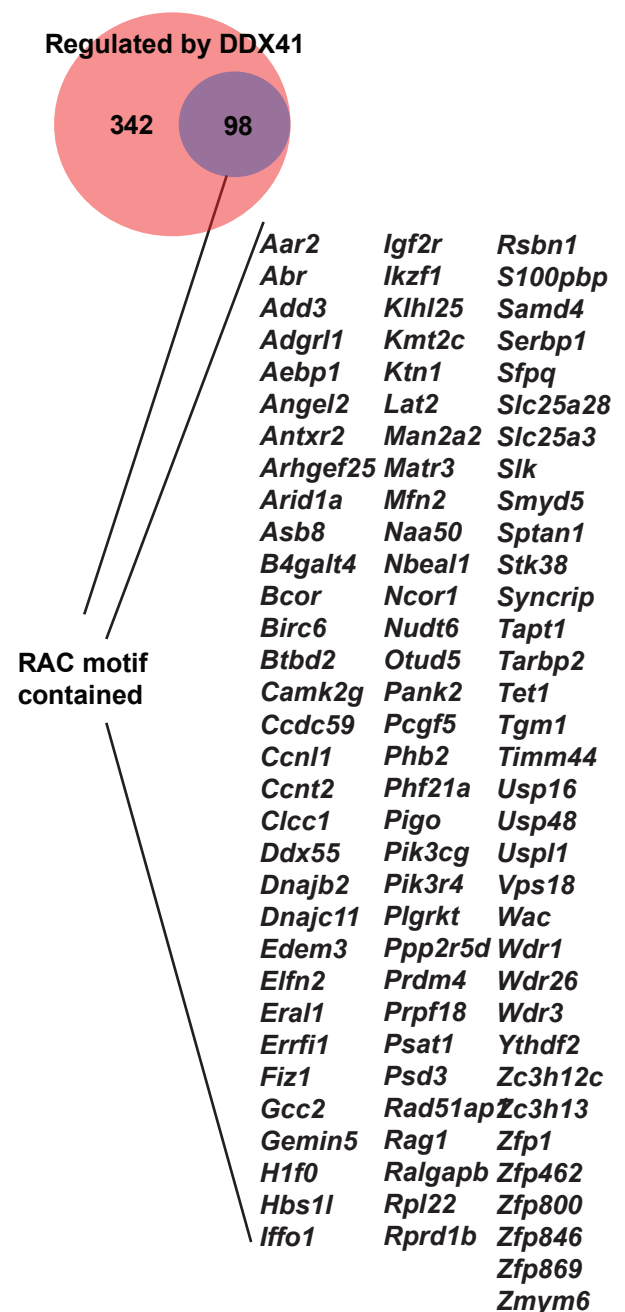
Differentiated cell populations were analyzed using FlowJo v10.8.0 software (BD Biosciences). Quantitative flow analysis results were presented as the mean $\pm$ SEM. Statistics were calculated using one-way ANOVA, followed by Šídák test; *, $P \leq 0.01$; **, $P \leq 0.01$; ***, $P \leq 0.001$; ****, $P \leq 0.0001$. (n=8). Exact p-values were supplied in Source Data file.

a

Predicted m6A motif (RRACH)
(probability >0.7)

**b**

Predicted m6A motif (RAC)
(probability >0.7)



Supplementary Fig.14 Predicted m⁶A motifs, RRACH or RAC (R = A or G, H = A, C, or U) in sequences of genes corresponding to transcripts regulated by DDX41 using the genome-wide motif dataset from im6A modeling by Luo et al (2022). Among the genes regulated by DDX41, 72 genes (16%) contain the RRACH motif (**a**), and 98 genes (22%) contain the RAC motif (**b**). Prediction probability >0.7.