The RNA helicase DDX39B activates FOXP3 RNA splicing to control T regulatory cell fate

Minato Hirano[1,2,†], Gaddiel Galarza-Muñoz[1,3,†], Geraldine Schott[1,§], Liuyang Wang[4], Chloe Nagasawa[1,5], Alejandro L. Antonia[4], Vaibhav Jain[6], Xiaoying Yu[7], Steven G. Widen[1], Farren B.S. Briggs[8], Simon G. Gregory[4,6,9], Dennis C. Ko[4,10], W. Samuel Fagg[1,11], Shelton S. Bradrick[1,12,‡,*], and Mariano A. Garcia-Blanco[1,12,13,*,§§].

[1]Department of Biochemistry and Molecular Biology, University of Texas Medical Branch, Galveston, Texas, USA.

[2]National Research Center for the Control ad Prevention of Infectious Disease, Nagasaki University, Nagasaki, Japan.

[3]Autoimmunity Biologic Solutions, 2200 Market Street, Galveston, Texas, USA.

[4]Department of Molecular Genetics and Microbiology, Duke University, Durham, North Carolina, USA.

[5]Human Pathophysiology and Translational Medicine Program, Institute for Translational Sciences, University of Texas Medical Branch, Galveston, Texas, USA.

[6]Duke Molecular Physiology Institute, Duke University, Durham, North Carolina, USA.

[7]Department of Preventive Medicine and Population Health, University of Texas Medical Branch, Galveston, Texas, USA.

[8]Department of Population and Quantitative Health Sciences, School of Medicine, Case Western Reserve University, Cleveland, Ohio, USA.

[9]Department of Neurology, Duke University, Durham, North Carolina, USA.

[10]Division of Infectious Diseases, Department of Medicine, Duke University, Durham, North Carolina, USA.

[11]Transplant Division, Department of Surgery, University of Texas Medical Branch, Galveston, Texas, USA.

[12] Institute of Human Infections and Immunity, University of Texas Medical Branch, Galveston, Texas, USA.

[13] Department of Internal Medicine, University of Texas Medical Branch, Galveston, Texas, USA.

*Correspondence to: Mariano A. Garcia-Blanco (maragarc@utmb.edu) or Shelton S. Bradrick (ssbradri@utmb.edu).

[§§]Lead contact.

[†]These authors contributed equally.

[§]Present address: 2654 Quivira Trce, League City, Texas, USA

38 ‡Present address: Infectious Diseases, Surveillance & Diagnostics Division, MRIGlobal, 425
39 Volker Blvd, Kansas City, MO 64110.
40

Summary: Genes associated with increased susceptibility to multiple sclerosis (MS) have been identified, but their functions are incompletely understood. One of these genes codes for the RNA helicase DExD/H-Box Polypeptide 39B (DDX39B), which shows genetic and functional epistasis with interleukin-7 receptor-α gene (*IL7R*) in MS-risk. Based on evolutionary and functional arguments we postulated that DDX39B enhances immune tolerance decreasing MS risk. Consistent with such a role we show that DDX39B controls the expression of many MS susceptibility genes and important immune-related genes. Among these we identified *Forkhead Box P3* (*FOXP3*), which codes for the master transcriptional factor in $CD4^+/CD25^+$ T regulatory cells. DDX39B knockdown led to loss of immune-regulatory and gain of immune-effector expression signatures. Splicing of *FOXP3* introns, which belong to a previously unrecognized subclass of introns with C-rich polypyrimidine tracts, was exquisitely sensitive to DDX39B levels. Given the importance of FOXP3 in autoimmunity, this work cements DDX39B as an important guardian of immune tolerance.

Keywords: Multiple Sclerosis, Autoimmunity, RNA helicase, DDX39B, Splicing, FOXP3, T regulatory cells

Autoimmune diseases are caused by a combination of environmental and genetic factors. Genetic factors associated with increased susceptibility to multiple sclerosis (MS), an autoimmune disease of the central nervous system, have been identified, but their mechanisms of action are incompletely understood (Briggs, 2019; International Multiple Sclerosis Genetics, 2019). We established that the association between MS risk and the interleukin-7 receptor-α gene (*IL7R*) is mediated by alternative splicing of *IL7R* transcripts (Gregory et al., 2007). The disease associated allele of the single nucleotide polymorphism (SNP) rs6897932 in exon 6 of IL7R strengthens an exonic splicing silencer (ESS) and increases skipping of exon 6 leading to increased production of soluble IL7R (sIL7R) (Evsyukova et al., 2013; Gregory et al., 2007; Schott et al., 2021). Elevated levels of sIL7R have been shown to exacerbate the inducible murine MS-like model experimental autoimmune encephalomyelitis (EAE) and are proposed to increase the bioavailability of IL7 (Lundstrom et al., 2013). Given the importance of exon 6 splicing we identified RNA binding proteins that bind it and impact its splicing, and among these identified the RNA helicase DExD/H-Box Polypeptide 39B (DDX39B) (Evsyukova et al., 2013; Galarza-Munoz et al., 2017).

DDX39B, known to RNA biologists as UAP56, plays roles in RNA splicing and nucleocytoplasmic transport (Shen, 2009), but was first discovered by immunologists, who named it BAT1 (Spies et al., 1989) and linked it to autoimmune diseases (Degli-Esposti et al., 1992). We connected the RNA biology and immunology roles of DDX39B showing that it activates splicing of exon 6 and represses production of sIL7R (Galarza-Munoz et al., 2017). Furthermore, *DDX39B* shows genetic and functional epistasis with *IL7R* in enhancing MS risk (Galarza-Munoz et al., 2017). Based on these studies we proposed that DDX39B plays protective role in MS and other autoimmune diseases (Galarza-Munoz et al., 2017). The alternative splicing

of IL7R exon 6 is conserved only in primates, while *DDX39B* is found in the major histocompatibility complex (MHC) class III region in all vertebrates (Schott and Garcia-Blanco, 2021), suggesting that DDX39B controls important immune modulators other than IL7R.

Here we show that DDX39B controls expression of gene products involved in autoimmunity including Forkhead Box P3 (FOXP3), a master regulator of the development, maintenance and function of $CD4^+/CD25^+$ regulatory Tcells (Tregs) (Georgiev et al., 2019; Hori, 2021; Josefowicz et al., 2012) and a repressor of autoimmune diseases (Bennett et al., 2001; Brunkow et al., 2001; Chatila et al., 2000; Wildin et al., 2001). Splicing of FOXP3 introns, which belong to a previously unknown subclass of introns with C-rich polypyrimidine tracts, is strongly dependent on DDX39B, making FOXP3 expression highly sensitive to the levels of this RNA helicase.

DDX39B controls expression of genes associated with multiple sclerosis risk. We had previously knocked down DDX39B with two independent shRNAs (Sh3 and Sh5) in primary human $CD4^+$ T cells from six healthy human donors and shown that DDX39B depletion led to increased skipping of IL7R exon 6 (Galarza-Munoz et al., 2017). To identify other RNAs affected by DDX39B in these immune relevant cells, we carried out RNA sequencing (RNAseq) of polyadenylated RNA from control and DDX39B depleted $CD4^+$ T cells from two of these donors (donors 1 and 4). More transcript level changes were detected with Sh3 treatment than with Sh5 treatment (Fig. S1A, Table S1), which was consistent with Sh3 treatment reducing DDX39B RNA levels more profoundly (Table S1) and DDX39B protein levels more consistently (Galarza-Munoz et al., 2017). Therefore, analysis of transcriptome alterations were first carried out by comparing cells depleted of DDX39B using Sh3 to cells treated with a control

shRNA; however, all results were confirmed with both shRNAs and in samples from multiple human donors.

We evaluated the overlap between the 762 transcripts that were significantly altered by Sh3-mediated DDX39B knockdown in $CD4^+$ T cells from both donors 1 and 4 (Fig. 1A, Table S2) and 558 putative MS susceptibility genes identified by the International Multiple Sclerosis Genetics Consortium (International Multiple Sclerosis Genetics, 2019). Expression of 421 of these MS susceptibility genes was detected in our RNAseq data and expression of 41 of these was significantly altered by DDX39B depletion (e.g., IL2RA, Fig. S1B). This overlap meant that DDX39B differentially expressed genes (DEG) were highly enriched among MS susceptibility genes ($p = 0.00013$) (Figs. 1B & S1C).

Based on eQTL information of the disease associated alleles (Consortium et al., 2017), we classified 250 of the 558 MS susceptibility genes as potentially 'pathogenic' since the disease-associated allele also associates with higher gene expression in lymphoblastoid cells, and 262 genes as potentially 'protective' since the disease-associated allele associates with lower expression of the gene (Fig. S1C & Table S3). We found significantly more overlap than expected by chance between MS pathogenic genes and protective genes with genes upregulated and downregulated upon DDX39B knockdown, respectively ($p = 0.00015$ and $p = 0.03$; Fig. S1D). In contrast, there was no significant overlap when the directional pairing was flipped to compare MS pathogenic genes with those downregulated by DDX39B knockdown ($p = 0.7$) or MS protective genes with those upregulated by DDX39B knockdown ($p = 0.1$) (Fig. S1D).These data strongly suggested a shift of gene expression signature from protective to pathogenic upon DDX39B depletion, and provide support for a protective role for DDX39B in MS risk.

DDX39B controls expression of FOXP3. Among the 762 transcripts that were significantly altered by Sh3-mediated DDX39B knockdown in T cells from both donors (Fig. 1A, Table S2) we identified Forkhead Box P3 (FOXP3) transcripts. Indeed, FOXP3 transcripts were among 122 high confidence targets of DDX39B significantly reduced in $CD4^+$ T cells from both donors 1 and 4 with both Sh3 and Sh5 (Fig. S1E, Table S1). Given the importance of FOXP3 in the development, maintenance and function of Tregs (Georgiev et al., 2019; Josefowicz et al., 2012), and its strong association with autoimmune diseases (Dominguez-Villar and Hafler, 2018) we investigated this further.

Consistent with the measurement from the RNAseq, RT-qPCR showed FOXP3 RNA was reduced in DDX39B-depleted primary $CD4^+$ T cells from the two aforementioned donors and those from four additional healthy donors (Fig. 1C). Importantly, FOXP3 protein was decreased in $CD4^+$ T cells lysates that had measurable levels of total protein from four of these donors (Figs. 1C & S1F). In blood-derived $CD4^+$ T cells FOXP3 is predominantly expressed in $CD4^+/CD25^+$ Tregs (Josefowicz et al., 2012). To address the effect of DDX39B knockdown in Tregs we depleted DDX39B in the MT-2 Treg-like human cell line (Hamano et al., 2015) and in primary human Tregs isolated from two healthy donors (donors 7 and 8) (Liu et al., 2006). As observed in $CD4^+$ T cells, knockdown of DDX39B led to loss of expression of FOXP3 RNA and protein in MT-2 cells (Figs. 1D & S1G) and in primary human Tregs (Figs. 1E, S1H & S1I). Our data showed that FOXP3 RNA and protein expression were highly sensitive to DDX39B depletion.

DDX39B depletion disrupts Treg-specific gene expression. Given the effect of DDX39B knockdown on FOXP3 expression we predicted important alterations in immune relevant gene expression networks (Fig. 2A) and tested this prediction in primary $CD4^+$ T cells, Treg-like MT-

2 cells and primary Tregs depleted of DDX39B using various approaches. Using Gene Set Enrichment Analysis (GSEA) (Subramanian et al., 2005), we identified gene sets enriched in genes differentially expressed upon DDX39B knockdown in $CD4^+$ T cells (Table S4). Among 648 gene sets enriched for genes whose expression was decreased upon DDX39B knockdown (nominal p-value ≤ 0.05), we found many related to immune function, including a gene set of human homologues of murine FOXP3 targets identified by Gavin et al (Gavin et al., 2007) (Figs. 2B & S2A).

To confirm regulation of FOXP3-driven gene expression network by DDX39B in an independent data set, we explored the ConnectivityMap (CMap) database provided by The Broad Institute (Subramanian et al., 2017). CMap is a large-scale catalog of transcriptional responses of human cells to chemical and genetic perturbations that enables the identification of conditions causing similar transcriptional responses. We found that the transcriptional profile of DDX39B knockdown (ID: CGS001-7919) showed significant overlap with that of FOXP3 knockdown (ID: CGS001-50943, Table S5). These results suggested that the gene expression networks of DDX39B and FOXP3 are tightly linked, and are consistent with our model that DDX39B regulates FOXP3 expression.

We evaluated the effect of DDX39B knockdown on FOXP3 targets directly, and established that IL2RA (CD25) and IL10 RNAs were significantly downregulated in DDX39B depleted $CD4^+$ T cells and primary Tregs (Fig. 2C). Other FOXP3 targets behaved differently in these two cell populations, for instance EBI3, ICAM1, and TNFRSF9 were consistently downregulated by DDX39B depletion in $CD4^+$ T cells (Fig. S2B) but not in Tregs (Fig. S2C), suggesting the existence of FOXP3 independent pathways altered by DDX39B. To complement gene expression analysis, we examined the effect of depleting DDX39B on expression of immune

markers and secretion of immune modulators. Cell surface expression of IL2RA (CD25) was decreased in MT-2 cells depleted of DDX39B (Fig. 2D) and levels of secreted IL10 decreased upon knockdown in $CD4^+$ T cells and Tregs (Fig. 2E).

We explored the effect of DDX39B knockdown on the expression of important transcriptional regulators of different T cell lineages (Fig. 2A). DDX39B knockdown in MT-2 cells and primary Tregs decreased expression of TBX21, a transcriptional regulator of T_H1 cells, but increased expression of GATA3, a transcriptional regulator of T_H2 cells (Fig. 2F). DDX39B depletion did not alter expression of RORC, a regulator of T_H17 cells (data not shown). Given the complex composition of T cell populations (Josefowicz et al., 2012) these results have to be carefully interpreted; nonetheless, they suggest the acquisition of gene expression associated with T_H2 cells. The data presented above show that DDX39B activates FOXP3 expression and therefore expression of downstream immune mediators.

Low DDX39B causes FOXP3 intron retention. We investigated the mechanistic basis for the dependence of FOXP3 RNA accumulation on DDX39B. Since DDX39B plays important roles in constitutive splicing (Fleckner et al., 1997; Kistler and Guthrie, 2001; Shen et al., 2008), alternative splicing (Galarza-Munoz et al., 2017; Nakata et al., 2017), and nucleo-cytoplasmic transport (Gatfield et al., 2001; Huang et al., 2018; Jensen et al., 2001; Luo et al., 2001) we investigated which of these roles was affected in FOXP3 transcripts. We depleted DDX39B in the Treg-like MT-2 cells and biochemically fractionated cytoplasm, nucleoplasm and chromatin (Fig. 3A) to determine which cellular compartment(s) showed reduced FOXP3 transcripts. In control conditions, FOXP3 transcripts were disproportionally found to be associated with chromatin or in the nucleoplasm relative to transcripts coding for the translation elongation factor EEF1A1, and this association was increased for FOXP3 but not EEF1A1 RNAs upon

DDX39B knockdown (Fig. 3B). A significant reduction of FOXP3 RNAs was clearly observed in all three cellular fractions (Fig. 3C). These data indicate that the DDX39B effect on FOXP3 transcripts occurred early during their biogenesis.

Given this early effect, we analyzed changes in FOXP3 RNA splicing. The RNAseq data suggested that DDX39B knockdown in $CD4^+$ T cells led to increased retention of FOXP3 introns (Fig. 4A). The DDX39B-sensitive retention of FOXP3 introns was confirmed using FOXP3 intron-specific RT-qPCR from total primary $CD4^+$ T cell RNA (Fig. 4B), total and chromatin-associated MT-2 RNA (Figs. 4C & S3A), and total Treg RNA (Fig. S3B). Retention of *FOXP3* introns 2, 4, 6, 7, 9, and 11 in DDX39B depleted MT-2 cells was rescued by expression of recombinant DDX39B (Figs. 4D & 4E). FOXP3 protein expression was also rescued although to a lesser magnitude than intron retention (Fig. 4D). We concluded that DDX39B is required for efficient intron removal for several FOXP3 introns and this results in a dramatic decrease in overall levels of FOXP3 RNA and protein.

To ascertain if changes in intron retention were seen for other DDX39B targets we examined global changes in alternative splicing upon DDX39B knockdown in our RNAseq data. First and foremost, we noted that only a limited number of splicing events detected by RNAseq were significantly altered by DDX39B depletion (Fig. S3C). As with transcript level changes, we noted more alternative splicing events changing with Sh3 than Sh5, and more with Donor 1 than Donor 4 (Fig. S3C). We detected very few events in common for all conditions, perhaps because of the inefficient knockdown of DDX39B observed in the libraries from cells treated with Sh5 and the well-documented heterogeneity of alternative splicing between individuals (Wang et al., 2008). Therefore, most of our analysis focused on events changed by Sh3 in $CD4^+$ T cells from Donors 1 and 4 (Fig. S3C & Table S6). Retained introns were the most frequently observed type

of RNA splicing event changed (n = 784) with increased retention in 397 introns and decreased retention in 387 (Fig. S3C). The second most common changes were in cassette exon use (n = 488), followed by alternative use of 5' and 3' splice sites (Fig. S3C). We conclude that DDX39B regulates levels of splicing of a subset of genes and exerts its control primarily by modulating intron retention as observed for FOXP3 transcripts.

***FOXP3* introns have C-rich py tracts**. Since removal of *FOXP3* introns was sensitive to DDX39B levels, we compared *FOXP3* splice sites to those found in introns of other protein coding genes on the X chromosome (Fig 5A). 5' splice sites in *FOXP3* were only modestly weaker than those found in the average gene, whereas 3' splice sites were considerably weaker (Table S7) and this weakness was driven by the composition of the polypyrimidine tract (py tract) (Fig. 5A). *FOXP3* introns contain cytosine (C)-rich py tracts, which on the average are more likely to have a C at each position, except positions -19 and -20 (measured relative to the guanine (G) in the 3' splice site AG marked as position -1) (Figs. 5A & 5B). This contrasts with our analysis of py tracts in other genes on the X chromosome (Figs. 5A & 5B) and previous analysis for all human introns spliced by the major (U2 snRNP dependent) spliceosome (Yeo and Burge, 2004), where the average position in the py tract is most likely to be occupied by uracil (U).

Remarkably, C-rich py tracts in *FOXP3* introns are conserved in mammals from both monotreme and theriiformes subclasses, which diverged over 200 million years ago (Tarver et al., 2016) (Fig. 5C). Furthermore, the amphibian *Xenopus laevis FOXP3* has a U-poor py tract, although in this case the Us are replaced by all three other nucleotides (Fig. 5C). The zebrafish *Danio rerio* has two *FOXP3* paralogs and py tracts in one of these (*FOXP3a*) trend towards the C-rich py tracts of mammals, while the other (*FOXP3b*) has U-rich tracts (Fig 5C). An

examination of *FOX* family genes revealed that close relatives of *FOXP3* diverge in their py tracts, *FOXP4* shares C-rich py tracts, while *FOXP1* and *FOXP2* have U-rich tracts (Fig. 5D). *RBFOX2*, a RNA binding protein and not a member of the FOX family of transcription factors, is shown as a phylogenetic outlier (Fig. 5D). C-rich py tracts appear to be the ancestral form in the family as py tracts in the more divergent *FOXN1* are more like those in FOXP3 and FOXP4 (Fig. 5D). The evolutionary conservation of C-rich py tracts in these *FOX* family genes suggests important regulatory function.

C-rich py tracts determine sensitivity of FOXP3 introns to DDX39B depletion. Since we had previously shown that replacement of U residues with C in intronic py tracts inhibits splicing of model pre-mRNA substrates *in vitro* (Roscigno et al., 1993), we posited that the C-rich py tracts in FOXP3 introns would make these inefficient and highly sensitive to DDX39B levels. To test this we made reporter constructs where FOXP3 introns interrupted a Renilla luciferase ORF, and splicing efficiency could be inferred by luciferase activity and measured directly using RT-PCR (Fig. 6A). FOXP3 introns, but not an efficient human *ßglobin* intron 2, markedly reduced luciferase expression and splicing of the luciferase reporter (data not shown). Importantly, splicing reporters containing *FOXP3* introns were markedly dependent on DDX39B (Figs. 6B & S4A-C). While conversion of the 5' splice site in FOXP3 introns 7 and 11 to consensus only modestly relieved DDX39B dependency, replacement of their py tracts with U-rich tracts made these introns insensitive to DDX39B knockdown (Figs. 6B, S4B-C). These findings indicated that weak C-rich py tracts are necessary for the strong DDX39B dependency of FOXP3 introns.

To determine whether or not other DDX39B-sensitive intron retention events shared sequence features with FOXP3 introns, we analyzed our RNAseq data from DDX39B-depleted $CD4^+$ T cells. When 500 randomly selected unaffected introns were compared to 397 introns

with increased retention upon DDX39B knockdown with Sh3 in either donor 1 or 4 we found no difference in the 5' splice site maximum entropy score and a modest but statistically significant decrease in the 3' splice site maximum entropy score (Fig. S4D). Since the lower 3' splice site score for FOXP3 introns was driven by the C-rich py tract, we examined the composition of the py tract in DDX39B-sensitive introns. We found that while the frequency of U residues in DDX39B sensitive events was significantly lower than in unaffected events, the frequency of C was significantly higher (Fig. 6C). There was a statistically significant but very small increase in G residues in these C-rich py tracts (Fig. 6C). An equivalent analysis interrogating all intron retention events altered with DDX39B knockdown with both Sh3 and Sh5 led to same conclusion: DDX39B sensitive introns are enriched for introns with C-rich py tracts (Fig. S4E). We concluded that among introns with strong dependence on DDX39B there is significant enrichment for C-rich py tracts.

To address whether or not C-rich py tracts are sufficient to make an intron DDX39B sensitive, we analyzed splicing of introns 11, 14 and 19 in the *FOXP1* gene (Fig. S4F), which encodes a homologue of *FOXP3* that contains introns with U-rich py tracts (Figs. 5D & S4F). Retention of *FOXP1* introns 11, 14 and 19 was not increased with DDX39B knockdown in MT-2 cells (Fig. S4F). We modified the *FOXP3* intron 11 luciferase reporter replacing the intron with the first and last 75 nucleotides of *FOXP1* intron 19. As expected the *FOXP1* intron 19 reporter was not sensitive to DDX39B knockdown (Fig S4G). We replaced the *FOXP1* intron 19 py tract with that of *FOXP3* intron 11 and noted that this C-rich tract did not confer DDX39B sensitivity on *FOXP1* intron 19 (Fig S4G). *En masse* analysis of the 387 introns that were less retained upon DDX39B depletion also suggested that the presence of a C-rich py tract is not sufficient to confer DDX39B-dependency. Among these introns we detected an increase in the frequency of C-rich

py tracts, albeit of smaller magnitude than among more retained introns (Fig. S4H). Therefore a C-rich py tract was required but not sufficient to impart DDX39B-dependency on any intron, indicating the *FOXP3* introns have other elements that cooperate with the C-rich py tract to confer DDX39B-dependency.

To further test our conclusions on the role of C-rich py tracts, we interrogated the behavior of introns in other X chromosome genes selected only because they contained either C-rich or U-rich py tracts. Of twelve introns with C-rich py tracts, in *FAM3A*, *PORCN*, *RBM10*, *RENBP*, *CFP* and *G6PD*, all but two, both in the *CFP* gene, were more retained upon DDX39B knockdown in MT-2 cells (Fig. S4I). In contrast, four U-rich py tract introns in *FMR1* and *DDX3X* were tested and were not more retained upon DDX39B depletion (Fig S4I), thereby supporting the requirement of C-rich py tracts for DDX39B-dependency.

Collectively, the data presented in this section show that C-rich py tracts are required for the strong DDX39B-dependency of *FOXP3* introns and suggest the existence of a sub-class of introns where the C-rich py tract is required for said dependency.

Splicing of FOXP3 introns requires the ATPase activity of DDX39B but not its helicase activity. The py tract is recognized early in the splicing reaction by the U2 small nuclear RNA auxiliary factor (U2AF) composed of U2AF1 ($U2AF^{35}$) and U2AF2 ($U2AF^{65}$) (Ruskin et al., 1988; Zamore and Green, 1989). U2AF2 binds preferentially to U-rich sequences in the py tract *in vitro* (Singh et al., 1995; Zamore et al., 1992) and this is recapitulated by crosslinking-immunoprecipitation experiments *in vivo* (Wu and Fu, 2015). U2AF2 binds DDX39B, which licenses the U2AF-mediated recruitment of U2 snRNP to the neighboring branchpoint sequence to form the pre-splicesome (Shen et al., 2008). We conjectured that the exquisite dependency of C-rich py tract FOXP3 introns on DDX39B was mediated by its ability to promote pre-

spliceosome formation. Pre-spliceosome formation requires the ATPase activity of DDX39B, but not its helicase activity, which is required for later steps in spliceosome assembly (Shen et al., 2008). While expression of wildtype (WT) DDX39B rescued splicing of FOXP3 intron 11 in DDX39B depleted cells, two ATPase defective mutants, K95A and E197A (Shen et al., 2007), did not rescue splicing (Figs. 7 & S5). On the other hand, the helicase defective DDX39B mutant D199A that retains ATPase activity and supports pre-spliceosome formation (Shen et al., 2008; Shen et al., 2007) rescued FOXP3 intron 11 splicing (Figs. 7 & S5). Of note, the data in Fig. 7 indicate that the DDX39B D199A mutant rescued FOXP3 intron 11 splicing even better than the WT DDX39B. This result could be explained by the fact that while the D199A lacks helicase activity, it has higher ATPase activity than WT DDX39B (Shen et al., 2007). These results indicate that the C-rich py tract of FOXP3 intron 11 is highly dependent on the DDX39B ATPase activity and this unique dependency may distinguish its spliceosome assembly pathway from more conventional introns. Nine of eleven FOXP3 introns belong to the C-rich py tract subclass, and are all likely to share this requirement for the DDX39B ATPase activity, which explains why overall FOXP3 expression is exquisitely sensitive to DDX39B levels.

Discussion

DDX39B (also known as BAT1) was proposed to be an anti-inflammatory factor (Allcock et al., 2001) and we have shown that *DDX39B* is genetically associated with MS, an autoimmune disease (Galarza-Munoz et al., 2017). Previously we discovered a molecular mechanism that partially explained these phenomena. DDX39B is a potent activator of IL7R exon 6 splicing and a repressor of the production of sIL7R (Galarza-Munoz et al., 2017), which exacerbates the MS-like EAE (Lundstrom et al., 2013). Complementing this molecular analysis, we demonstrated

that alleles in DDX39B and IL7R are epistatic relative to MS risk and individuals homozygous for the risk alleles at both loci have an approximately 3-fold higher chance of developing MS (Galarza-Munoz et al., 2017). While this connection between DDX39B and autoimmunity was compelling, evolutionary arguments suggested that this RNA helicase would play other roles in immunity. Our exploration of this suggestion led to the discovery, described above, that DDX39B is exquisitely required for the splicing of FOXP3 transcripts. Low levels of DDX39B lead to low levels of FOXP3 and by inference low numbers of Tregs. CRISPR screens to discover regulators of FOXP3 did not identify DDX39B (Cortez et al., 2020; Loo et al., 2020; Schumann et al., 2020), but this is likely because knockout of DDX39B results in lethality.

The connection between FOXP3 and autoimmunity is well established: mutations in FOXP3 cause a systemic and severe autoimmune syndrome in humans (IPEX) (Bennett et al., 2001; Chatila et al., 2000; Wildin et al., 2001) and the equivalent (scurfy) in mice (Brunkow et al., 2001). While there is evidence supporting a prominent role for FOXP3 in MS (Fletcher et al., 2009; Sambucci et al., 2018), genetic variation within FOXP3 has yet to be associated with MS largely due to the fact that the X chromosome has been understudied in human genetic association studies. The role of Tregs, which depend on FOXP3 transcriptional control (Hori et al., 2003), on autoimmunity has been carefully documented (Dominguez-Villar and Hafler, 2018). Furthermore, it is clear that high expression of FOXP3 is required for full Treg function and there is a FOXP3 dose-dependent effect on immune supressive activity (Wan and Flavell, 2007). Therefore, given the strong dependence of FOXP3 expression on DDX39B levels, it is likely that humans with low levels of DDX39B will have reduced FOXP3 expression and by extension low levels of Treg cell development, maintenance and suppressive function. Furthermore, given our data that DDX39B knockdown promotes GATA3 expression it is

356 possible that low levels of DDX39B could promote the conversion of Tregs to T_H2-like self-
357 reactive effector cells (Komatsu et al., 2014; Noval Rivas et al., 2015; Zhou et al., 2009). The
358 DDX39B control of FOXP3 expression cements its role in immune regulation.

359 There are other intriguing connections between DDX39B and immunity. First, we show that
360 putative MS susceptibility genes are regulated by DDX39B and some are likely regulated
361 independently of effects on FOXP3. Second, DDX39B regulates the nucleocytoplasmic transport
362 of human circular RNAs (circRNAs) larger than 1200 nucleotides (Huang et al., 2018). There is
363 an emerging realization that circRNAs play important, although heretofore incompletely
364 understood roles, in immunity and autoimmune disease (Zhou et al., 2019). Third, *DDX39B*
365 resides in the class III region of the MHC and many of its neighbors code for proteins involved
366 in RNA transactions (e.g., *DXO* and *SKIV2L*) (Lehner et al., 2004; Schott and Garcia-Blanco,
367 2021). This overrepresentation of genes involved in RNA metabolism within the class III region
368 is found in all vertebrates we have examined, suggesting an important immune function for these
369 RNA binding proteins. These observations, first revealed by studying DDX39B, suggest new and
370 interesting connections between RNA and immunity.

371 DDX39B, which is also known as UAP56, was characterized as a factor required for
372 formation of splicing complexes (Fleckner et al., 1997; Shen et al., 2008). Here we describe a
373 subclass of introns that have C-rich py tracts and depend on high levels of DDX39B for splicing.
374 Although these py tracts are poor binding sites for U2AF2 (Singh et al., 1995), we observed that
375 these introns depend on this splicing factor (Hirano et al., data not shown), and thus propose that
376 DDX39B driven U2 snRNP binding to the branchpoint sequence stabilizes U2AF2 at these
377 suboptimal sites. We suggest that for these introns the U2 snRNP complex may be the first
378 splicing commitment complex, which would be distinct from U2 snRNP-independent

commitment complexes in introns with U-rich py tracts (Jamison et al., 1992; Li et al., 2019; Seraphin and Rosbash, 1989). Our studies indicate that these introns are exceedingly sensitive to DDX39B ATPase activity, but are impervious to its helicase activity. Shen et al showed that DDX39B ATPase activity, but not helicase activity, is required for binding of U2AF and recruitment of U2 snRNP to the branchpoint sequence of the 3' splice site to form the pre-spliceosome (Shen et al., 2008). This suggests that once bound to the branchpoint sequence and py tract in these C-rich introns, the U2 snRNP-U2AF complex does not require further remodeling by DDX39B. In contrast to this, activation of IL7R exon 6 requires DDX39B helicase activity (Galarza-Munoz et al., 2017). Our data thus indicate distinct functions for DDX39B in different splicing events, which implies multiple pathways to spliceosome assembly as has been suggested before (Kistler and Guthrie, 2001; Newnham and Query, 2001). An intriguing possibility is that FOXP3 introns, and perhaps many introns with C-rich py tracts, are actually prone to intron 'detention' rather than retention and that certain stimuli may free them from DDX39B dependency (Boutz et al., 2015).

Acknowledgments

We thank Drs. Paul Boutz (Rochester) and Harinder Singh (Pittsburgh) for excellent suggestions on an early draft of this manuscript. We acknowledge members of our laboratories for excellent discussions and comments to improve the science and the manuscript (MGB, SSB, SGG, DCK). WSF thanks John Paul Donohue (RNA Center for Molecular Biology, UC-Santa Cruz) for expert advice on the analysis of RNA sequencing data. MGB thanks Drs. Michael Sheetz and Linda Kenney (UTMB) for use of their house in Santa Fe, NM to write the final draft of this manuscript.

Funding: We acknowledge support from the Uehara Foundation Fellowship and McLaughlin Postdoctoral Fund (MH), NIH F32 NS087899 (GGM), Duke Neurology startup and Stone family funds (SGG), KL2 TR001441-07 (WSF), R21AI133305 (DCK and LW), Duke MGM Start-up funds (DCK), R01 CA204806 (MGB), UTMB startup funds (MGB), and P01 AI150585 (MGB).

Author contributions: Conceptualization: MH, GGM, WSF, LY, DCK, SSB, and MGB contributed to conceptualization of the manuscript. Data acquisition and initial interpretation: MH contributed to Figs. 1,2,5-7 and related supplementary material, and worked with XY to carry out statistical analyses of data involving donor samples. GGM contributed to Figs. 1-4 and related supplementary materials. LW contributed data to Fig. 1 and related supplementary materials. AA contributed data to Fig. 2. XY contributed to all statistical analysis of donor samples in Figs. 1, 2 and related supplementary materials. SW carried out the bulk RNA sequencing and contributed to its analysis. GS contributed Figs. 1 and 4. CN contributed to Fig. S4. WSF contributed to Figs. 1, 3 and 4, and related supplementary material. SSB contributed to Figs. 3 and 6 and related supplementary material. Writing, Review, Editing and Visualization: MH and MGB wrote the first draft and designed the figures, which were created primarily by

MH with contributions from GGM and many other authors. All the other authors contributed to editing and discussion. MGB assembled edits from co-authors and prepared the final draft of the manuscript. Supervision and analysis: All authors contributed to analysis and interpretation of the data, WSF and SW analyzed the raw RNA seq data, DCK, FB, & SGG supervised analysis of data in Fig. 1. SSB and MGB supervised all the studies and participated in the analysis of all the data presented.

Competing interests: GGM and MGB acknowledge that they have significant ownership in Autoimmunity Biologic Solutions, Inc (Galveston, TX), which is commercializing therapies that target the IL7R pathway in autoimmune diseases. While we do not believe this represents a conflict of interest it can lead to the perception of said conflict.

Data and materials availability: There will be no restriction on any material described in this manuscript. The bulk RNA sequencing datasets have been uploaded to the Gene Expression Omnibus with accession number GSE145773. All data are available in the main text, supplementary materials or will be made available upon request.

Materials and Methods

Cells

MT-2 cells were kindly provided by Bryan R. Cullen (Duke University) and grown at 37°C in Roswell Park Memorial Institute (RPMI) 1640 medium (Thermo Fisher Scientific) with 10% (v/v) fetal bovine serum (FBS) (Genesee Scientific) and 1% (v/v) penicillin-streptomycin (Thermo Fisher Scientific). Human embryonic kidney 293T/17 (HEK 293T/17; ATCC CRL-11268) and HeLa (ATCC CCL-2) cell lines were obtained from the Duke University Cell Culture Facility. The HeLa Flp-In T-Rex cell line (HeLa-Flp-In) was kindly provided by Dr. E. Dobrikova (Duke University). HEK293T/17, HeLa, and HeLa-Flp-In cells were cultured at 37°C in Dulbecco's Modified Eagle Medium (DMEM) (Thermo Fisher Scientific) with 10% (v/v) FBS, 1% (v/v) penicillin-streptomycin and 2.5 μg/ml Plasmocin (InvivoGen). HeLa_Flp-In cell lines expressing wild-type or mutant versions of DDX39B (K95A, E197A and D199A) were grown as described above except media was supplemented with 2.5 μg/ml blasticidin (InvivoGen) and 200 μg/ml Hygromycin B (Thermo Fisher Scientific). All cells were free of Mycoplasma contamination as confirmed by routine testing using the MycoAlert Mycoplasma Detection Kit (Lonza).

Primary human $CD4^+$ T cells were isolated and cultured as described previously (Galarza-Munoz et al., 2017). In brief, peripheral blood mononuclear cells (PBMCs) from healthy donors were isolated from whole blood or buffy coats using the Ficoll-Plaque Gradient method. Blood samples were collected at Duke University following the institutional IRB protocol (#

Pro00070584) and buffy coats were purchased from the New York Blood Center. $CD4^+$ T cells were further isolated using $CD4^+$ T cell isolation kit (Miltenyi Biotec). The isolated cells were cultured in RPMI medium supplemented with 20% (v/v) FBS and 100 ng/mL human recombinant IL-2 (Peprotech). Two days prior to transduction, $CD4^+$ T cells or MT-2 cells were activated with media containing 50 ng/mL anti-CD3 (eBioscience) and 100 ng/mL anti-CD28 (BD Biosciences) antibodies.

Primary human T regulatory cells (Tregs) were isolated from PBMCs using $CD4^+$ $CD25^+$ $CD127^{dim/-}$ Regulatory T Cell Isolation Kit (Miltenyi Biotec). The isolated Treg cells were expanded in culture for 14 days with TexMACS medium (Miltenyi Biotec) supplemented with 5% (v/v) Human AB Serum (Sigma), 500 IU/mL recombinant human IL2 (PeproTech) and Treg Expansion kit (Miltenyi Biotec) which contains beads pre-loaded with anti-CD3 and anti-CD28 antibodies.

Plasmids

shRNA pLKO.1 plasmids expressing non-targeting control shRNA (NTC: SHC002) or anti-DDX39B shRNAs (Sh3: TRCN0000286976; and Sh5: TRCN0000294383) were purchased from Millipore Sigma. The construct pLCE-DDX39B was generated by cloning the coding sequence of *DDX39B* into the pLCE vector. Splicing reporter plasmids (pcDNA3.1-RLuc_FOXP3 Intron 7, pcDNA3.1-RLuc_FOXP3 Intron 9, pcDNA3.1-RLuc_FOXP3 Intron 11, and pcDNA3.1-RLuc_HGB1 Intron 2) were constructed by inserting introns 7, 9 or 11 of FOXP3 (NC_000023), or intron 2 of HGB1 (NC_000011) within the open reading frame of *Renilla Luciferase* (RLuc)

476 at nucleotide positions 599, 601, 408 or 253, respectively, in the pcDNA3.1 vector (pcDNA3.1-
477 RLuc). Splice sites (SS) mutations were introduced in the corresponding parental plasmid using
478 In-Fusion HD Cloning kit (Takara Bio) to generate pcDNA3.1-RLuc_FOXP3 Intron 7 5'-SS G-
479 U, pcDNA3.1-RLuc_FOXP3 Intron 7 3'-SS PolyU, pcDNA3.1-RLuc_FOXP3 Intron 11 5'-SS
480 G-U, and pcDNA3.1-RLuc_FOXP3 Intron 11 3'-SS PolyU. To construct pcDNA3.1-
481 RLuc_FOXP1 Intron 19 WT and 3' C-rich py tract mutant, synthesized DNA fragments with the
482 FOXP1 intron sequences (NG_028243.1) were purchased from Integrated DNA Technologies
483 and were subcloned into the pcDNA3.1-RLuc plasmid. Point mutants DDX39B K95A, E197A
484 and D199A were introduced in the parental pcDNA5/FRT/TO-DDX39B plasmid (Galarza-
485 Munoz et al., 2017) using QuickChange Lightning Mutagenesis Kit (Agilent Technologies). All
486 constructs and mutations were confirmed by Sanger sequencing. Primers used to generate these
487 constructs are shown in Table S8.

488

489 **Establishment of stable cell lines**

490 Generation of the inducible HeLa cell line expressing wild-type DDX39B or DDX39B mutant
491 D199A was described previously (Galarza-Munoz et al., 2017). HeLa cell lines stably expressing
492 DDX39B mutants (K95A and E197A) were constructed similarly. In brief, HeLa Flp-In T-Rex
493 cells were co-transfected with these constructs in the pcDNA5/FRT/TO plasmid and pOG44
494 plasmid, which encodes Flp recombinase, using Lipofectamine 2000 (Thermo Fisher Scientific).
495 Transfected cells were selected with 2.5 μg/ml blasticidin and 200 μg/ml Hygromycin B for 15
496 days. Expression of the transgene was induced by addition of 1 μg/mL doxycycline in the
497 culturing media.

498

Lentiviral packaging

Lentiviral packaging of shRNA constructs was conducted in 293T/17 cells using Lipofectamine 2000 (Thermo Fisher Scientific). In brief, 1.0×10^7 293T/17 cells were seeded in 15-cm dishes in DMEM media and cultured overnight (three 15-cm dishes per construct). Cells were co-transfected with 17 μg of the corresponding shRNA pLKO.1 vector, pLCE-DDX39B or pLCE-GFP, 17 μg of packaging plasmid (pCMVR8.74) and 7 μg of VSV-G envelope plasmid (pMD2.G) in serum-free media and 160 μL of Lipofectamine 2000. The media was replaced with 20 mL fresh DMEM media 18 hr post-transfection. After 72 hr, supernatants were collected, filtered through 0.45 mm filters and concentrated to 6 mL in Amicon Ultra 100K centrifugal filter units (Millipore Sigma). Concentrated lentiviral particles were used immediately or stored at -80°C.

Lentiviral transduction of human primary $CD4^+$ T cells, MT-2 cells and primary Tregs

Transduction of human primary $CD4^+$ T cells with either NTC or anti-DDX39B shRNAs was described previously.(Galarza-Munoz et al., 2017) In brief, 4.0×10^6 activated primary $CD4^+$ T cells from each donor were transduced in T25 flasks with lentiviruses encoding control (NTC) or DDX39B-targeting (Sh3 or Sh5) shRNAs for 3 days, and transduced cells were selected under medium supplemented with 1.5 μg/mL puromycin for 4 days. Cells were then cultured for 24 hours in the absence of puromycin and collected for functional analyses (8 days after initial transduction). Transduction of activated MT-2 cells (1.0×10^7 cells) was carried out as for primary $CD4^+$ T cells. Activated primary Tregs (1.0×10^5 cells) were transduced in 96 well plates and cultured for 7 days with half media exchange every 2 days (without puromycin

selection). In all cell types, depletion of DDX39B protein was confirmed by western blot using anti-UAP56 (DDX39B) antibody (ab18106, Abcam) and α-Tubulin (AB_1904178, Cell Signaling Technology) or PTBP1 (anti-PTB rabbit serum (Wagner and Garcia-Blanco, 2002)) as loading control.

DDX39B rescue in MT-2 cells

To rescue DDX39B expression in MT-2 cells, 1.0×10^7 cells were co-transduced with the following combination of shRNAs and expression plasmids for 3 days: (i) NTC shRNA + pLCE-GFP (control), (ii) Sh5 shRNA + pLCE-GFP (knockdown), and (iii) Sh5 shRNA + pLCE-DDX39B (rescue). The cells were then selected with RPMI medium supplemented with 1.5 μg/mL puromycin for 2 days, followed by 24-hour culture in the absence of puromycin, and collection of cell lysates for RNA and protein analyses (6 days after initial transduction). We used the DDX39B shRNA Sh5 since it targets the 3' untranslated region of DDX39B mRNA, and thus it depletes the endogenous DDX39B mRNAs but not the trans-gene DDX39B mRNAs. Depletion or rescue of DDX39B protein was confirmed by western blot using anti-UAP56 (DDX39B) antibody (ab18106, Abcam) and α-Tubulin (AB_1904178, Cell Signaling Technology) as loading control.

RNA-seq library preparation, sequencing and analyses

Total RNA was isolated from control (NTC) or DDX39B-depleted (Sh3 or Sh5) $CD4^+$ T cells using ReliaPrep RNA Cell Miniprep System (Promega), and treated in-column with DNase I following the manufacturer's recommendations. Poly-A^+ RNA was enriched from 1 μg of total

RNA and used as template to generate libraries using the Illumina TruSeq platform as recommended by the manufacturer. Libraries were sequenced on a 2x100 paired-end format on an Illumina Hi-Seq 1500. Reads were aligned to the human GRCh38 reference with program STAR version 2.5.2b, using the parameters recommended for the ENCODE consortium (Consortium, 2011) (Dobin et al., 2013). The STAR genome index was built from the GENCODE primary genome assembly and the corresponding primary annotation file. Gfold software, version 1.1.4, with default parameters was used to count reads per gene and estimate expression differences between treatments. The GENCODE V24 basic annotation file was used for the gene counts. In order to determine significantly changed transcript abundance, an initial cutoff was used to discard transcripts averaging an RPKM of less than 2, then with a secondary cutoff of |GFold value| $\geq$ 0.3 or $\geq$ 0.1 (for Sh3 or Sh5, respectively, compared to NTC from the same donor). A lower cutoff was used for Sh5 since it was less effective than Sh3 at knocking down DDX39B. Finally, only transcripts fulfilling these parameters in both Donor 1 and Donor 4 were included as significantly changed (see Table S1).

Splicing analysis was carried out using Vast-tools program version 0.2.1 (Irimia et al., 2014) by aligning the paired-end reads to the Vast-tools human database (vastdb.hsa.7.3.14) using the default parameters. The Sh3-treated sample from each donor were compared using the differential function of vast-tools to determine changes in splicing relative to NTC in the same donor. Subsequently, we used the minimum value of change percent spliced in at 95% confidence interval value (MV|dPSI|; https://github.com/vastgroup/vast-tools) to filter for significantly changed events using 0.15 as cutoff (Table S6).

566 **MS susceptibility gene enrichment analysis**

567 The study by the International Multiple Sclerosis Genetics (IMSG) Consortium (International
568 Multiple Sclerosis Genetics, 2019) identified single nucleotide polymorphisms (SNPs)
569 genetically and/or functionally associated with increased MS risk, from which a map of MS
570 susceptibility genes was determined. We created a functional classification of the MS
571 susceptibility genes (the genes with exonic or intronic SNPs, or cis-eQTL, MS_Susceptibility:
572 558 genes) (see supplemental tables 7 & 19 in (International Multiple Sclerosis Genetics, 2019)),
573 in lymphocytes based on expression quantitative trait loci (eQTL). Gene expression profiles of
574 the MS susceptibility genes in EBV-transformed lymphocytes were obtained from GTExPortal
575 (Consortium et al., 2017) and merged with data of odds ratios (OR) of MS risk to generate risk
576 SNP and susceptibility gene pairs. Pairs of SNPs and genes with missing expression profiles
577 were excluded from the analysis. A list of 539 pairs of SNPs and genes was generated (Table
578 S3). If the MS risk allele (OR > 1) showed higher expression of the corresponding susceptibility
579 gene (Normalized expression: NES < 0) in the lymphoblastoid cells, that gene was defined as a
580 pathogenic MS gene (MS_Pathogenic: 250 genes). If the MS risk allele (OR >1) showed lower
581 expression of the corresponding susceptibility gene (NES > 0), that gene was defined as a
582 protective MS gene (MS_Protective: 262 genes).

583

584 In order to estimate enrichment of differentially expressed genes (DEGs) in pre-determined MS
585 gene sets, we used a simulation method. We determined DEGs in two donors with the shRNA
586 that effectively knocked down DDX39B, Sh3, based on the Gfold method (Feng et al., 2012)
587 with a cutoff of Gfold value change > 0.3. The observed number of DEGs found in the MS gene
588 set, k, was calculated by intersecting DEGs with each MS gene set (MS_Susceptibility,

589 MS_Pathogenic, or MS_Protective). We then calculated enrichment using a resampling method.
590 We constructed a sampling distribution for the null hypothesis (X) by randomly resampling n_1
591 genes from N, and counting the number of genes overlapping with each MS gene list n_2 over
592 100,000 permutations. An empirical p value was calculated based on the fraction of instances
593 that simulated (X) was greater than or equal to the actual observed value of k for each MS gene
594 set.

595

596 **RT-qPCR analysis of RNA expression**

597 Total RNA was isolated from control or DDX39B-depleted cells using ReliaPrep RNA Cell
598 Miniprep System (Promega) or Direct-zol RNA kit (Zymo Research), and treated in-column with
599 DNase I following the manufacturer's recommendations. Reverse transcription was conducted
600 with random primers using the High Capacity cDNA Reverse Transcription Kit (Thermo Fisher
601 Scientific). DDX39B and FOXP3 RNA levels were measured by real-time quantitative PCR
602 (RT-qPCR) using PowerUP SYBR Green Master Mix (Thermo Fisher Scientific) in a
603 StepOnePlus Real-Time PCR System (Thermo Fisher Scientific). *EEF1A1* was used as
604 normalization control since its expression was not affected by DDX39B depletion.

605

606 Genes whose expression was found to be differentially expressed upon DDX39B depletion in the
607 RNAseq dataset in primary $CD4^+$ T cells were analyzed by RT-qPCR in control (NTC) or
608 DDX39B-depleted (Sh3 and Sh5) primary $CD4^+$ T cells (from 6 donors), MT-2 cells and primary
609 Tregs (from 2 donors). Target genes and primers are shown in Table S8. The data were

normalized to *EEF1A1* expression. Differentially expressed genes were determined by comparing levels in DDX39B-depleted cells versus control cells.

Western blot analysis of DDX39B and FOXP3 protein expression

Whole cell lysates from control (NTC) or DDX39B-depleted (Sh3 and Sh5) cells were collected in 1 X RIPA buffer (150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, and 50 mM Tris-HCl at pH 7.5) freshly supplemented with 1X protease inhibitors (Roche). The cell lysates were quantified by bradford and equal amount of total protein was loaded per lane on NuPAGE 4%–12% Bis-Tris pre-cast gels (Life Technologies), transferred to nitrocellulose membranes (Whatman), and blotted using standard protocols with antibodies against DDX39B (ab18106, Abcam), FOXP3 (AB_467554, Thermo Fisher Scientific), and either PTB or α-Tubulin (AB_1904178, Cell Signaling Technology) as loading control.

Gene-set Enrichment Analysis (GSEA)

To conduct GSEA analysis with our RNAseq data set in primary $CD4^+$ T cells, we generated a gene list from the control condition (NTC: Donor1-NTC, Donor4-NTC) and a gene list from the DDX39B knockdown condition (KD: Donor1-Sh3, Donor4-Sh3). All the genes that passed the initial cutoff (RPKM of ≥2 in both NTC libraries) were used. GSEA was conducted with the C2 curated genesets collections of Molecular Signatures Database (Subramanian et al., 2005). Among 5501 genesets in the C2 collection, 730 genesets passed the cutoff of the nominal p-value

($p < 0.05$). Of these, 648 and 82 genesets showed enrichment to NTC and KD groups, respectively (Table S4).

Flow Cytometry analysis

8.8×10^6 MT-2 cells were cultured in 6 well plates and transduced with lentiviruses as described above. Control (NTC) or DDX39B-depleted (Sh3 and Sh5) MT-2 cells were fixed and permeabilized with True-Nuclear Transcription Factor Buffer Set (BioLegend). The permeabilized cells were stained with antibodies against FOXP3 conjugated with Alexa Fluor 647 (AB_2539534, Thermo Fisher Scientific), CD25 (IL2RA) conjugated with Alexa Fluor 488 (AB_493044, BD Biosciences) or fluorophore-matched isotype control. For each sample, 10,000 events were recorded on a Guava Easy-Cyte System (Millipore Sigma) and forward scatter (FCC), side scatter (SSC) and fluorescence intensity parameters were extracted. Mean fluorescence intensity (MFI) for each target was determined and plotted using FlowJo software (BD Biosciences).

Measuring IL10 in DDX39B-depleted cells

Secreted IL10 from control (NTC) or DDX39B-depleted (Sh3 and Sh5) primary $CD4^+$ T cells and Tregs was measured by Luminex cytokine profiling analysis of supernatants of cultured cells. In brief, 24 hours prior to collection, the cell medium was replaced with fresh medium free of supplements. At the time of collection, supernatants were collected on ice and stored at -80°C until analysis. Multiplex cytokine profiling was carried out on a Luminex MAGPIX system

(Luminex) using the Milliplex Human Premixed 41 Plex Immunology Multiplex panel (Millipore Sigma) following the manufacturer's recommendations.

Subcellular Fractionation of MT-2 cells

Subcellular fractionation of MT-2 cells was conducted following a protocol described previously (Bhatt et al., 2012) with minor modification. Control (NTC) or DDX39B-depleted (Sh3 and Sh5) MT-2 cells were washed with Phosphate-buffered saline (PBS), then lysed in NP40 cytoplasmic lysis buffer [0.075% (v/v) NP-40, 20 mM Tris-HCl pH7.5, 150 mM NaCl, 1mM DTT, and 1X protease inhibitor] for 2.5 minutes on ice. Cell lysates were then layered on top of a sucrose cushion [24% (w/w) sucrose, 20 mM Tris-HCl pH7.5, 150 mM NaCl, 1mM DTT, and 1X protease inhibitor] and centrifuged at 14,000 rpm for 10 min. For further separation of nucleoplasm and chromatin, the pelleted nuclei were treated with glycerol nucleoplasm lysis buffer [50% (v/v) Glycerol, 20 mM Tris-HCl pH7.5, 75 mM NaCl, 0.5 mM EDTA, 1 mM DTT and 1X protease inhibitor], mixed gently, and treated with urea nuclei lysis buffer [1% (v/v) NP-40, 1M Urea, 20 mM HEPES pH7.5, 1 mM DTT, 7.5 mM MgCl2, 0.2 mM EDTA and 1X protease inhibitor] for 2 minutes on ice, followed by centrifugation at 14,000 rpm for 2 min. Supernatants were collected as nucleoplasm fraction and the pellets were collected as the chromatin fraction. Successful separation of subcellular compartments was confirmed by western blot with antibodies against markers of the different compartments: cytoplasmic α-Tubulin (AB_1904178, Cell Signaling Technology), nucleoplasmic Nucleolin (AB_533406, Bethyl Laboratories), and chromatin-associated hnRNP C (AB_627731, Santa Cruz Biotechnology). RNA was isolated from each fraction using Direct-zol RNA kit (Zymo Research), and 200 ng of RNA from each compartment was used for reverse transcription with random primers as above.

The abundance of FOXP3 and EEF1A1 RNAs were quantified in each fraction by RT-qPCR and normalized to total RNA as above. The percentage of EEF1A1 and FOXP3 RNA in each compartment was calculated by dividing by the corresponding total signal from the three compartments.

RT-qPCR analysis of retained introns in FOXP3, FOXP1 and other RNAs

To quantify intron retention events, we designed RT-qPCR primers to amplify total FOXP3 transcripts (constitutive exon 11), spliced FOXP3 transcripts (spanning the exon-exon junction between constitutive exons 10 and 11) and retained intron-containing FOXP3 transcripts: intron 2 (spanning the intron 2/exon 3 junction), intron 4 (spanning the intron 4/exon 5 junction), intron 6 (spanning the exon 6/intron 6 junction), intron 7 (spanning the intron 7/exon 8 junction), intron 9 (spanning the exon 9/intron 9 junction) or intron 11 (spanning the intron 11/exon 12 junction) (Table S8). Likewise, we designed primers to amplify total FOXP1 transcripts (within constitutive exon 21) and FOXP1 transcripts with retained introns: intron 11 (spanning intron 11/exon 12 junction), intron 14 (spanning intron 14/exon 15 junction) or intron 19 (spanning intron 19/exon 20 junction). Retained intron values were normalized to the corresponding total FOXP3 or FOXP1 RNA, and the data for DDX39B-depleted (Sh3 or Sh5) primary $CD4^+$ T cells, MT-2 cells or primary Tregs is presented as fold-change over control (NTC). For the rescue experiments of endogenous DDX39B-depletion in MT-2 cells, cells transduced with NTC and GFP lentiviruses were used as controls. Detections of retained introns in FAM3A, PORCN, RBM10, RENBP, CFP and G6PD (C-rich py tracts) or FMR1 and DDX3X (U-rich py tracts) were conducted in the same manner. All primers used for this analysis are shown in Table S8.

Phylogenetic analysis of the *FOXP3* gene

The evolutionary history of the *FOXP3* gene was inferred by using the Maximum Likelihood method based on the Tamura-Nei model (Tamura and Nei, 1993), and a phylogenic tree was generated with MEGA7 software (Kumar et al., 2016). The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 15 genomic sequences of *FOXP3*, *FOXP3a* and *FOXP3b* gene from different vertebrate organisms: Human_ENSG00000049768, Gorilla_ENSGGOG00000000913, Macaque_ENSMMUG00000008624, Tarsier_ENSTSYG00000035932, Microbat_ENSMLUT00000005714.2, Mouse_ENSMUSG00000039521, Rat_ENSRNOG00000011702, Cow_ENSBTAG00000013279, Dog_ENSCAFG00000015934, Elephant_ENSLAFG00000003504, Opossum_ENSMODG00000009847, Platypus_ENSOANG00000013584, Xenopus_ENSXETG00000031498, Zebrafish_FOXP3a_ENSDARG00000055750, and Zebrafish_FOXP3b_ENSDARG00000078279.

Likewise, the evolutionary history of the *FOX* family gene was inferred by using the Maximum Likelihood method as described above. The analysis involved 12 protein-coding sequences of human FOX family genes (FOXA1: ENST00000250448.3, FOXD1: ENST00000615637.3, FOXE1: ENST00000375123.4, FOXF1: ENST00000262426.6, FOXH1: ENST00000377317.4, FOXN1: ENST00000226247.2, FOXO1: ENST00000379561.6, FOXP1: ENST00000318789.9,

720 FOXP2: ENST00000350908.9, FOXP3: ENSG0000004976, FOXP4: ENST00000373063.7, and
721 RBFOX2: ENST00000449924.6).

722

723 **FOXP3 splicing reporter assays**

724 2.0×10^5 HeLa cells were transfected in 12 well plates with control (NTC: all-stars non-targeting
725 control, Qiagen) or DDX39B-targeting (D11: Hs_BAT1_11, and D13: Hs_BAT1_13, Qiagen)
726 siRNAs using Lipofectamine RNAiMAX Transfection Reagent (Thermo Fisher Scientific). After
727 two days of culture, the Renilla luciferase (RLuc) splicing reporter plasmids were co-transfected
728 with a Firefly luciferase plasmid (pGL3 control-FLuc, Promega) as transfection control. After 24
729 hours, cell lysates were collected for measurements of luciferase activity and RNA isolation.
730 RLuc and FLuc activities were measured by Dual-Luciferase Reporter Assay System (Promega),
731 and the data are presented as RLuc/FLuc.

732 To directly quantify the splicing efficiency of the reporters, RNA was isolated and used for
733 reverse transcription as described above. Spliced and unspliced reporter transcripts were
734 measured by endpoint PCR (Platinum Taq DNA polymerase, Thermo Fisher Scientific) with
735 primers specific to the RLuc coding sequence. PCR amplicons were detected by electrophoresis
736 on 6% non-denaturing polyacrylamide/TBE gels with SYBR Gold Nucleic Acid Gel Stain
737 (Thermo Fisher Scientific). Splicing efficiency was measured by densitometry analysis of the
738 PCR amplicons in image J (Schindelin et al., 2012; Schneider et al., 2012).

739

740 For the splicing-rescue assay, HeLa-Flp-In cells stably expressing either wild-type or mutant
741 (K95A, E197A or D199A) DDX39B *trans*-genes were transfected with control siRNA (NTC) or

DDX39B-targeting siRNA D13, which targets the 3' untranslated region (3'UTR) of DDX39B mRNA. After two days of culture, expression of DDX39B *trans*-genes in NTC- or D13-treated cells was either induced by 1 µg/mL doxycycline or not (doxycycline withheld) to generate the following four conditions: NTC –doxycycline (control), NTC +doxycycline (overexpression), D13 –doxycycline (knockdown) and D13 +doxycycline (rescue). The cells were then transfected with pcDNA3.1-RLuc_FOXP3 Intron 11 and harvested 24 hours after for analyses. DDX39B protein expression was confirmed by western blot, and the splicing efficiency of the reporters was evaluated by endpoint PCR as described above.

Py tract sequence analysis of retained introns

5' SS (from -3 to + 6 nt at the 5' Exon-Intron junction) and 3' SS (from -20 to +3 nt at the 3' Intron-Exon junction) sequences of *FOXP3*, *FOX* family genes and X chromosome genes (2155 genes) were obtained from Human genome assembly (GRCh38.p13) using the UCSC table browser (Karolchik et al., 2004; Kent et al., 2002) and Galaxy platform (Afgan et al., 2018). Sequence logos of the splice sites of *FOX* family genes and X chromosomal genes (total 2155 genes) were generated by WebLogo3 (Crooks et al., 2004). Sequences of 3' SS (from -20 to -3 nt) were defined as poly-pyrimidine tract (Py tract) sequence, and nucleotide probability at each of these positions was determined per gene. The sequence logo for a given gene shows the average at each position of the Py tract from all introns for that particular gene.

For comparison of Py tracts of introns that were DDX39B-sensitive (retained in the knockdown) versus DDX39B-insensitive (unchanged in the knockdown), we compared the 402 introns that

were significantly retained versus 500 events randomly selected from the total unchanged introns. 5' SS and 3' SS sequences for these events were obtained using the Galaxy platform, and MaxEntScore for these were determined by MaxEntScan (Yeo and Burge, 2004). The nucleotide probability at each position of the Py tract region was determined as above.

Statistical analysis

In all figures, error bars represent standard deviation (S.D.). Asterisks denote level of statistical significance: * $p < 0.05$, ** $p < 0.01$, *** $p<0.001$ and **** $p < 0.0001$. Statistical analyses were conducted in Prism8 (GraphPad Software) or SAS (SAS Institute). For the statistical analysis of experiments using $CD4^+$ T cells from individual donors, One-way repeated measures ANOVA was used. To specifically compare each experimental (Sh3 and Sh5) group to control (NTC) after a significant Omnibus f test, Dunnett's test was performed to control family-wise type I error rate due to multiple comparisons. For the experiments other than $CD4^+$ T cells, One-way ANOVA model followed by aforementioned Dunnett's tests were conducted to compare between experimental (Sh3 and Sh5) and control (NTC). For the rescue assay of splicing, One-way ANOVA models followed by Tukey's tests were used to perform pairwise comparisons among all tested conditions. For the sequence analysis of retained introns, outliers of Max entropy scores were identified and removed (ROUT, Q= 0.5%) (Motulsky and Brown, 2006). Then, Mann-Whitney tests or Kruskal-Wallis tests were conducted.

Statistical analysis of RT-qPCR data in $CD4^+$ T cell were carried out using a linear mixed model on the delta CT values (dCT). For each gene the linear mixed model includes treatment [Control, DDX39B Knockdown1 (Sh3), DDX39B Knockdown2 (Sh5)] as fixed effect, and random

787 intercept to account for the heterogeneity of the individuals and correlations among repeated
788 measures from the same individual. The fold-change differences [delta-delta CT (ddCT) values]
789 and their standard error (SE) between treatment groups versus the control were estimated from
790 the model. The p-values were adjusted by False Discovery Rate (FDR) method, which controls
791 the false discovery rate due to multiplicity in hypothesis testing. To facilitate the interpretation
792 using relative expression, the point estimate of the ddCT values were converted to 2^{-ddCT}.
793
794

References

Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Cech, M., Chilton, J., Clements, D., Coraor, N., Gruning, B.A., *et al.* (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res *46*, W537-W544.

Allcock, R.J., Williams, J.H., and Price, P. (2001). The central MHC gene, BAT1, may encode a protein that down-regulates cytokine production. Genes Cells *6*, 487-494.

Bennett, C.L., Christie, J., Ramsdell, F., Brunkow, M.E., Ferguson, P.J., Whitesell, L., Kelly, T.E., Saulsbury, F.T., Chance, P.F., and Ochs, H.D. (2001). The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of FOXP3. Nat Genet *27*, 20-21.

Bhatt, D.M., Pandya-Jones, A., Tong, A.J., Barozzi, I., Lissner, M.M., Natoli, G., Black, D.L., and Smale, S.T. (2012). Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. Cell *150*, 279-290.

Boutz, P.L., Bhutkar, A., and Sharp, P.A. (2015). Detained introns are a novel, widespread class of post-transcriptionally spliced introns. Genes Dev *29*, 63-80.

Briggs, F. (2019). Unraveling susceptibility to multiple sclerosis. Science *365*, 1383-1384.

Brunkow, M.E., Jeffery, E.W., Hjerrild, K.A., Paeper, B., Clark, L.B., Yasayko, S.A., Wilkinson, J.E., Galas, D., Ziegler, S.F., and Ramsdell, F. (2001). Disruption of a new forkhead/winged-helix protein, scurfin, results in the fatal lymphoproliferative disorder of the scurfy mouse. Nat Genet *27*, 68-73.

Chatila, T.A., Blaeser, F., Ho, N., Lederman, H.M., Voulgaropoulos, C., Helms, C., and Bowcock, A.M. (2000). JM2, encoding a fork head-related protein, is mutated in X-linked autoimmunity-allergic disregulation syndrome. J Clin Invest *106*, R75-81.

Consortium, E.P. (2011). A user's guide to the encyclopedia of DNA elements (ENCODE). PLoS Biol *9*, e1001046.

Consortium, G.T., Laboratory, D.A., Coordinating Center -Analysis Working, G., Statistical Methods groups-Analysis Working, G., Enhancing, G.g., Fund, N.I.H.C., Nih/Nci, Nih/Nhgri, Nih/Nimh, Nih/Nida, *et al.* (2017). Genetic effects on gene expression across human tissues. Nature *550*, 204-213.

Cortez, J.T., Montauti, E., Shifrut, E., Gatchalian, J., Zhang, Y., Shaked, O., Xu, Y., Roth, T.L., Simeonov, D.R., Zhang, Y., *et al.* (2020). CRISPR screen in regulatory T cells reveals modulators of Foxp3. Nature *582*, 416-420.

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome Res *14*, 1188-1190.

Degli-Esposti, M.A., Leelayuwat, C., and Dawkins, R.L. (1992). Ancestral haplotypes carry haplotypic and haplospecific polymorphisms of BAT1: possible relevance to autoimmune disease. Eur J Immunogenet *19*, 121-127.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.

Dominguez-Villar, M., and Hafler, D.A. (2018). Regulatory T cells in autoimmune disease. Nat Immunol *19*, 665-673.

838 Evsyukova, I., Bradrick, S.S., Gregory, S.G., and Garcia-Blanco, M.A. (2013). Cleavage and
839 polyadenylation specificity factor 1 (CPSF1) regulates alternative splicing of interleukin 7
840 receptor (IL7R) exon 6. RNA *19*, 103-115.
841 Feng, J., Meyer, C.A., Wang, Q., Liu, J.S., Shirley Liu, X., and Zhang, Y. (2012). GFOLD: a
842 generalized fold change for ranking differentially expressed genes from RNA-seq data.
843 Bioinformatics *28*, 2782-2788.
844 Fleckner, J., Zhang, M., Valcarcel, J., and Green, M.R. (1997). U2AF65 recruits a novel human
845 DEAD box protein required for the U2 snRNP-branchpoint interaction. Genes Dev *11*, 1864-
846 1872.
847 Fletcher, J.M., Lonergan, R., Costelloe, L., Kinsella, K., Moran, B., O'Farrelly, C., Tubridy, N.,
848 and Mills, K.H. (2009). CD39+Foxp3+ regulatory T Cells suppress pathogenic Th17 cells and
849 are impaired in multiple sclerosis. J Immunol *183*, 7602-7610.
850 Galarza-Munoz, G., Briggs, F.B.S., Evsyukova, I., Schott-Lerner, G., Kennedy, E.M., Nyanhete,
851 T., Wang, L., Bergamaschi, L., Widen, S.G., Tomaras, G.D., *et al.* (2017). Human Epistatic
852 Interaction Controls IL7R Splicing and Increases Multiple Sclerosis Risk. Cell *169*, 72-84 e13.
853 Gatfield, D., Le Hir, H., Schmitt, C., Braun, I.C., Kocher, T., Wilm, M., and Izaurralde, E.
854 (2001). The DExH/D box protein HEL/UAP56 is essential for mRNA nuclear export in
855 Drosophila. Curr Biol *11*, 1716-1721.
856 Gavin, M.A., Rasmussen, J.P., Fontenot, J.D., Vasta, V., Manganiello, V.C., Beavo, J.A., and
857 Rudensky, A.Y. (2007). Foxp3-dependent programme of regulatory T-cell differentiation. Nature
858 *445*, 771-775.
859 Georgiev, P., Charbonnier, L.M., and Chatila, T.A. (2019). Regulatory T Cells: the Many Faces
860 of Foxp3. J Clin Immunol *39*, 623-640.
861 Gregory, S.G., Schmidt, S., Seth, P., Oksenberg, J.R., Hart, J., Prokop, A., Caillier, S.J., Ban, M.,
862 Goris, A., Barcellos, L.F., *et al.* (2007). Interleukin 7 receptor alpha chain (IL7R) shows allelic
863 and functional association with multiple sclerosis. Nat Genet *39*, 1083-1091.
864 Hamano, R., Wu, X., Wang, Y., Oppenheim, J.J., and Chen, X. (2015). Characterization of MT-2
865 cells as a human regulatory T cell-like cell line. Cell Mol Immunol *12*, 780-782.
866 Hori, S. (2021). FOXP3 as a master regulator of Treg cells. Nat Rev Immunol *21*, 618-619.
867 Hori, S., Nomura, T., and Sakaguchi, S. (2003). Control of regulatory T cell development by the
868 transcription factor Foxp3. Science *299*, 1057-1061.
869 Huang, C., Liang, D., Tatomer, D.C., and Wilusz, J.E. (2018). A length-dependent evolutionarily
870 conserved pathway controls nuclear export of circular RNAs. Genes Dev *32*, 639-644.
871 International Multiple Sclerosis Genetics, C. (2019). Multiple sclerosis genomic map implicates
872 peripheral immune cells and microglia in susceptibility. Science *365*.
873 Irimia, M., Weatheritt, R.J., Ellis, J.D., Parikshak, N.N., Gonatopoulos-Pournatzis, T., Babor,
874 M., Quesnel-Vallieres, M., Tapial, J., Raj, B., O'Hanlon, D., *et al.* (2014). A highly conserved
875 program of neuronal microexons is misregulated in autistic brains. Cell *159*, 1511-1523.
876 Jamison, S.F., Crow, A., and Garcia-Blanco, M.A. (1992). The spliceosome assembly pathway
877 in mammalian extracts. Mol Cell Biol *12*, 4279-4287.
878 Jensen, T.H., Boulay, J., Rosbash, M., and Libri, D. (2001). The DECD box putative ATPase
879 Sub2p is an early mRNA export factor. Curr Biol *11*, 1711-1715.
880 Josefowicz, S.Z., Lu, L.F., and Rudensky, A.Y. (2012). Regulatory T cells: mechanisms of
881 differentiation and function. Annu Rev Immunol *30*, 531-564.
882 Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D., and Kent,
883 W.J. (2004). The UCSC Table Browser data retrieval tool. Nucleic Acids Res *32*, D493-496.

884 Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler,
885 D. (2002). The human genome browser at UCSC. Genome Res *12*, 996-1006.
886 Kistler, A.L., and Guthrie, C. (2001). Deletion of MUD2, the yeast homolog of U2AF65, can
887 bypass the requirement for sub2, an essential spliceosomal ATPase. Genes Dev *15*, 42-49.
888 Komatsu, N., Okamoto, K., Sawa, S., Nakashima, T., Oh-hora, M., Kodama, T., Tanaka, S.,
889 Bluestone, J.A., and Takayanagi, H. (2014). Pathogenic conversion of Foxp3+ T cells into TH17
890 cells in autoimmune arthritis. Nat Med *20*, 62-68.
891 Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics
892 Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol *33*, 1870-1874.
893 Lehner, B., Semple, J.I., Brown, S.E., Counsell, D., Campbell, R.D., and Sanderson, C.M.
894 (2004). Analysis of a high-throughput yeast two-hybrid system and its use to predict the function
895 of intracellular proteins encoded within the human MHC class III region. Genomics *83*, 153-167.
896 Li, X., Liu, S., Zhang, L., Issaian, A., Hill, R.C., Espinosa, S., Shi, S., Cui, Y., Kappel, K., Das,
897 R., *et al.* (2019). A unified mechanism for intron and exon definition and back-splicing. Nature
898 *573*, 375-380.
899 Liu, W., Putnam, A.L., Xu-Yu, Z., Szot, G.L., Lee, M.R., Zhu, S., Gottlieb, P.A., Kapranov, P.,
900 Gingeras, T.R., Fazekas de St Groth, B., *et al.* (2006). CD127 expression inversely correlates
901 with FoxP3 and suppressive function of human CD4+ T reg cells. J Exp Med *203*, 1701-1711.
902 Loo, C.S., Gatchalian, J., Liang, Y., Leblanc, M., Xie, M., Ho, J., Venkatraghavan, B.,
903 Hargreaves, D.C., and Zheng, Y. (2020). A Genome-wide CRISPR Screen Reveals a Role for
904 the Non-canonical Nucleosome-Remodeling BAF Complex in Foxp3 Expression and Regulatory
905 T Cell Function. Immunity *53*, 143-157 e148.
906 Lundstrom, W., Highfill, S., Walsh, S.T., Beq, S., Morse, E., Kockum, I., Alfredsson, L., Olsson,
907 T., Hillert, J., and Mackall, C.L. (2013). Soluble IL7Ralpha potentiates IL-7 bioactivity and
908 promotes autoimmunity. Proc Natl Acad Sci U S A *110*, E1761-1770.
909 Luo, M.L., Zhou, Z., Magni, K., Christoforides, C., Rappsilber, J., Mann, M., and Reed, R.
910 (2001). Pre-mRNA splicing and mRNA export linked by direct interactions between UAP56 and
911 Aly. Nature *413*, 644-647.
912 Motulsky, H.J., and Brown, R.E. (2006). Detecting outliers when fitting data with nonlinear
913 regression - a new method based on robust nonlinear regression and the false discovery rate.
914 BMC Bioinformatics *7*, 123.
915 Nakata, D., Nakao, S., Nakayama, K., Araki, S., Nakayama, Y., Aparicio, S., Hara, T., and
916 Nakanishi, A. (2017). The RNA helicase DDX39B and its paralog DDX39A regulate androgen
917 receptor splice variant AR-V7 generation. Biochem Biophys Res Commun *483*, 271-276.
918 Newnham, C.M., and Query, C.C. (2001). The ATP requirement for U2 snRNP addition is
919 linked to the pre-mRNA region 5' to the branch site. RNA *7*, 1298-1309.
920 Noval Rivas, M., Burton, O.T., Wise, P., Charbonnier, L.M., Georgiev, P., Oettgen, H.C.,
921 Rachid, R., and Chatila, T.A. (2015). Regulatory T cell reprogramming toward a Th2-cell-like
922 lineage impairs oral tolerance and promotes food allergy. Immunity *42*, 512-523.
923 Roscigno, R.F., Weiner, M., and Garcia-Blanco, M.A. (1993). A mutational analysis of the
924 polypyrimidine tract of introns. Effects of sequence differences in pyrimidine tracts on splicing. J
925 Biol Chem *268*, 11222-11229.
926 Ruskin, B., Zamore, P.D., and Green, M.R. (1988). A factor, U2AF, is required for U2 snRNP
927 binding and splicing complex assembly. Cell *52*, 207-219.

928 Sambucci, M., Gargano, F., De Rosa, V., De Bardi, M., Picozza, M., Placido, R., Ruggieri, S.,
929 Capone, A., Gasperini, C., Matarese, G., *et al.* (2018). FoxP3 isoforms and PD-1 expression by T
930 regulatory cells in multiple sclerosis. Sci Rep *8*, 3674.
931 Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch,
932 S., Rueden, C., Saalfeld, S., Schmid, B., *et al.* (2012). Fiji: an open-source platform for
933 biological-image analysis. Nat Methods *9*, 676-682.
934 Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of
935 image analysis. Nat Methods *9*, 671-675.
936 Schott, G., Galarza-Munoz, G., Trevino, N., Chen, X., Weirauch, M., Gregory, S.G., Bradrick,
937 S.S., and Garcia-Blanco, M.A. (2021). U2AF2 binds IL7R exon 6 ectopically and represses its
938 inclusion. RNA.
939 Schott, G., and Garcia-Blanco, M.A. (2021). MHC Class III RNA Binding Proteins and
940 Immunity. RNA Biol *18*, 640-646.
941 Schumann, K., Raju, S.S., Lauber, M., Kolb, S., Shifrut, E., Cortez, J.T., Skartsis, N., Nguyen,
942 V.Q., Woo, J.M., Roth, T.L., *et al.* (2020). Functional CRISPR dissection of gene networks
943 controlling human regulatory T cell identity. Nat Immunol *21*, 1456-1466.
944 Seraphin, B., and Rosbash, M. (1989). Identification of functional U1 snRNA-pre-mRNA
945 complexes committed to spliceosome assembly and splicing. Cell *59*, 349-358.
946 Shen, H. (2009). UAP56- a key player with surprisingly diverse roles in pre-mRNA splicing and
947 nuclear export. BMB Rep *42*, 185-188.
948 Shen, H., Zheng, X., Shen, J., Zhang, L., Zhao, R., and Green, M.R. (2008). Distinct activities of
949 the DExD/H-box splicing factor hUAP56 facilitate stepwise assembly of the spliceosome. Genes
950 Dev *22*, 1796-1803.
951 Shen, J., Zhang, L., and Zhao, R. (2007). Biochemical characterization of the ATPase and
952 helicase activity of UAP56, an essential pre-mRNA splicing and mRNA export factor. J Biol
953 Chem *282*, 22544-22550.
954 Singh, R., Valcarcel, J., and Green, M.R. (1995). Distinct binding specificities and functions of
955 higher eukaryotic polypyrimidine tract-binding proteins. Science *268*, 1173-1176.
956 Spies, T., Blanck, G., Bresnahan, M., Sands, J., and Strominger, J.L. (1989). A new cluster of
957 genes within the human major histocompatibility complex. Science *243*, 214-217.
958 Subramanian, A., Narayan, R., Corsello, S.M., Peck, D.D., Natoli, T.E., Lu, X., Gould, J., Davis,
959 J.F., Tubelli, A.A., Asiedu, J.K., *et al.* (2017). A Next Generation Connectivity Map: L1000
960 Platform and the First 1,000,000 Profiles. Cell *171*, 1437-1452 e1417.
961 Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A.,
962 Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., *et al.* (2005). Gene set enrichment
963 analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc
964 Natl Acad Sci U S A *102*, 15545-15550.
965 Tamura, K., and Nei, M. (1993). Estimation of the number of nucleotide substitutions in the
966 control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol *10*, 512-526.
967 Tarver, J.E., Dos Reis, M., Mirarab, S., Moran, R.J., Parker, S., O'Reilly, J.E., King, B.L.,
968 O'Connell, M.J., Asher, R.J., Warnow, T., *et al.* (2016). The Interrelationships of Placental
969 Mammals and the Limits of Phylogenetic Inference. Genome Biol Evol *8*, 330-344.
970 Wagner, E.J., and Garcia-Blanco, M.A. (2002). RNAi-mediated PTB depletion leads to
971 enhanced exon definition. Mol Cell *10*, 943-949.
972 Wan, Y.Y., and Flavell, R.A. (2007). Regulatory T-cell functions are subverted and converted
973 owing to attenuated Foxp3 expression. Nature *445*, 766-770.

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. Nature *456*, 470-476.

Wildin, R.S., Ramsdell, F., Peake, J., Faravelli, F., Casanova, J.L., Buist, N., Levy-Lahad, E., Mazzella, M., Goulet, O., Perroni, L., *et al.* (2001). X-linked neonatal diabetes mellitus, enteropathy and endocrinopathy syndrome is the human equivalent of mouse scurfy. Nat Genet *27*, 18-20.

Wu, T., and Fu, X.D. (2015). Genomic functions of U2AF in constitutive and regulated splicing. RNA Biol *12*, 479-485.

Yeo, G., and Burge, C.B. (2004). Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J Comput Biol *11*, 377-394.

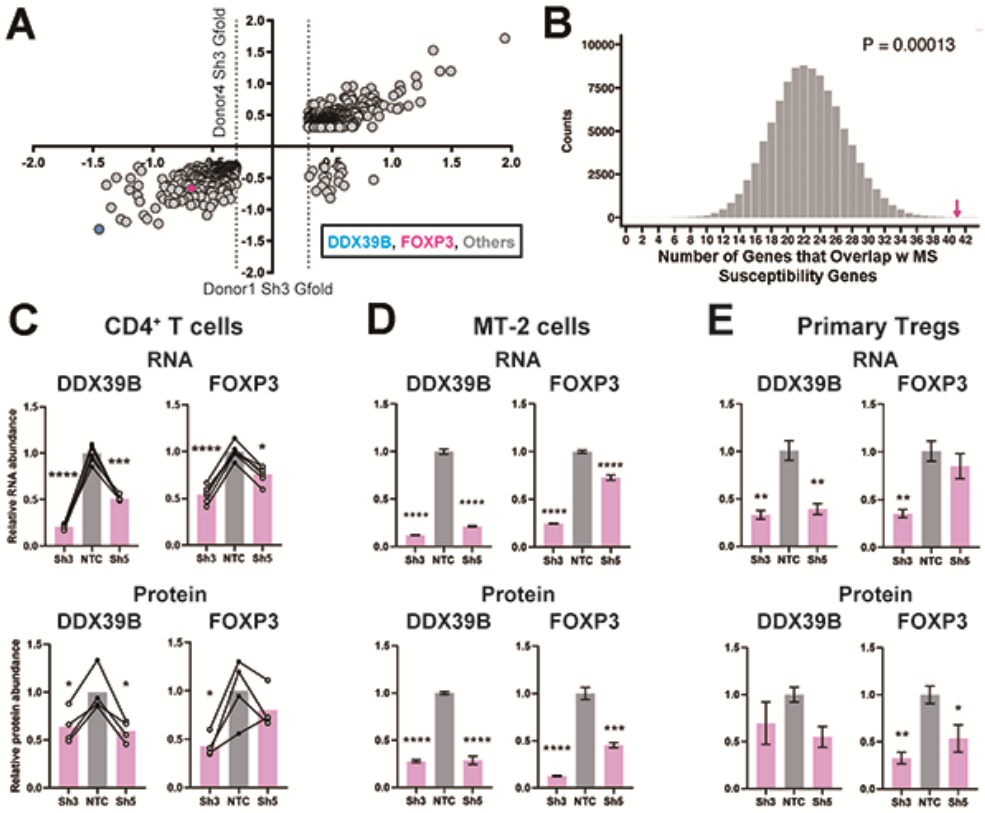
Zamore, P.D., and Green, M.R. (1989). Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. Proc Natl Acad Sci U S A *86*, 9243-9247.

Zamore, P.D., Patton, J.G., and Green, M.R. (1992). Cloning and domain structure of the mammalian splicing factor U2AF. Nature *355*, 609-614.

Zhou, X., Bailey-Bucktrout, S.L., Jeker, L.T., Penaranda, C., Martinez-Llordella, M., Ashby, M., Nakayama, M., Rosenthal, W., and Bluestone, J.A. (2009). Instability of the transcription factor Foxp3 leads to the generation of pathogenic memory T cells in vivo. Nat Immunol *10*, 1000-1007.

Zhou, Z., Sun, B., Huang, S., and Zhao, L. (2019). Roles of circular RNAs in immune regulation and autoimmune diseases. Cell Death Dis *10*, 503.

999 **Figures and Figure Legends**



1000

1001 **Fig. 1. Loss of FOXP3 expression in DDX39B-depleted T cells.** (**A**) RNA abundance changes
1002 between control (NTC) and DDX39B-depleted (Sh3) $CD4^+$ T cells from two healthy individuals
1003 (Donor 1 and Donor 4) identified by RNAseq. Data points for DDX39B (cyan) and FOXP3
1004 (magenta) are indicated. (**B**) Enrichment analysis of MS susceptibility genes in DEGs following
1005 DDX39B depletion. Resampling 100,000 times resulted in a distribution of the number of genes
1006 that overlap by chance, with 23 being the most common result. The observed overlap of 41
1007 (Magenta arrow) demonstrates substantial enrichment (empirical p=0.00013). (**C-E, upper**
1008 **panels**) Levels of DDX39B RNA and FOXP3 RNA, normalized to EEF1A1 RNA levels, after
1009 DDX39B depletion in $CD4^+$ T cells (C) Treg-like MT-2 cells (D), or primary Tregs (E). (**C-E,**
1010 **lower panels**) Levels of DDX39B and FOXP3 relative to Tubulin after DDX39B depletion in

1011 $CD4^+$ T cells (C), MT-2 cells (D), or primary Tregs (E). Connected dots indicate samples from
1012 the same donor. In all figures the error bars indicate standard deviation. *: $p < 0.05$, **: $p < 0.01$,
1013 ***: $p<0.001$ and ****: $p < 0.0001$.

1014 **Legends for supplementary figures associated with Figure 1.**

1015 **Fig. S1A. Transcript level changes upon DDX39B depletion in primary human $CD4^+$ T**
1016 **cells.** DDX39B was depleted in primary $CD4^+$ T cells from two healthy donors (donors 1 and 4)
1017 using DDX39B targeting (Sh3 and Sh5) shRNAs. RNAseq was used to examine transcript level
1018 changes between these cells and those treated with control (NTC) shRNA. The figure illustrates
1019 the comparison of the transcripts whose expression changed significantly (up or down) for each
1020 donor and shRNA combination. Criteria for inclusion were (a) transcripts were required to have
1021 RPKM ≥ 2 and (b) transcript level changes between conditions had |Gfold values| ≥ 0.3 for Sh3
1022 or ≥ 0.1 for Sh5.

1023 **Fig. S1B. Examples of MS susceptibility genes differentially expressed upon DDX39B**
1024 **knockdown.** The heat map shows expression of five genes between control (NTC) and DDX39B
1025 depleted (Sh3) T cells.

1026 **Fig. S1C. Schematic diagram of the enrichment analysis of MS-susceptibility genes among**
1027 **genes whose expression is altered by DDX39B knockdown.** List of odds ratio (OR) of Risk-
1028 SNP and Susceptibility genes were obtained from International Multiple Sclerosis Genetics
1029 Consortium (2019). Detail of the classification is described in the Methods section. The list of
1030 the MS_Pathogenic and MS_Protective genes is shown in Table S3.

Fig. S1D. Enrichment analysis of the MS_Pathogenic or MS_Protective genes. Resampling 100,000 times resulted in a distribution of the number of genes that overlap by chance with experimental overlap indicated by magenta arrow. P value was calculated based on the fraction of instances. Details of the analysis are described in the Methods section.

Fig. S1E. Coverage tracks of reads mapping to the *FOXP3* gene in control and DDX39B-depleted CD4$^+$ T cells. The figure shows coverage of RNAseq reads mapping to the FOXP3 gene in control (NTC) or DDX39B-depleted (Sh3 or Sh5) CD4$^+$ T cells from two healthy donors (Donors 1 and 4).

Fig. S1F. Western blot analysis of FOXP3 expression upon DDX39B depletion in CD4$^+$ T cells. FOXP3 protein abundance was analyzed by western blot in CD4$^+$ T cells from four donors transduced with either control (NTC) or DDX39B targeting (Sh3 or Sh5) shRNAs. The data were normalized to a-tubulin. Densitometry quantification of these data is shown in Fig. 1C.

Fig. S1F Source Data. Two versions of source data files have been uploaded for all gels/blots shown in this figure: (1) the original files of the full raw unedited gels or blots and; (2) figures with the uncropped gels or blots with the relevant bands clearly labelled.

Fig. S1G. Western blot analysis of FOXP3 expression upon DDX39B depletion in MT-2 cells. DDX39B was depleted in MT-2 cells by lentivirus transduction with either non-targeting control (NTC) or DDX39B targeting (Sh3 or Sh5) shRNAs. DDX39B and FOXP3 protein abundance was measured by western blot and normalized to a-tubulin. Densitometry quantification of these data is shown in Fig. 1D.

Fig. S1G Source Data. Two versions of source data files have been uploaded for all gels/blots shown in this figure: (1) the original files of the full raw unedited gels or blots and; (2) figures with the uncropped gels or blots with the relevant bands clearly labelled.

Fig. S1H. Quantification of DDX39B and FOXP3 RNA upon DDX39B depletion in primary Tregs. Human primary Tregs from Donor 8 were transduced with either non-targeting control (NTC) or DDX39B targeting (Sh3 or Sh5) shRNAs. DDX39B and FOXP3 RNA abundance was measured by RT-qPCR and normalized to EEF1A1 RNA expression.. ***: $p<0.001$. The result for donor 7 is shown in Fig. 1E

Fig. S1I. DDX39B and FOXP3 expression upon DDX39B depletion in primary Tregs. DDX39B was depleted in Tregs from two donors (Donors 7 and 8) via lentivirus transduction with either non-targeting control (NTC) or DDX39B targeting (Sh3 or Sh5) shRNAs. DDX39B and FOXP3 protein abundance was measured by western blot and normalized to PTBP1. Quantification of data for Donor 7 is shown in Fig. 1E.

Fig. S1I Source Data. Two versions of source data files have been uploaded for all gels/blots shown in this figure: (1) the original files of the full raw unedited gels or blots and; (2) figures with the uncropped gels or blots with the relevant bands clearly labelled.

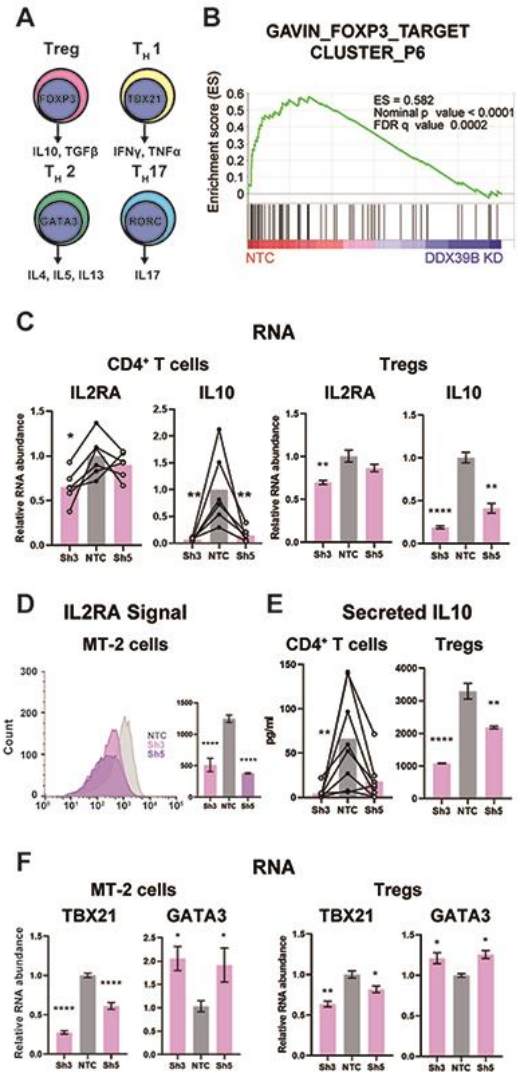


Fig. 2. DDX39B depletion causes loss of Treg gene expression signature. (**A**) Schematic of genes expressed in different T cell lineages: Tregs, T_H17, T_H1, and T_H2 effector cells. (**B**) Enrichment of the FOXP3 target gene-list in the control (NTC) over DDX39B-depleted (Sh3) $CD4^+$ T cells from two individual donors (Donor 1 and Donor 4). (**C**) RNA abundance of IL2RA (CD25) and IL10 relative to EEF1A1 in DDX39B-depleted $CD4^+$ T cells (Left) and primary Tregs (Right). (**D**) FOXP3 expression and surface expression of IL2RA (CD25) by flow cytometry in DDX39B-depleted MT-2 cells. (**E**) IL10 secretion in DDX39B-depleted $CD4^+$ T cells (Left) and Tregs (Right). (**F**) RNA abundance of transcription factors regulating T cell

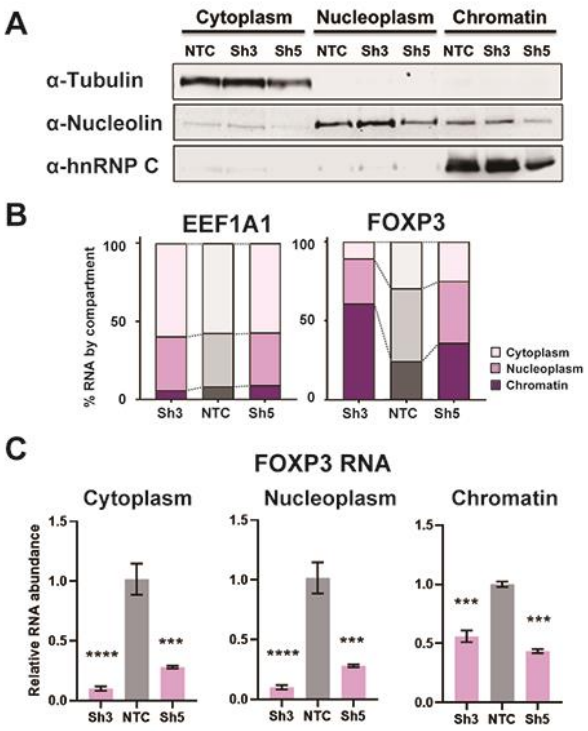
differentiation: TBX21 (T_H1), GATA3 (T_H2) in MT-2 cells (Left) and primary Tregs (Right). *: $p < 0.05$, **: $p < 0.01$, ***: $p<0.001$ and ****: $p < 0.0001$.

Legends for supplementary figures associated with Figure 2.

Fig. S2A. GSEA results of FOXP3-related gene sets enriched in normal over DDX39B depleted cells. GSEA was conducted comparing the transcriptome of control (NTC) and DDX39B-depleted (Sh3) $CD4^+$ T cells. Additional data for the GAVIN_FOXP3_TARGET_CLUSTER_P6 gene set are shown in Fig. 2B.

Fig. S2B. Effect of DDX39B depletion on expression of genes in $CD4^+$ T cells. Expression of genes with known immune function and *GAPDH* (as a control), normalized to *EEF1A1* expression, was measured by RT-qPCR in control (NTC) or in primary $CD4^+$ T cells depleted of DDX39B (Sh3 or Sh5) shRNAs. *: $p < 0.05$, **: $p < 0.01$, ***: $p<0.001$ and ****: $p < 0.0001$.

Fig. S2C. Effect of DDX39B depletion on expression of Treg genes. DDX39B was depleted in T regs from Donors 7 and 8, and expression of seven Treg-related genes (from Gavin et al., 2017) were measured by RT-qPCR and normalized to *EEF1A1* expression. *: $p < 0.05$, **: $p < 0.01$ and ****: $p < 0.0001$.



1096

1097 **Fig. 3. DDX39B depletion disturbs an early step in FOXP3 RNA biogenesis.** (**A**) Protein
1098 abundance of subcellular compartment markers in fractionated control (NTC) or DDX39B-
1099 depleted (Sh3 & Sh5) MT-2 cells. (**B**) Percent EEF1A1 or FOXP3 RNA in subcellular
1100 compartments. (**C**) FOXP3 RNA abundance relative to EEF1A1 in subcellular compartments
1101 upon DDX39B depletion.

1102 **Fig. 3A Source Data.** Two versions of source data files have been uploaded for all gels/blots
1103 shown in this figure: (1) the original files of the full raw unedited gels or blots and; (2) figures
1104 with the uncropped gels or blots with the relevant bands clearly labelled.

1105

1106 **Legends for supplementary figures associated with Figure 3.**

1107

Fig. S3A. Quantification of *FOXP3* retained introns in the chromatin fraction of MT-2 cells. Control (NTC) or DDX39B-depleted (Sh3 or Sh5) MT-2 cells were fractionated into cytoplasm, nucleoplasm and chromatin fractions. Retained *FOXP3* introns were quantified in the chromatin compartment by RT-qPCR using intron-specific primers and normalized to total FOXP3 RNA. **: $p < 0.01$, ***: $p<0.001$.

Fig. S3B. Quantification of *FOXP3* retained introns in primary Tregs. DDX39B was depleted in Tregs from two independent donors (donors 7 and 8) via lentivirus transduction with either DDX39B targeting (Sh3 or Sh5) shRNAs. Retained FOXP3 introns were quantified from total RNA of control cells treated with a non-targeting shRNA (NTC) or DDX39B depleted cells (Sh3 or Sh5) by RT-qPCR using intron-specific primers and normalized to total FOXP3 RNA. *: $p < 0.05$.

Fig. S3C. Alternative splicing events observed upon DDX39B depletion in $CD4^+$ T cells. Splicing analysis of RNAseq data was carried out using Vast-Tools. **i.** Total events changed between control and either of two knockdown conditions (Sh3 & Sh5) in two donors (donor 1 & 4). **ii.** Type of events of Sh3 knockdown in either donor are shown. Alt 5'SS, alternative 5' splice site; Alt 3'SS, alternative 3' splice site. **iii.** Intron retention events in either donor are divided into those showing more retention or less retention with DDX39B knockdown.

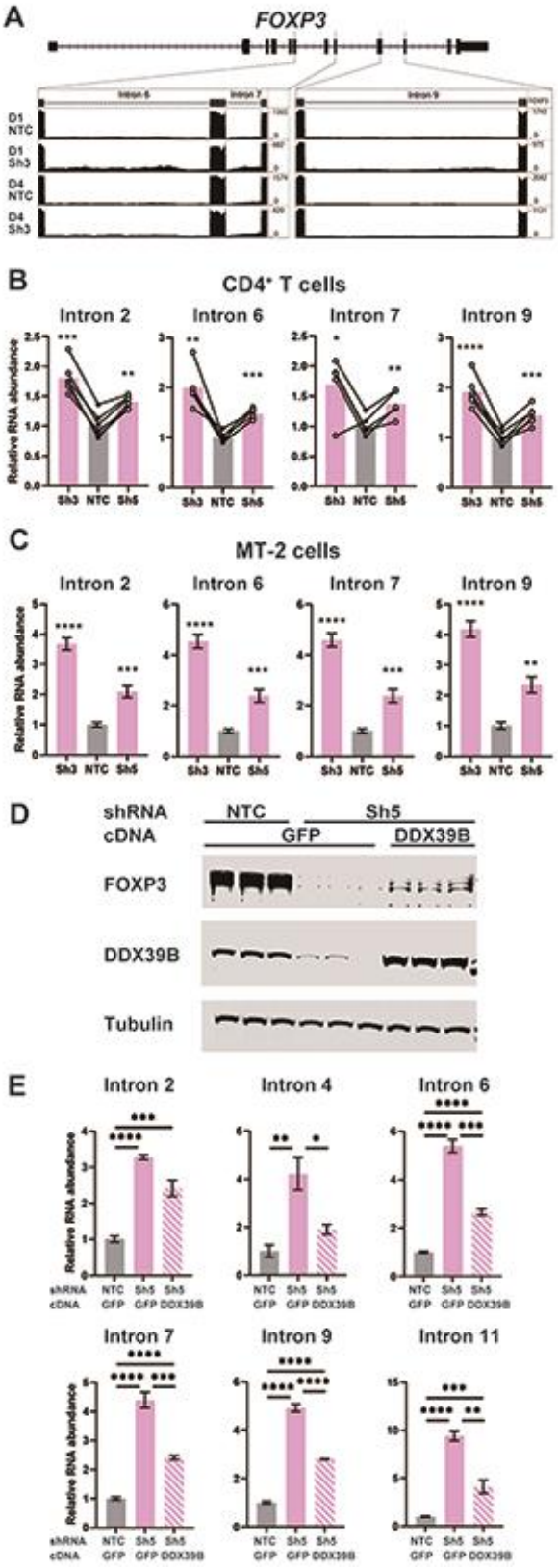


Fig.4. DDX39B depletion triggers the retention of FOXP3 introns. (**A**) RNAseq reads mapping to the FOXP3 genomic region in control (NTC) or DDX39B depleted (Sh3) $CD4^+$ T cells from Donors 1 and 4. Read counts for *FOXP3* introns 6, 7 and 9 are shown on the Y-axis in the two insets. (**B, C**) Abundance of FOXP3 RNA introns relative to total FOXP3 RNA after DDX39B depletion in $CD4^+$ T cells from six donors (B) or MT-2 cells (C). (**D-E**) Rescue of the DDX39B depletion (Sh3) by exogenous expression (GFP or DDX39B) in MT-2 cells. The abundance of FOXP3 and DDX39B relative to Tubulin (D) or FOXP3 RNA introns relative to total FOXP3 RNA (E) are shown.*: $p < 0.05$, **: $p < 0.01$, ***: $p<0.001$ and ****: $p < 0.0001$.

Fig. S4D Source Data. Two versions of source data files have been uploaded for all gels/blots shown in this figure: (1) the original files of the full raw unedited gels or blots and; (2) figures with the uncropped gels or blots with the relevant bands clearly labelled.

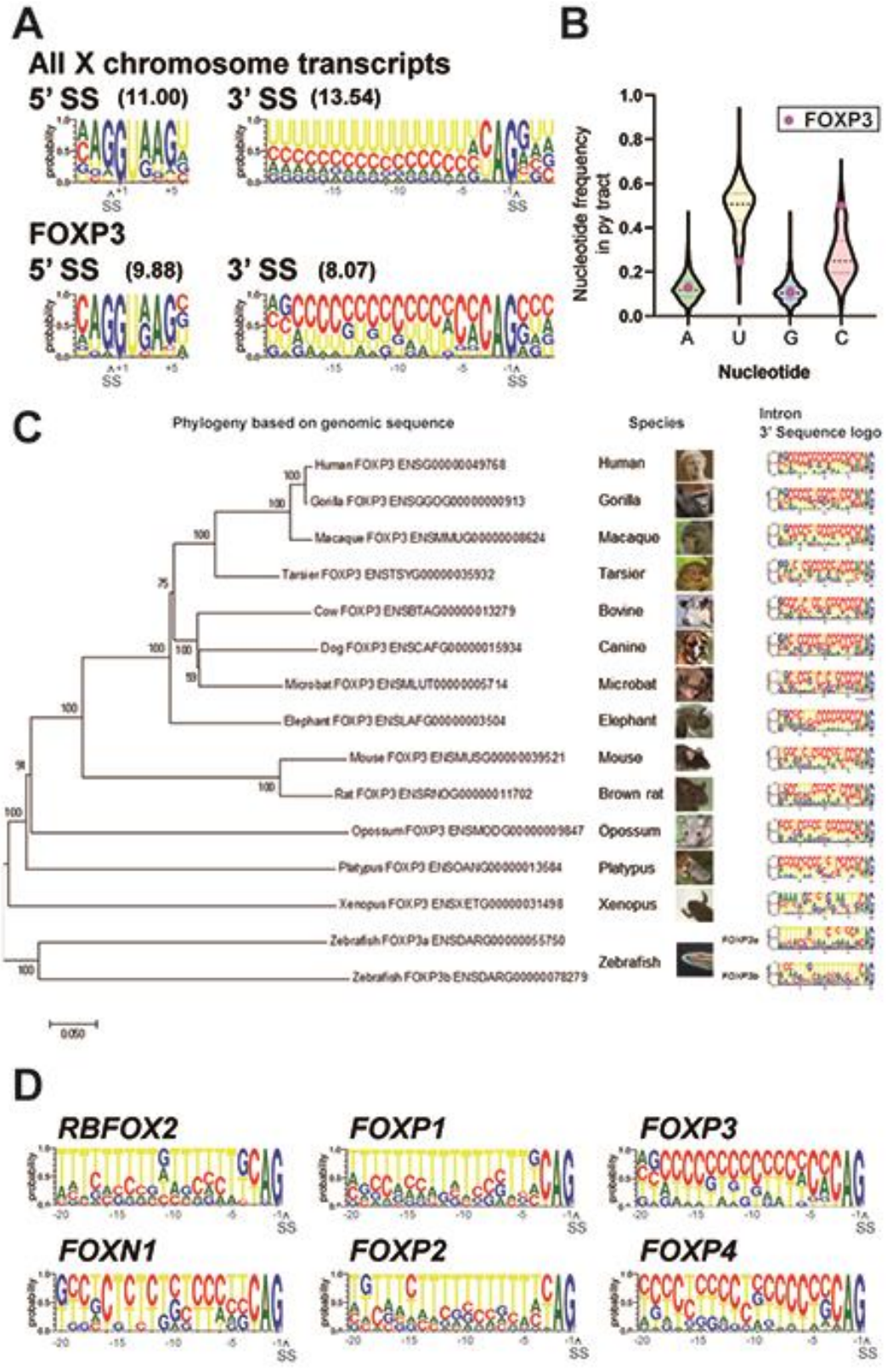


Fig. 5. *FOXP3* introns have conserved C-rich py tracts. (**A**) Sequence logos of the 5' and 3' splice sites (SS) of *FOXP3* versus all other chromosome X genes. 5' or 3' MaxEntScore based on the highest-probability sequence are shown. (**B**) Distribution of the py tract nucleotide frequency among chromosome X genes. Magenta dots indicate *FOXP3*. (**C**) Phylogenic tree and 3' SS sequence logos of the *FOXP3* among 14 species. Zebrafish has two *FOXP3* genes (*FOXP3a* and *FOXP3b*). (**D**) 3' SS sequence logos of FOX family genes. *RBFOX2* is shown as an outgroup.

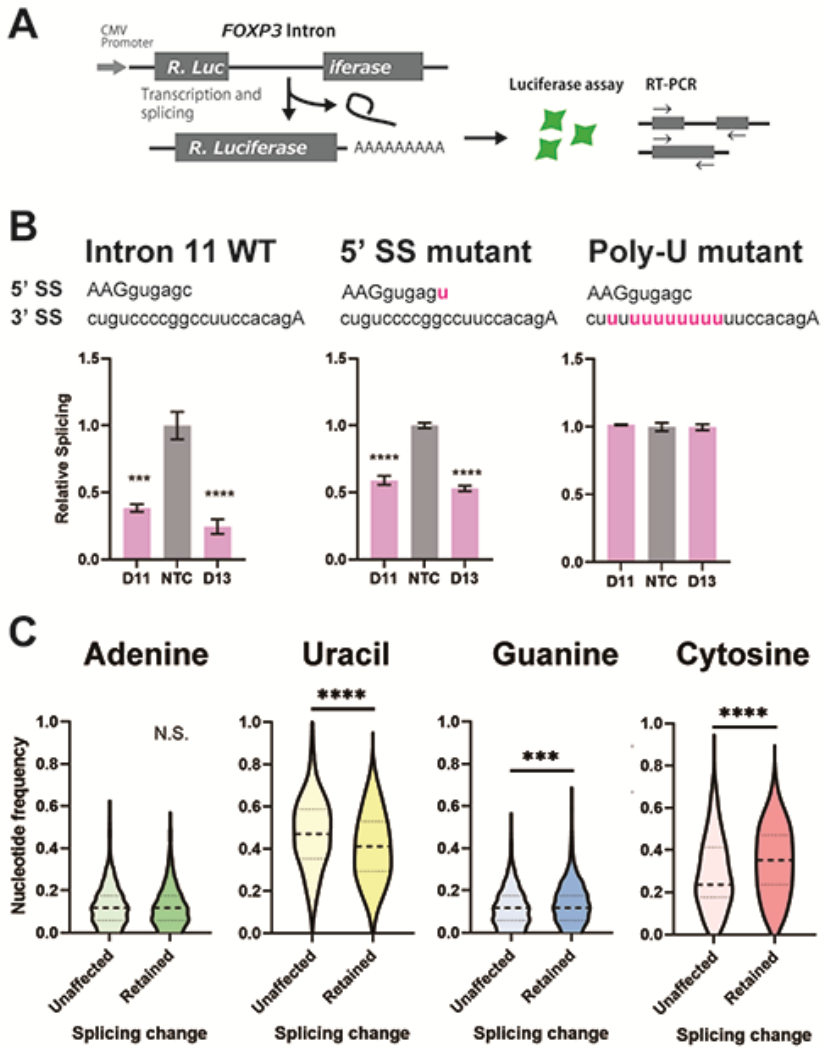


Fig. 6. C-rich py tracts in *FOXP3* introns determine DDX39B sensitivity. (**A**) Schematic diagram of the splicing reporter. (**B**) Relative splicing efficiency of the wild-type (WT) or mutant (5' SS mutant and Poly-U mutant) FOXP3 intron 11 splicing reporters after DDX39B depletion with D11 and D13 siRNAs. The sequence of 5' splice sites (5' SS) for these introns is shown indicating the three upstream exonic residues in capital letters and the six downstream intronic residues in minuscule letters. The sequence of the py tracts and 3' splice sites (3' SS) for these introns is shown indicating the intronic residues in minuscule letters and a single exonic residue in capital letters. (**C**) Nucleotide frequency in the py tract of introns that are insensitive

(unaffected) or sensitive (retained) to DDX39B depletion. *: $p < 0.05$, **: $p < 0.01$, ***: $p<0.001$ and ****: $p < 0.0001$.

Legends for supplementary figures associated with Figure 6.

Fig. S4A. Luciferase activity of the splicing reporters. Control (NTC) or DDX39B-depleted (D11 or D13 siRNAs) HeLa cells were co-transfected with a Firefly luciferase reporter (transfection control, FLuc) and Renilla luciferase (RLuc) splicing reporters with no intron, human *ßglobin* (HGB1) intron 2, or *FOXP3* introns 7, 9 or 11. Relative splicing efficiency of reporters was inferred by measuring luciferase activity (RLuc/FLuc) normalized to NTC. *: $p < 0.05$, **: $p < 0.01$, ***: $p<0.001$ and ****: $p < 0.0001$.

Fig. S4B. RT-PCR analysis of splicing efficiency of wild-type or mutant *FOXP3* Intron 11 RLuc reporters. Control (NTC) or DDX39B-depleted (D11 or D13) HeLa cells were transfected with *FOXP3* intron 11 RLuc reporters (WT, 5' SS consensus or 3' Poly-U mutant), and splicing efficiency was directly measured by endpoint RT-PCR. Quantification of the gels is shown in Fig. 6B.

Fig. S4B Source Data. Two versions of source data files have been uploaded for all gels/blots shown in this figure: (1) the original files of the full raw unedited gels or blots and; (2) figures with the uncropped gels or blots with the relevant bands clearly labelled.

Fig. S4C. RT-PCR analysis of splicing efficiency of wild-type or mutant *FOXP3* Intron 7 RLuc reporters. Control (NTC) or DDX39B-depleted (D11 or D13) HeLa cells were

transfected with *FOXP3* intron 7 RLuc reporters (WT, 5' SS consensus or 3' Poly-U mutant), and splicing efficiency was directly measured by endpoint RT-PCR. Quantification is shown on the right for the lower two panels; given the very low level of splicing with intron 7 WT reporter it could not be accurately quantified. ***: p<0.001 and ****: p < 0.0001.

Fig. S4C Source Data. Two versions of source data files have been uploaded for all gels/blots shown in this figure: (1) the original files of the full raw unedited gels or blots and; (2) figures with the uncropped gels or blots with the relevant bands clearly labelled.

Fig. S4D. Comparison of Max entropy splice sites (SS) score between unaffected and DDX39B-sensitive introns. Max entropy score of 5' SS and 3' SS of unaffected (500 randomly selected) and introns with increased retention upon DDX39B depletion (397 events) was determined using *MaxEntScan* (Yeo G and Burge CB (2004). Dashed lines within the violin plots denote median values. *: p < 0.05.

Fig. S4E. Comparison of Max entropy splice site (SS) scores (top panel) and nucleotide composition of py tracts (bottom panel) of unaffected introns or introns that were more retained in CD4$^+$ T cells from either donor 1 or 4 depleted of DDX39B by treatment with Sh3 or Sh5. *: p < 0.05, **: p < 0.01, ***: p<0.001 and ****: p < 0.0001.

Fig. S4F. Retention of *FOXP1* introns is not increased upon DDX39B depletion in MT-2 cells. *FOXP1* transcripts including introns 11, 14, or 19 were quantified by intron-specific RT-qPCR in control (NTC) and DDX39B-depleted (Sh3 or Sh5) MT-2 cells and normalized to total *FOXP1* transcripts. *: p < 0.05.

Fig. S4G. RT-PCR analysis of splicing efficiency of wild-type or mutant *FOXP1* Intron 19 RLuc reporters. Control (NTC) or DDX39B-depleted (D13) HeLa cells were transfected with *FOXP3* or *FOXP1* RLuc reporters. Splicing efficiency was directly measured by endpoint RT-PCR. Quantification is shown on the right. **: $p<0.01$ and ***: $p<0.001$.

Fig. S4G Source Data. Two versions of source data files have been uploaded for all gels/blots shown in this figure: (1) the original files of the full raw unedited gels or blots and; (2) figures with the uncropped gels or blots with the relevant bands clearly labelled.

Fig. S4H. Comparison of py tract composition between unaffected and less-retained introns. Nucleotide frequency of py tracts of unaffected (500 randomly selected) and those that are less-retained upon DDX39B knockdown (387) was calculated. Dashed lines denote median values. $p < 0.05$, **: $p < 0.01$, ***: $p<0.001$ and ****: $p < 0.0001$.

Fig. S4I. Detection of intron retention events in X chromosome genes with C-rich or U-rich py tracts. Introns of the transcripts with C-rich **(top)** or U-rich **(bottom)** py tract were quantified by intron-specific RT-qPCR in control (NTC) and DDX39B-depleted (Sh3 or Sh5) MT-2 cells and normalized to their corresponding total transcripts. *: $p < 0.05$, **: $p < 0.01$, ***: $p<0.001$ and ****: $p < 0.0001$.

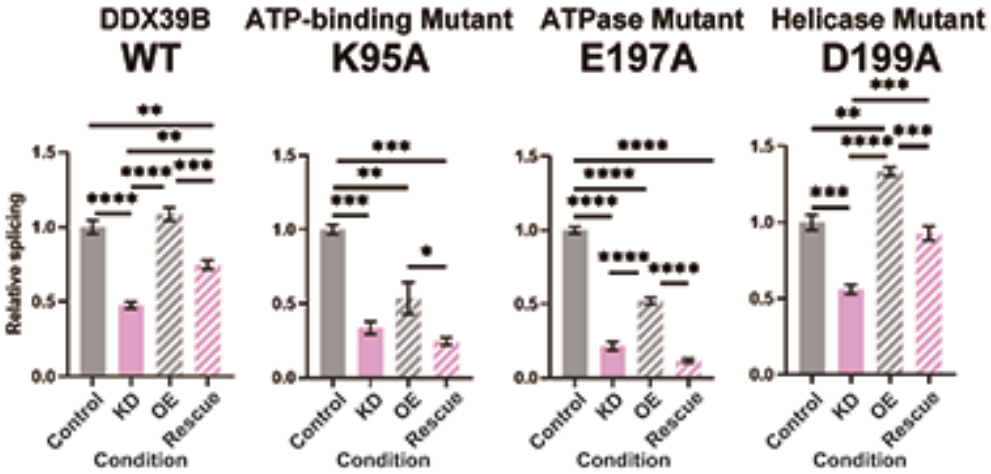


Fig. 7. *FOXP3* intron 11 splicing requires DDX39B ATPase activity but not its helicase activity. Rescue of the splicing of FOXP3 intron 11 reporter by induced expression of wild type (WT) or mutant (K95A, E197A and D199A) DDX39B. KD and OE indicate knockdown and over-expression, respectively. *: $p < 0.05$, **: $p < 0.01$, ***: $p<0.001$ and ****: $p < 0.0001$.

Legends for supplementary figures associated with Figure 6.

Fig. S5. Expression of WT or mutant DDX39B in stable cell lines. Stable HeLa cell lines with inducible expression of WT or mutant DDX39B *trans*-genes were transfected with control (NTC) or DDX39B (D13) siRNAs, and expression of a siRNA-resistant DDX39B transgene was induced with Doxycycline. DDX39B protein expression was quantified by western blot (left). Sanger sequencing of DDX39B RT-PCR amplicons from rescue with mutant DDX39B *trans*-genes (D13 + Doxycycline) shows preferential expression of DDX39B *trans*-genes under rescue conditions (right).

Fig. S5 Source Data. Two versions of source data files have been uploaded for all gels/blots shown in this figure: (1) the original files of the full raw unedited gels or blots and; (2) figures with the uncropped gels or blots with the relevant bands clearly labelled.

Fig. S1. DDX39 controls FOXP3 expression.

(Related to Figure 1 and to Tables S1, S2 and S3)

Fig. S1A

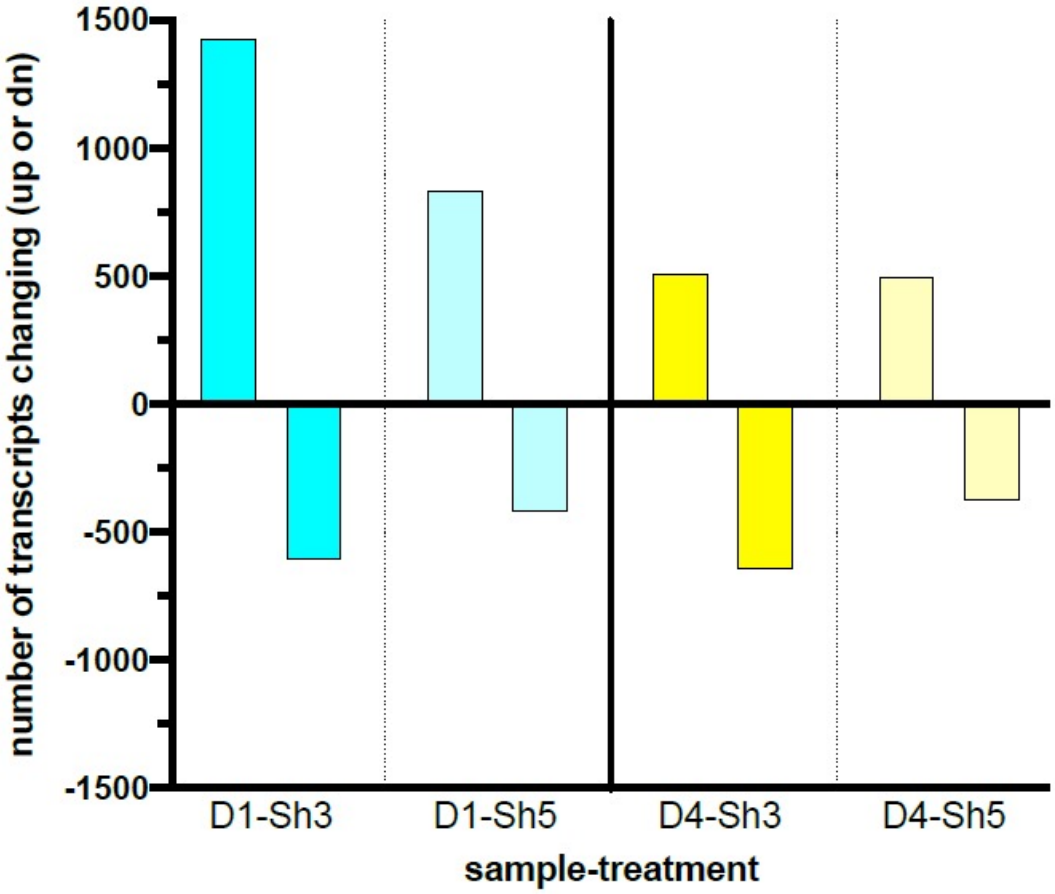


Fig. S1A. Transcript level changes upon DDX39B depletion in primary human CD4⁺ T cells. DDX39B was depleted in primary CD4⁺ T cells from two healthy donors (donors 1 and 4) using DDX39B targeting (Sh3 and Sh5) shRNAs. RNAseq was used to examine transcript level changes between these cells and those treated with control (NTC) shRNA. The figure illustrates the comparison of the transcripts whose expression changed significantly (up or down) for each donor and shRNA combination. Criteria for inclusion were (a) transcripts were required to have RPKM ≥ 2 and (b) transcript level changes between conditions had |Gfold values| ≥ 0.3 for Sh3 or ≥ 0.1 for Sh5.

Fig. S1B

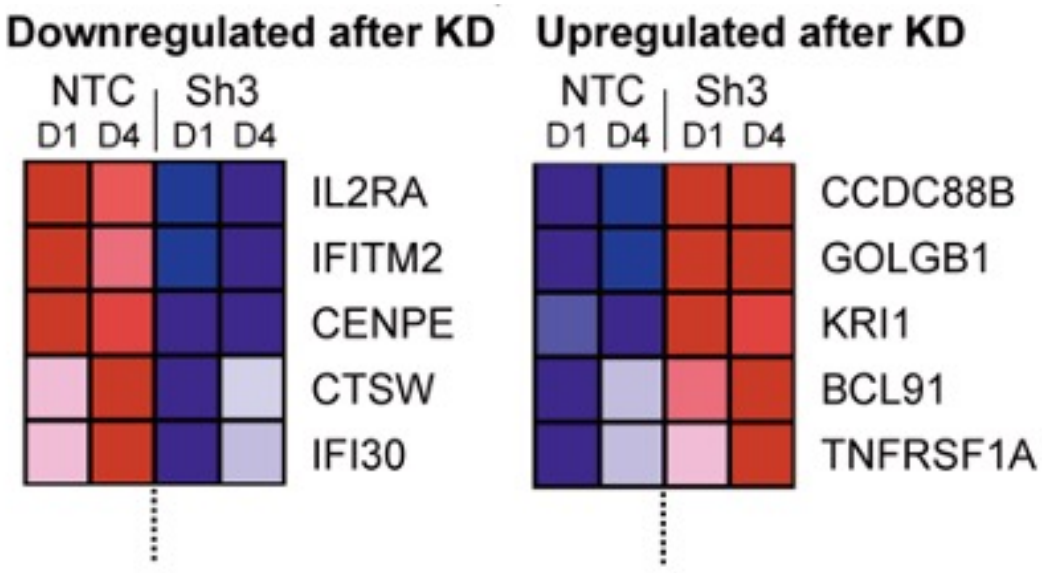


Fig. S1B. Examples of MS susceptibility genes differentially expressed upon DDX39B knockdown. The heat map shows expression of five genes between control (NTC) and DDX39B depleted (Sh3) T cells.

Fig. S1C

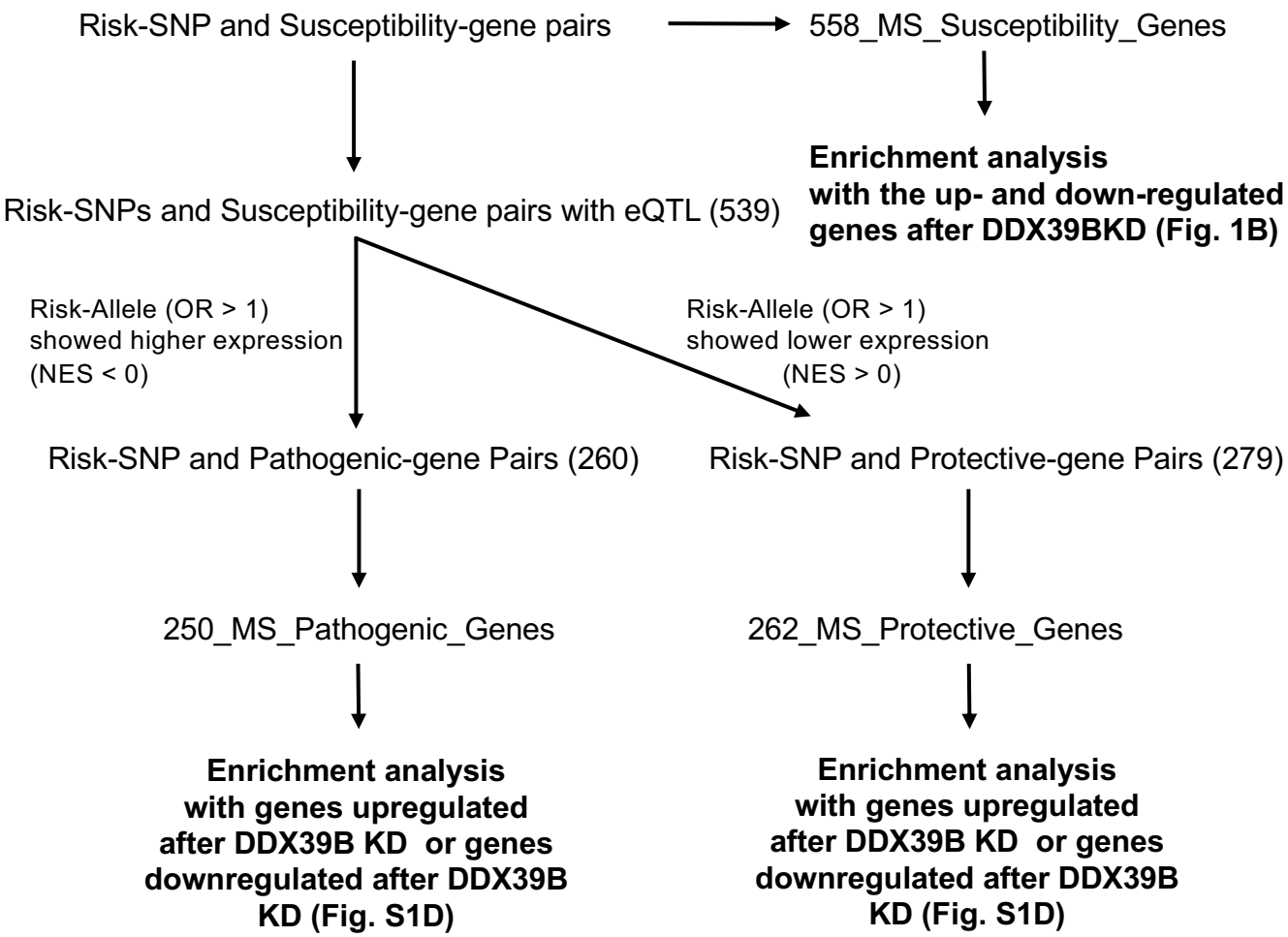


Fig. S1C. Schematic diagram of the enrichment analysis of MS-susceptibility genes among genes whose expression is altered by DDX39B knockdown. List of odds ratio (OR) of Risk-SNP and Susceptibility genes were obtained from International Multiple Sclerosis Genetics Consortium (2019). Detail of the classification is described in the Methods section. The list of the MS_Pathogenic and MS_Protective genes is shown in Table S3.

Fig. S1D

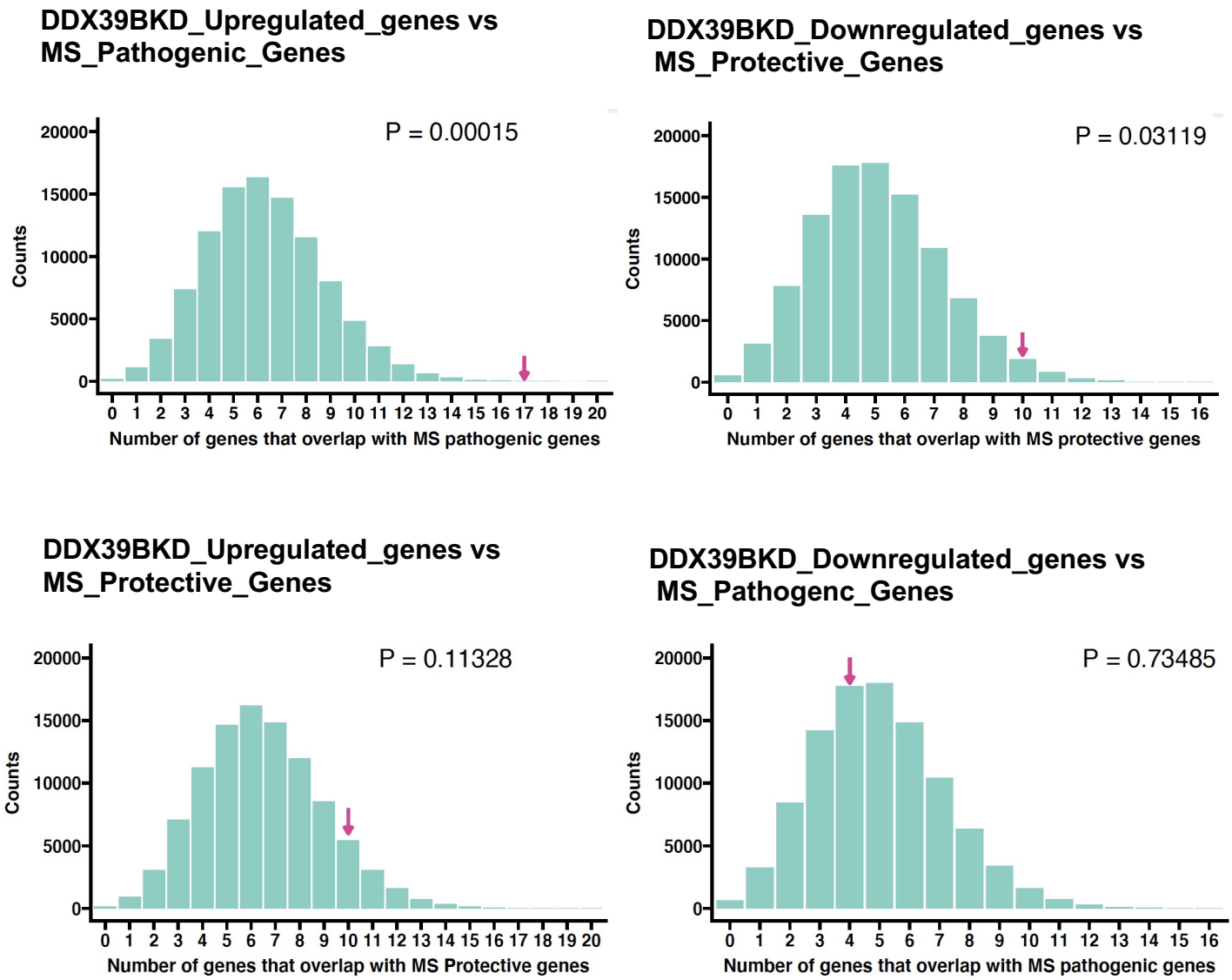


Fig. S1D. Enrichment analysis of the MS_Pathogenic or MS_Protective genes. Resampling 100,000 times resulted in a distribution of the number of genes that overlap by chance with experimental overlap indicated by magenta arrow. P value was calculated based on the fraction of instances. Details of the analysis are described in the Methods section.

Fig. S1E

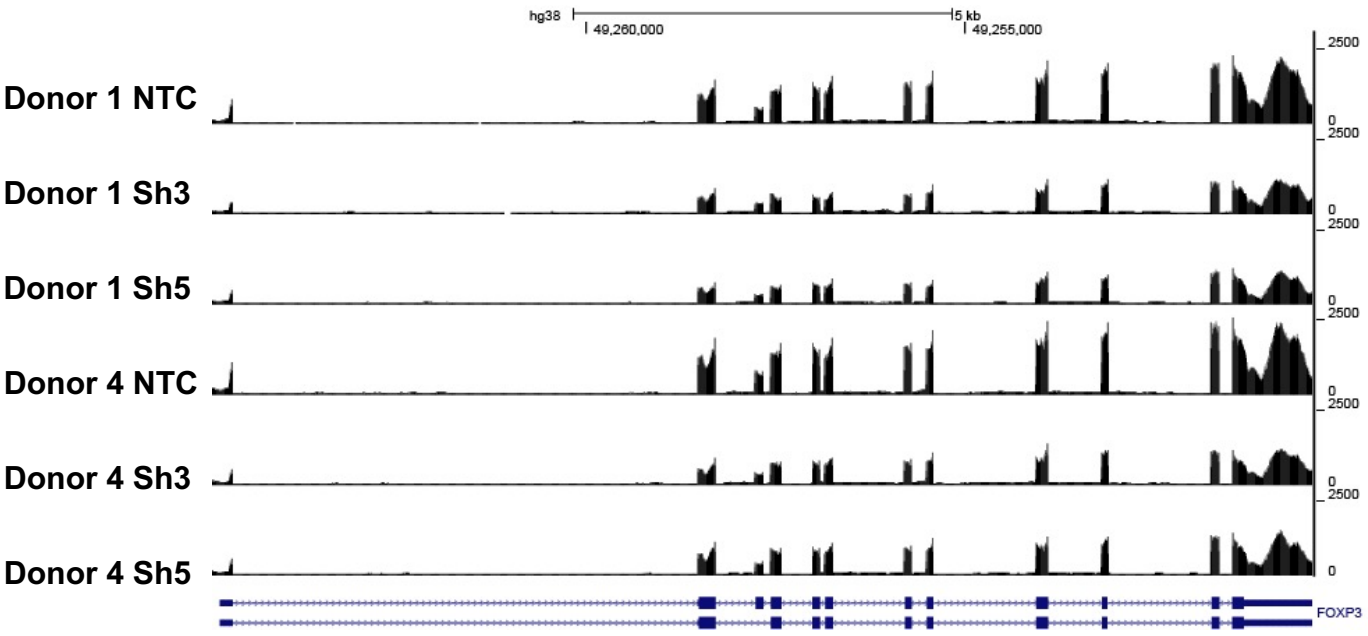


Fig. S1E. Coverage tracks of reads mapping to the *FOXP3* gene in control and DDX39B-depleted CD4⁺ T cells. The figure shows coverage of RNAseq reads mapping to the FOXP3 gene in control (NTC) or DDX39B-depleted (Sh3 or Sh5) CD4⁺ T cells from two healthy donors (Donors 1 and 4).

Fig. S1F

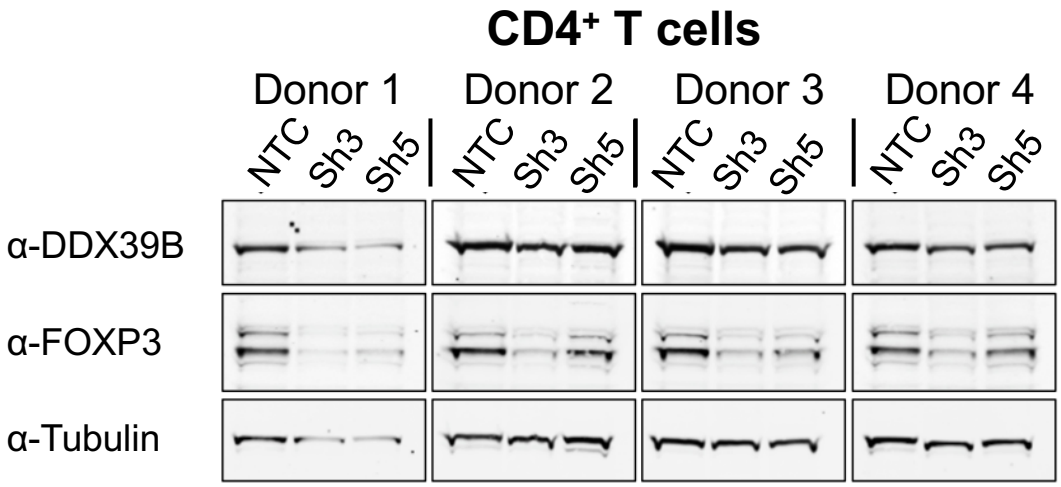


Fig. S1F. Western blot analysis of FOXP3 expression upon DDX39B depletion in CD4⁺ T cells. FOXP3 protein abundance was analyzed by western blot in CD4⁺ T cells from four donors transduced with either control (NTC) or DDX39B targeting (Sh3 or Sh5) shRNAs. The data were normalized to α-tubulin. Densitometry quantification of these data is shown in Fig. 1C.

Fig. S1G

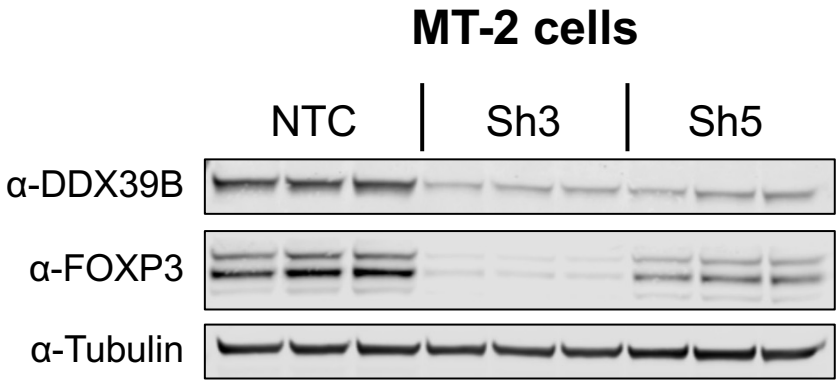


Fig. S1G. Western blot analysis of FOXP3 expression upon DDX39B depletion in MT-2 cells. DDX39B was depleted in MT-2 cells by lentivirus transduction with either non-targeting control (NTC) or DDX39B targeting (Sh3 or Sh5) shRNAs. DDX39B and FOXP3 protein abundance was measured by western blot and normalized to α-tubulin. Densitometry quantification of these data is shown in Fig. 1D.

Fig. S1H

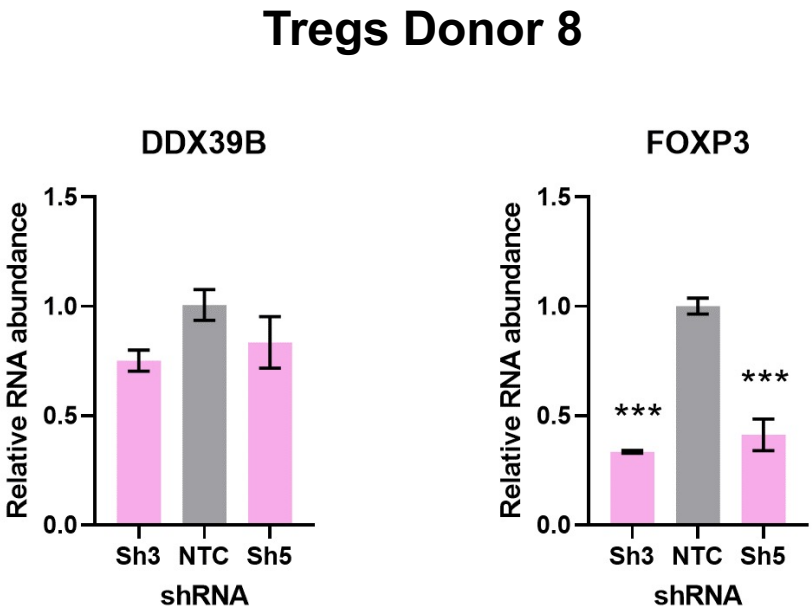


Fig. S1H. Quantification of DDX39B and FOXP3 RNA upon DDX39B depletion in primary Tregs. Human primary Tregs from Donor 8 were transduced with either non-targeting control (NTC) or DDX39B targeting (Sh3 or Sh5) shRNAs. DDX39B and FOXP3 RNA abundance was measured by RT-qPCR and normalized to EEF1A1 RNA expression.. ***: $p<0.001$. The result for donor 7 is shown in Fig. 1E

Fig. S1I

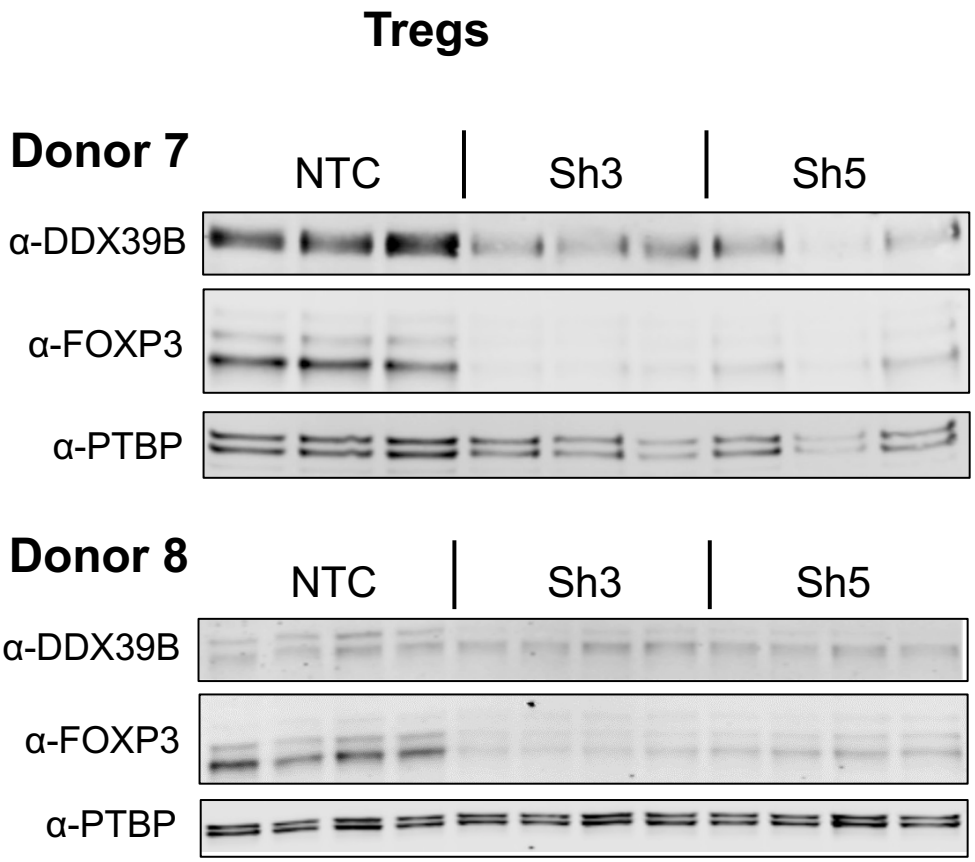


Fig. S1I. DDX39B and FOXP3 expression upon DDX39B depletion in primary Tregs. DDX39B was depleted in Tregs from two donors (Donors 7 and 8) via lentivirus transduction with either non-targeting control (NTC) or DDX39B targeting (Sh3 or Sh5) shRNAs. DDX39B and FOXP3 protein abundance was measured by western blot and normalized to PTBP1. Quantification of data for donor 7 is shown in Fig. 1E.

Fig. S2. FOXP3 pathways affected by DDX39B depletion.

Related to Fig. 2 and Table S3,

Fig. S2A

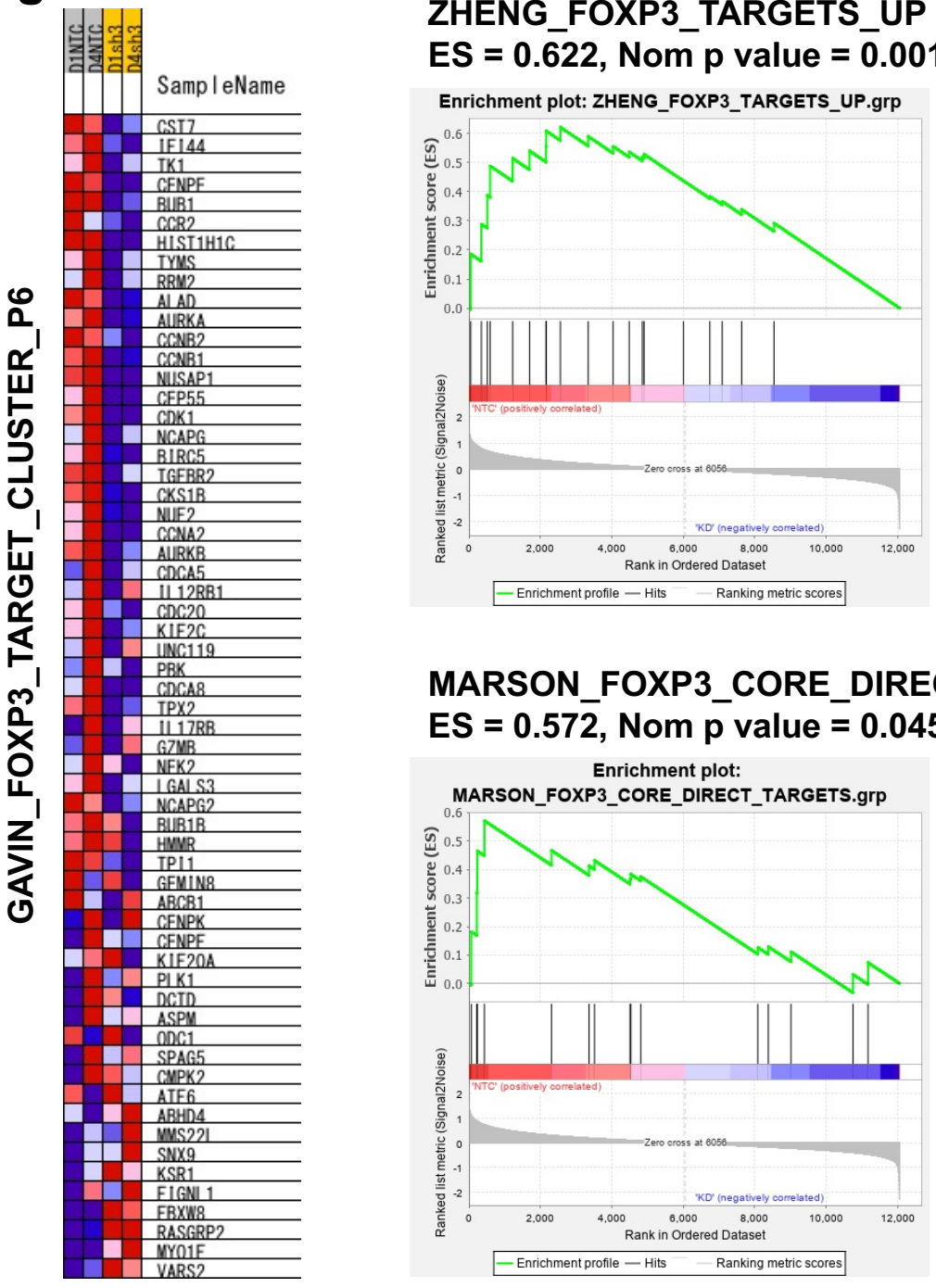


Fig. S2A. GSEA results of FOXP3-related gene sets enriched in normal over DDX39B depleted cells. GSEA was conducted comparing the transcriptome of control (NTC) and DDX39B-depleted (Sh3) CD4$^+$ T cells. Additional data for the GAVIN_FOXP3_TARGET_CLUSTER_P6 gene set are shown in Fig. 2B.

Fig. S2B

CD4+ T cells

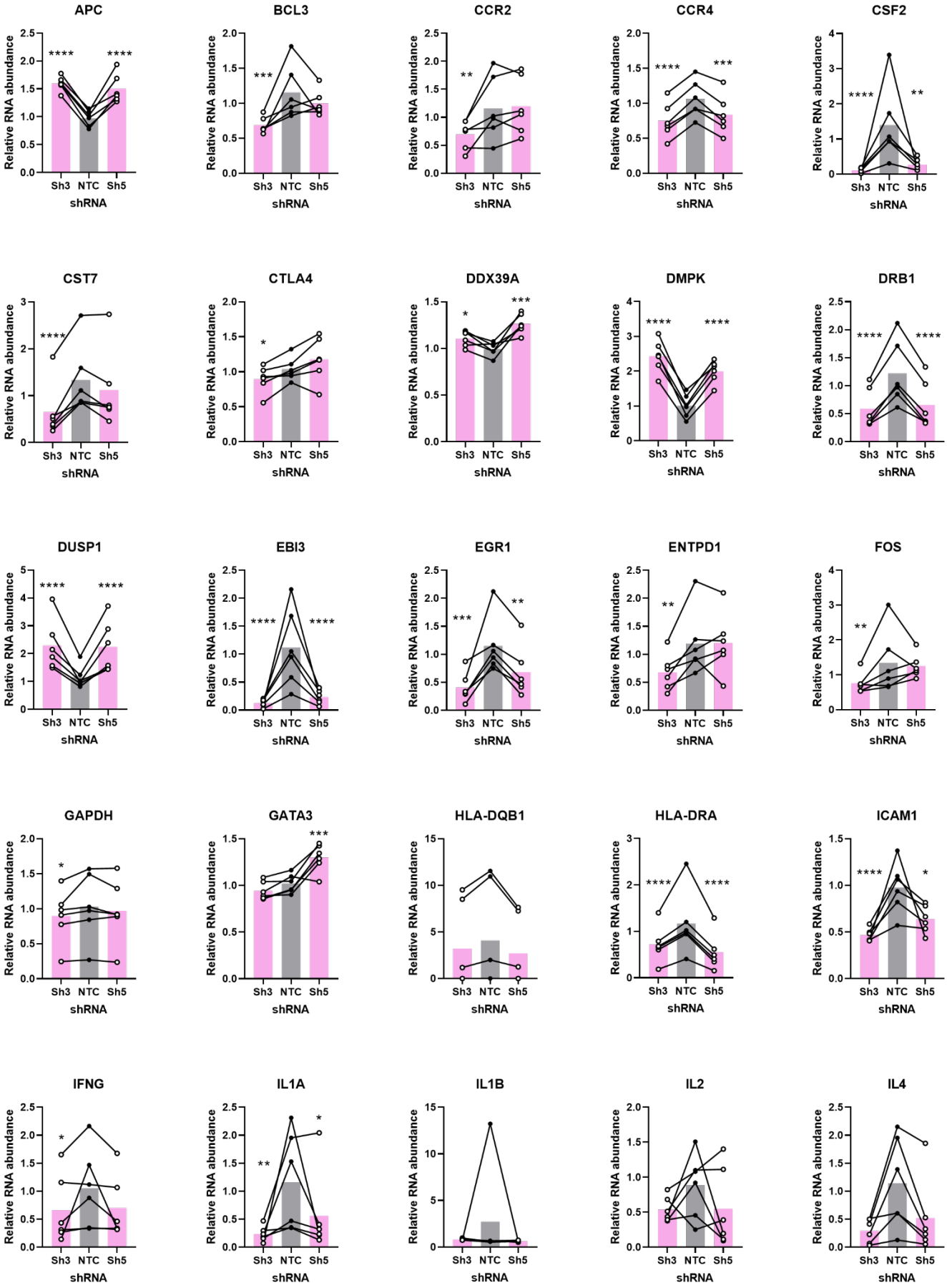


Fig. S2B (cont)

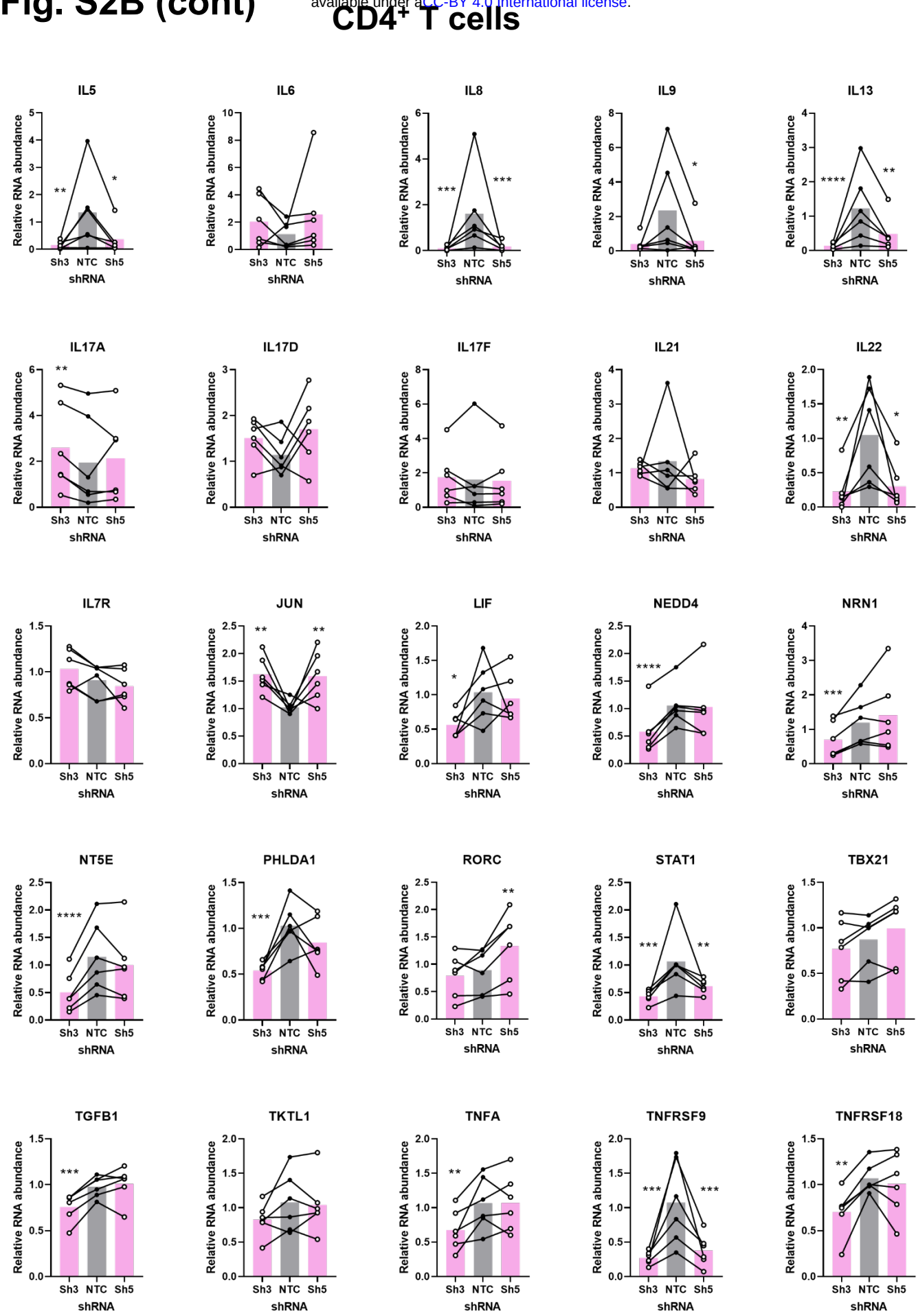


Fig. S2B. Effect of DDX39B depletion on expression of genes in CD4$^+$ T cells. Expression of genes with known immune function and *GAPDH* (as a control), normalized to *EEF1A1* expression, was measured by RT-qPCR in control (NTC) or in primary CD4$^+$ T cells depleted of DDX39B (Sh3 or Sh5) shRNAs. *: $p < 0.05$, **: $p < 0.01$, ***: $p<0.001$ and ****: $p < 0.0001$.

Fig. S2C

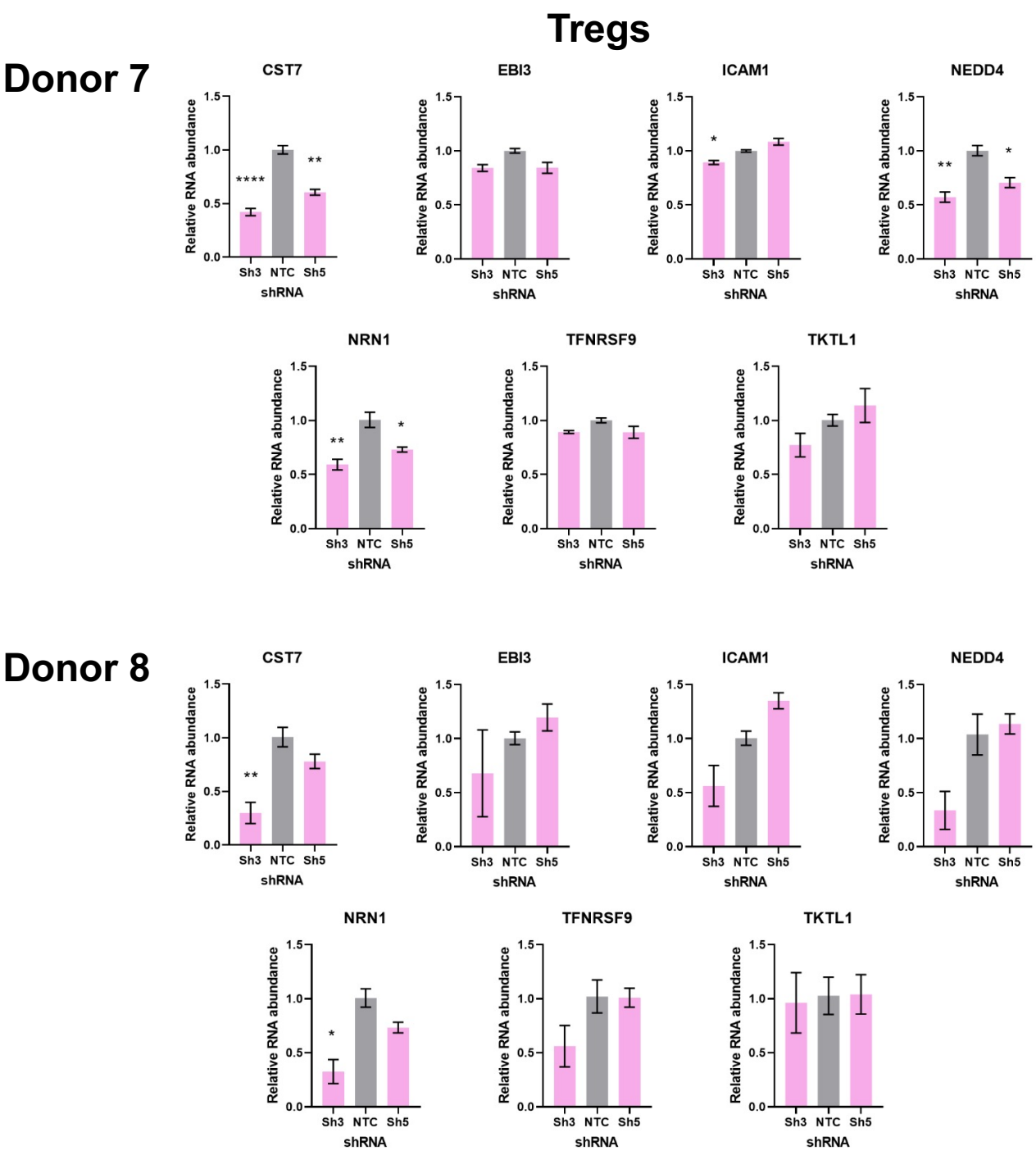


Fig. S2C. Effect of DDX39B depletion on expression of Treg genes. DDX39B was depleted in T regs from Donors 7 and 8, and expression of seven Treg-related genes (from Gavin et al., 2017) were measured by RT-qPCR and normalized to *EEF1A1* expression. *: $p < 0.05$, **: $p < 0.01$ and ****: $p < 0.0001$.

Fig. S3. Low DDX39B causes intron retention.

Related to Figure 4.

Fig. S3A

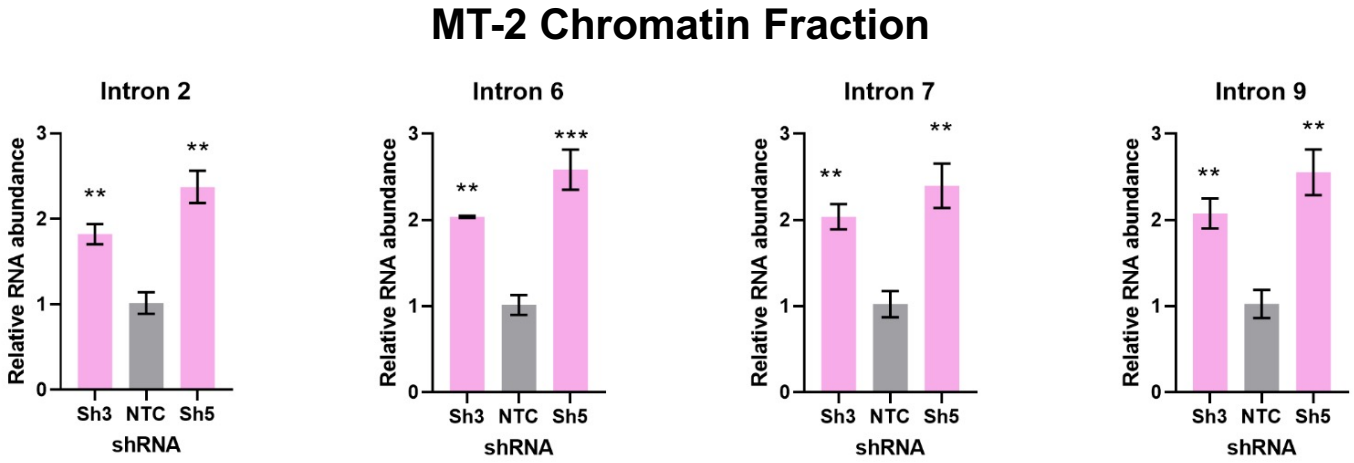


Fig. S3A. Quantification of *FOXP3* retained introns in the chromatin fraction of MT-2 cells. Control (NTC) or DDX39B-depleted (Sh3 or Sh5) MT-2 cells were fractionated into cytoplasm, nucleoplasm and chromatin fractions. Retained *FOXP3* introns were quantified in the chromatin compartment by RT-qPCR using intron-specific primers and normalized to total FOXP3 RNA. **: $p < 0.01$, ***: $p<0.001$.

Fig. S3B

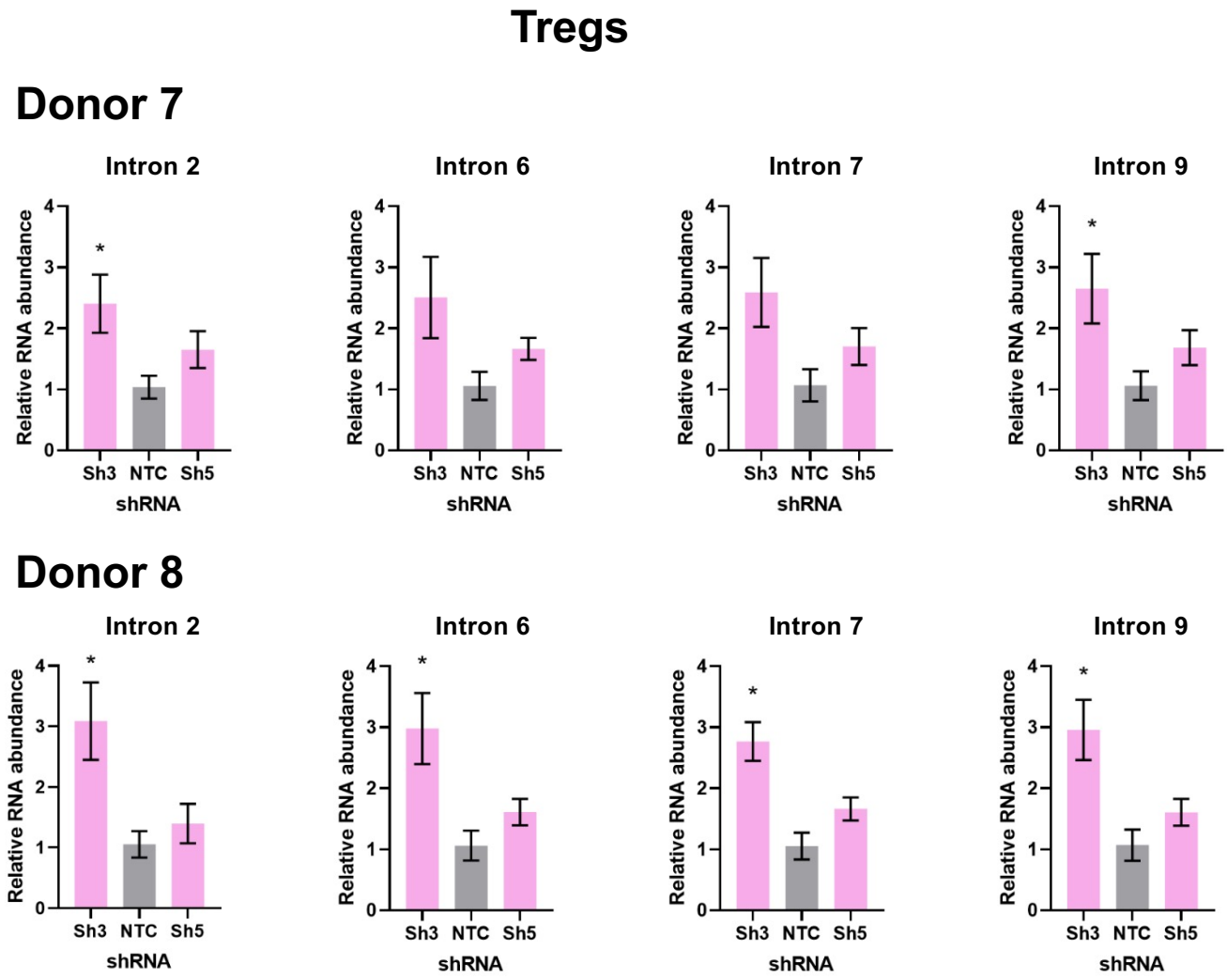


Fig. S3B. Quantification of *FOXP3* retained introns in primary Tregs. DDX39B was depleted in Tregs from two independent donors (donors 7 and 8) via lentivirus transduction with either DDX39B targeting (Sh3 or Sh5) shRNAs. Retained FOXP3 introns were quantified from total RNA of control cells treated with a non-targeting shRNA (NTC) or DDX39B depleted cells (Sh3 or Sh5) by RT-qPCR using intron-specific primers and normalized to total FOXP3 RNA. *: $p < 0.05$.

Fig. S3C

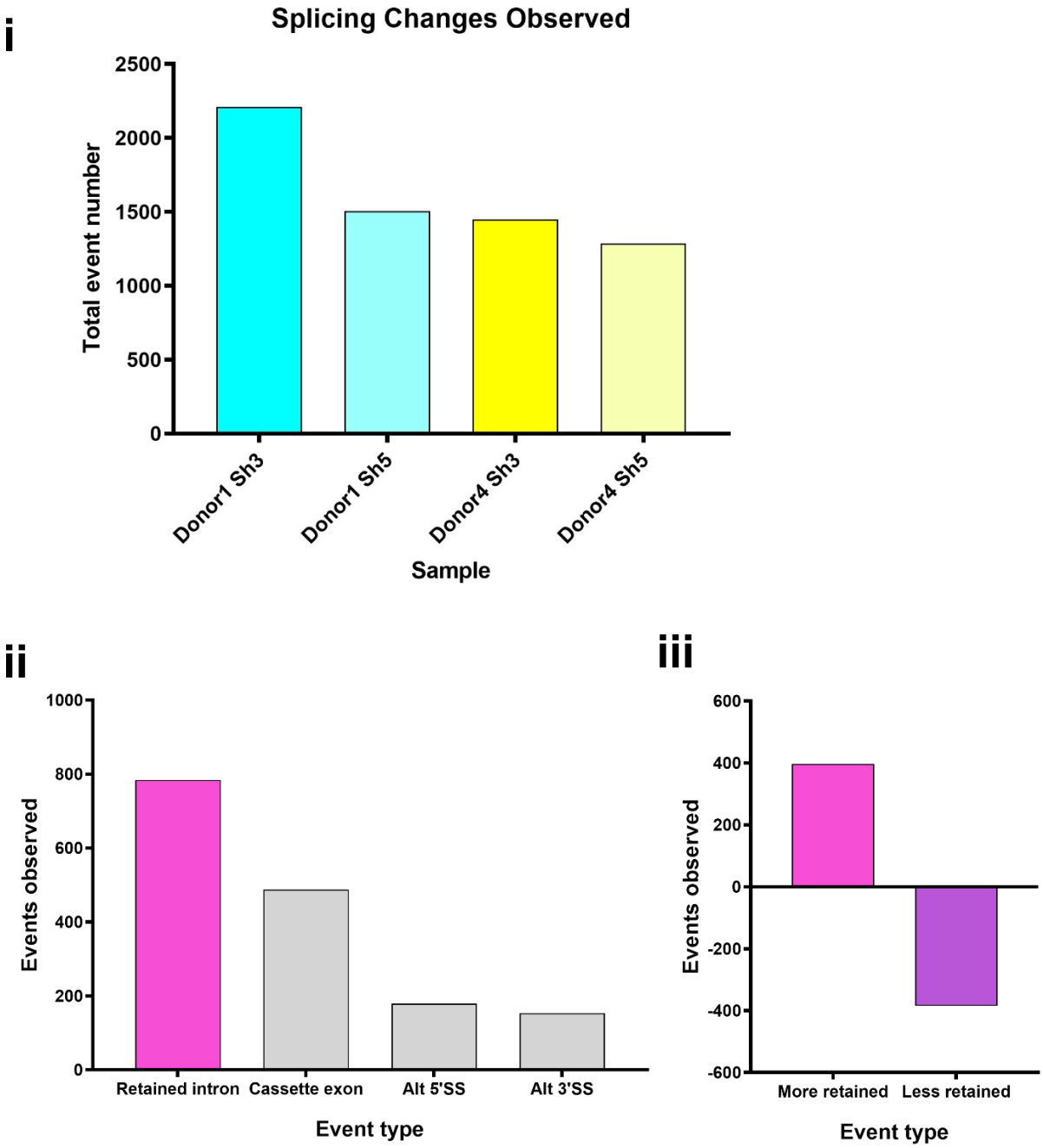


Fig. S3C. Alternative splicing events observed upon DDX39B depletion in CD4⁺ T cells. Splicing analysis of RNAseq data was carried out using Vast-Tools. **i.** Total events changed between control and either of two knockdown conditions (Sh3 & Sh5) in two donors (donor 1 & 4). **ii.** Type of events of Sh3 knockdown in either donor are shown. Alt 5'SS, alternative 5' splice site; Alt 3'SS, alternative 3' splice site. **iii.** Intron retention events in either donor are divided into those showing more retention or less retention with DDX39B knockdown.

Fig. S4. C-rich polypyrimidine tracts determine DDX39B sensitivity.

Related to Figure 6.

Fig. S4A

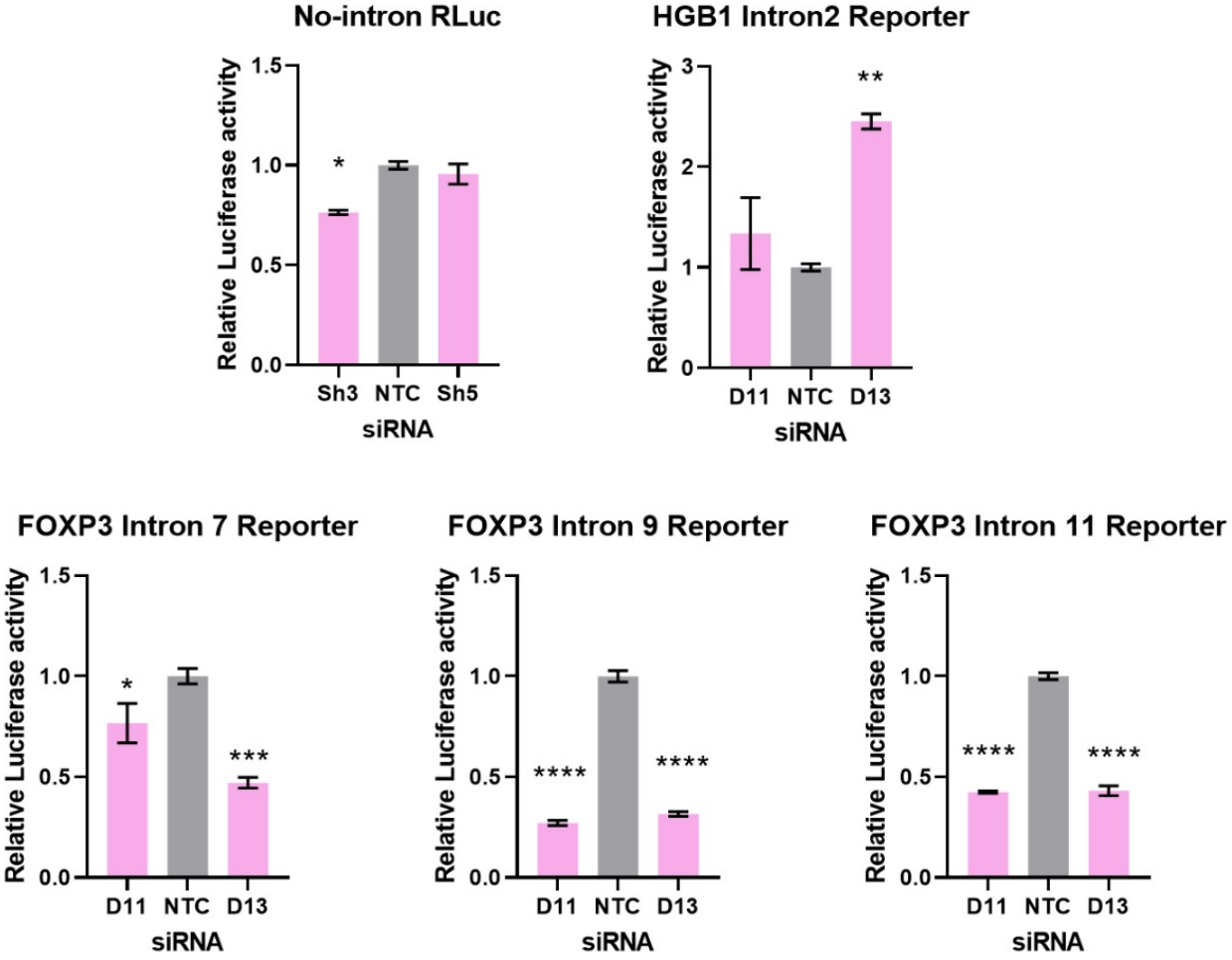


Fig. S4A. Luciferase activity of the splicing reporters. Control (NTC) or DDX39B-depleted (D11 or D13 siRNAs) HeLa cells were co-transfected with a Firefly luciferase reporter (transfection control, FLuc) and Renilla luciferase (RLuc) splicing reporters with no intron, human *ßglobin* (HGB1) intron 2, or *FOXP3* introns 7, 9 or 11. Relative splicing efficiency of reporters was inferred by measuring luciferase activity (RLuc/FLuc) normalized to NTC. *: $p < 0.05$, **: $p < 0.01$, ***: $p<0.001$ and ****: $p < 0.0001$.

Fig. S4B

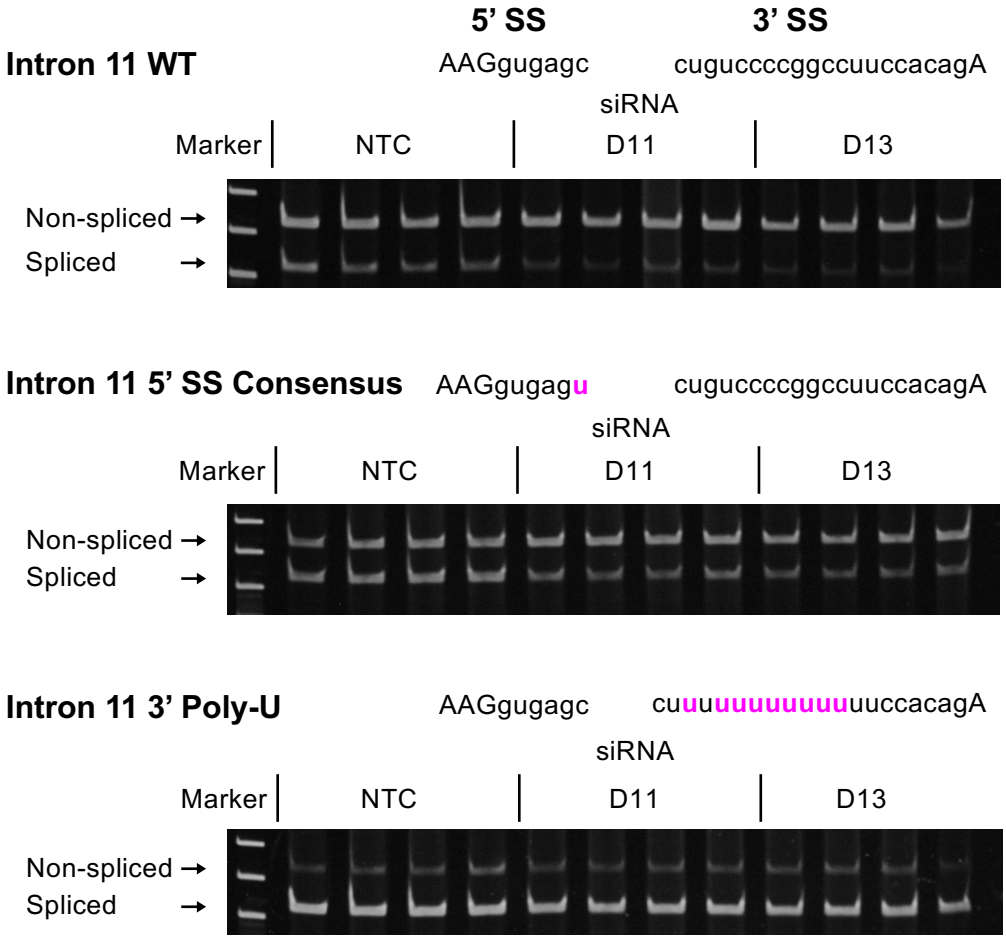


Fig. S4B. RT-PCR analysis of splicing efficiency of wild-type or mutant *FOXP3* Intron 11 RLuc reporters. Control (NTC) or DDX39B-depleted (D11 or D13) HeLa cells were transfected with *FOXP3* intron 11 RLuc reporters (WT, 5' SS consensus or 3' Poly-U mutant), and splicing efficiency was directly measured by endpoint RT-PCR. Quantification of the gels is shown in Fig. 6B.

Fig. S4C

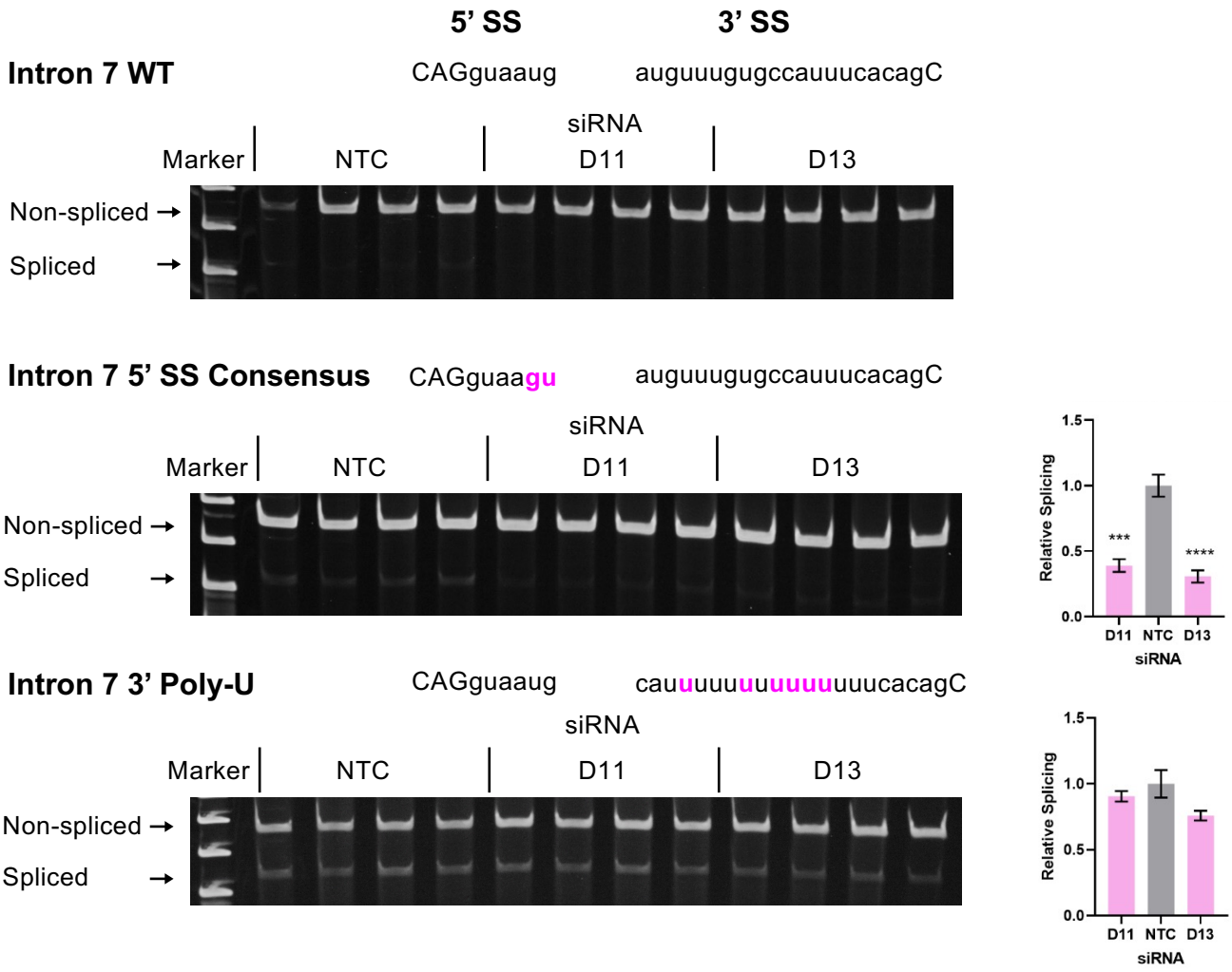


Fig. S4C. RT-PCR analysis of splicing efficiency of wild-type or mutant *FOXP3* Intron 7 RLuc reporters. Control (NTC) or DDX39B-depleted (D11 or D13) HeLa cells were transfected with *FOXP3* intron 7 RLuc reporters (WT, 5' SS consensus or 3' Poly-U mutant), and splicing efficiency was directly measured by endpoint RT-PCR. Quantification is shown on the right for the lower two panels; given the very low level of splicing with intron 7 WT reporter it could not be accurately quantified. ***: $p<0.001$ and ****: $p < 0.0001$.

Fig. S4D

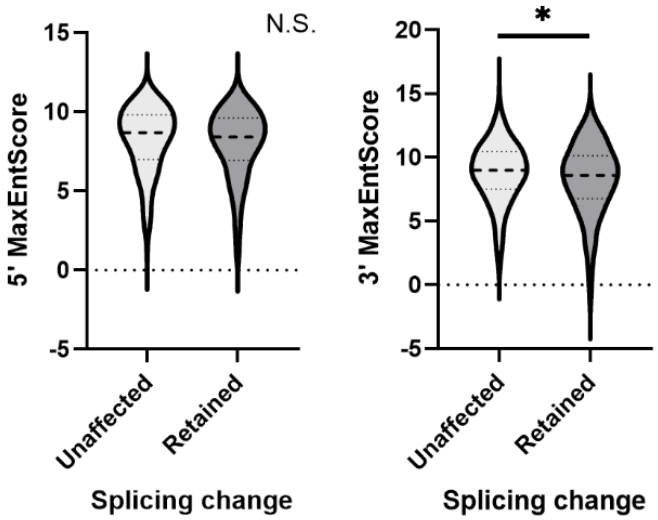


Fig. S4D. Comparison of Max entropy splice sites (SS) score between unaffected and DDX39B-sensitive introns. Max entropy score of 5' SS and 3' SS of unaffected (500 randomly selected) and introns with increased retention upon DDX39B depletion (397 events) was determined using *MaxEntScan* (Yeo G and Burge CB (2004). Dashed lines within the violin plots denote median values. *: $p < 0.05$.

Fig. S4E

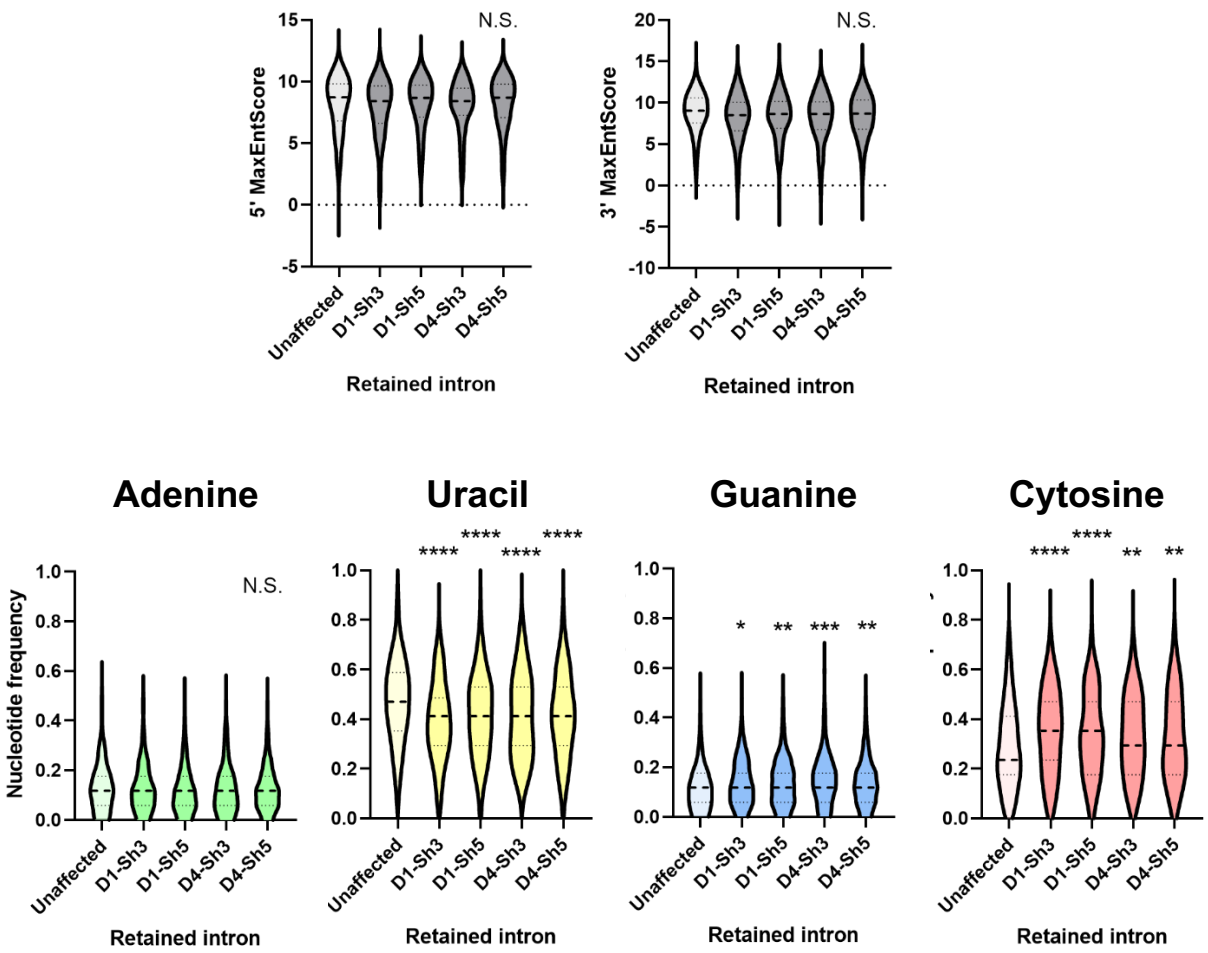


Fig. S4E. Comparison of Max entropy splice site (SS) scores (top panel) and nucleotide composition of py tracts (bottom panel) of unaffected introns or introns that were more retained in CD4⁺ T cells from either donor 1 or 4 depleted of DDX39B by treatment with Sh3 or Sh5. *: $p < 0.05$, **: $p < 0.01$, ***: $p<0.001$ and ****: $p < 0.0001$.

Fig. S4F

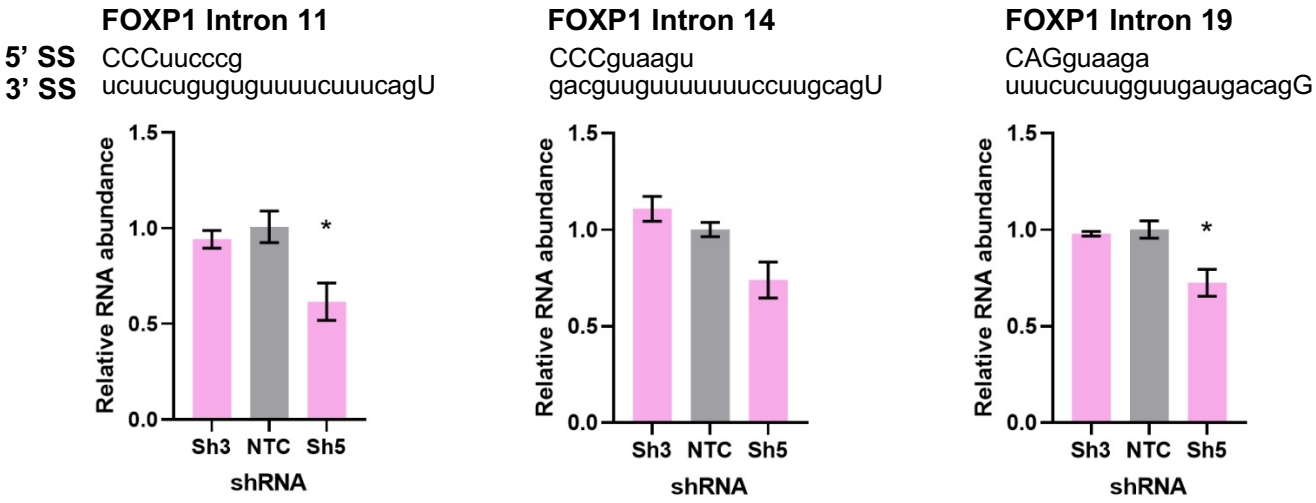


Fig. S4F. Retention of *FOXP1* introns is not increased upon DDX39B depletion in MT-2 cells. FOXP1 transcripts including introns 11, 14, or 19 were quantified by intron-specific RT-qPCR in control (NTC) and DDX39B-depleted (Sh3 or Sh5) MT-2 cells and normalized to total FOXP1 transcripts. *: $p < 0.05$.

Fig. S4G

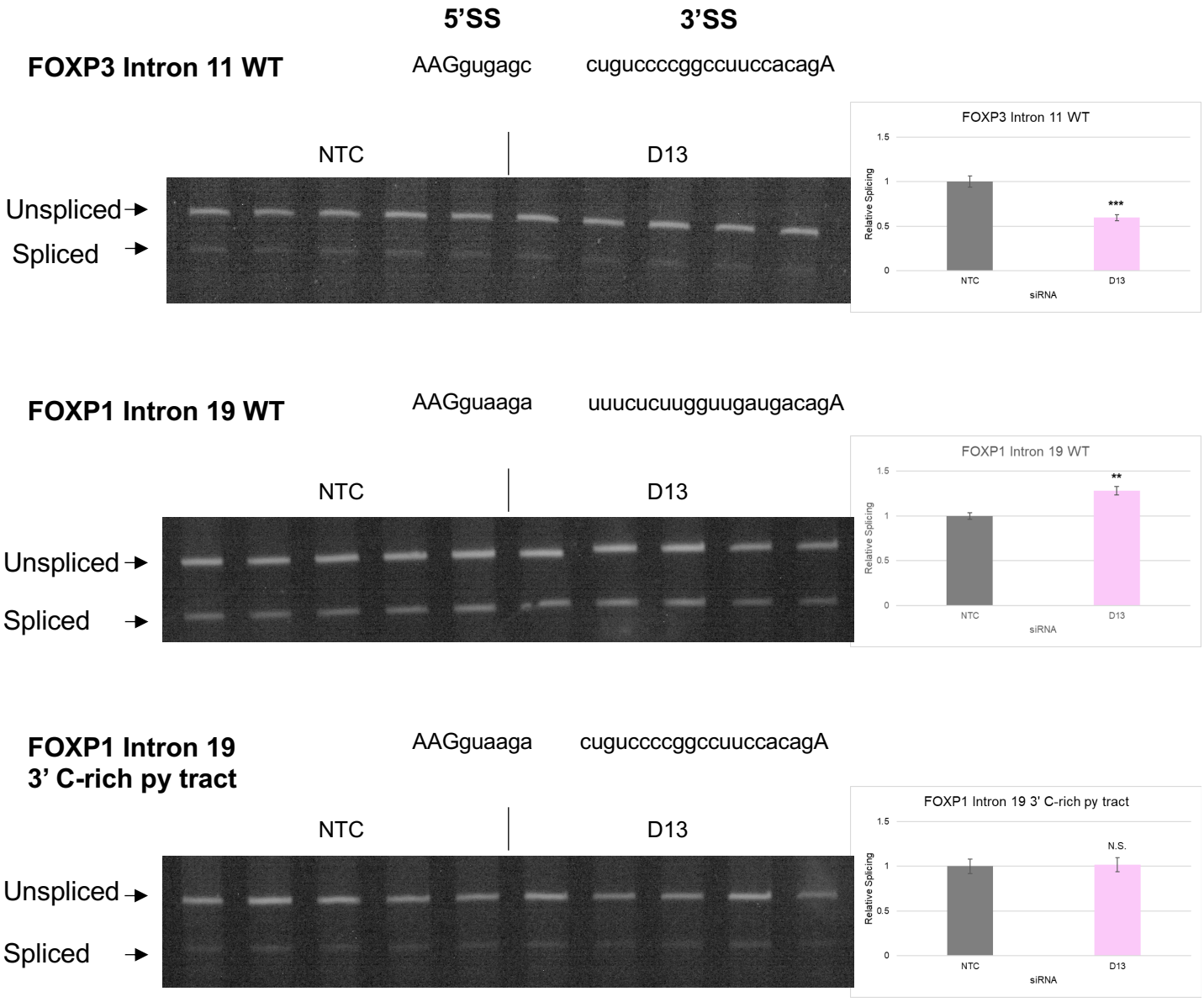


Fig. S4G. RT-PCR analysis of splicing efficiency of wild-type or mutant *FOXP1* Intron 19 RLuc reporters. Control (NTC) or DDX39B-depleted (D13) HeLa cells were transfected with *FOXP3* or *FOXP1* RLuc reporters. Splicing efficiency was directly measured by endpoint RT-PCR. Quantification is shown on the right. **: $p<0.01$ and ***: $p<0.001$.

Fig. S4H

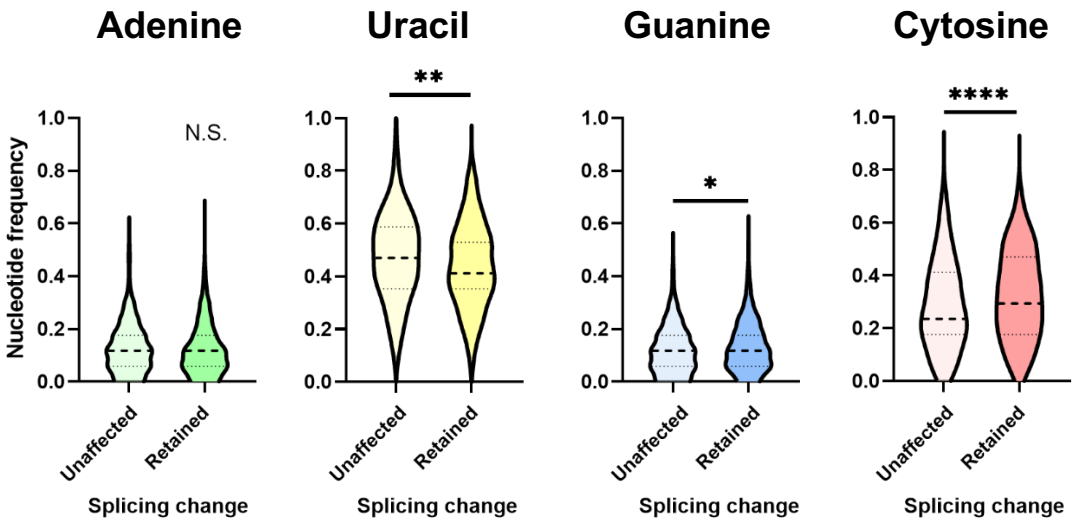
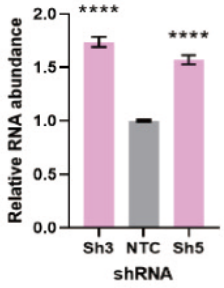


Fig. S4H. Comparison of py tract composition between unaffected and less-retained introns. Nucleotide frequency of py tracts of unaffected (500 randomly selected) and those that are less-retained upon DDX39B knockdown (387) was calculated. Dashed lines denote median values. p < 0.05, **: p < 0.01, ***: p<0.001 and ****: p < 0.0001.

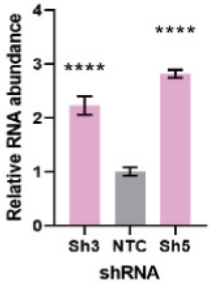
Fig. S4I

C-rich py tract genes

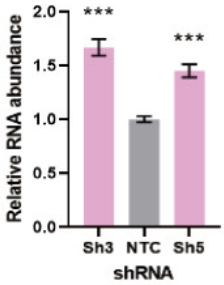
FAM3A Intron 6
5' SS ACGgugagg
3' SS gccccucccuccugacacagG



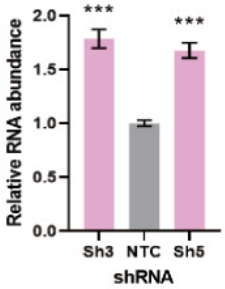
FAM3A Intron 7
GAGgugagu
cguggcaccugccccccagA



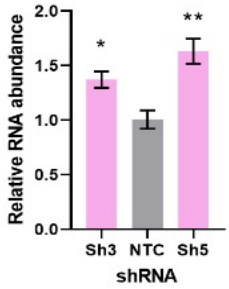
PORCN Intron 7
CAGguaaau
cccugccccuuucuccccagG



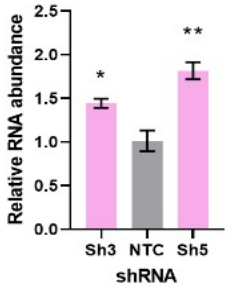
PORCN Intron 12
UUGgugagg
uauccaucugucccccacagG



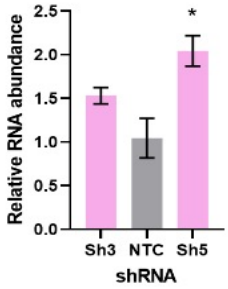
RBM10 Intron 5
5' SS AAGguacag
3' SS gccuggccugugccucacagU



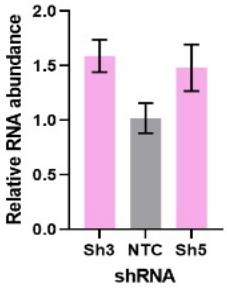
RBM10 Intron 12
CAGguacuc
cccucccaccugcccuucagG



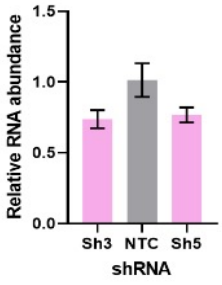
RENBPN Intron 3
CAGguaugg
uucccucccguacccgcagG



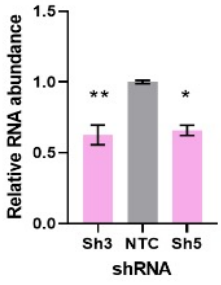
RENBP Intron 6
CAGggggg
cuucugcuccuucccgguagA



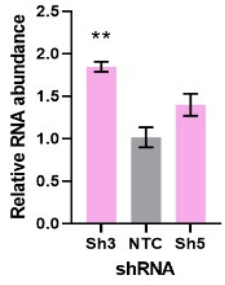
CFP Intron 5
5' SS CCAgugagu
3' SS caccucccaccccaccacagC



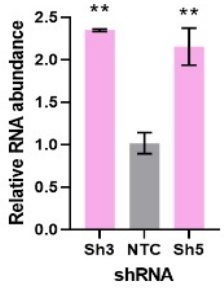
CFP Intron 7
CUGgucagc
cuguugcccugucucugcagU



G6PD Intron 5
GAUguaagg
cagcuccgggcucccagcagA

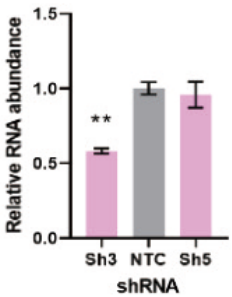


G6PD Intron 7
GAGggggg
ccaccuugccccucccugcagA

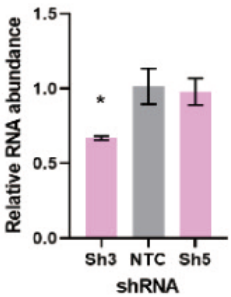


U-rich py tract genes

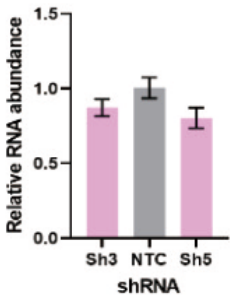
FMR1 Intron 5
5' SS AAUguaagu
3' SS uuuuuuuuuuaaauuucuagG



FMR1 Intron 9
UAG guaagu
uugauuuauuuauuucuuagG



DDX3X Intron 7
CAG guaagc
guuuguuuguuuuugaacagG



DDX3X Intron 15
GAG gugagg
acuuaauuuuuuucuuauagU

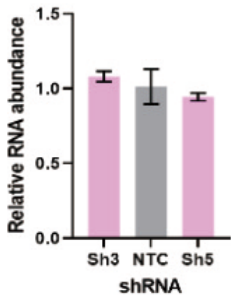


Fig. S4I. Detection of intron retention events in X chromosome genes with C-rich or U-rich py tracts. Introns of the transcripts with C-rich **(top)** or U-rich **(bottom)** py tract were quantified by intron-specific RT-qPCR in control (NTC) and DDX39B-depleted (Sh3 or Sh5) MT-2 cells and normalized to their corresponding total transcripts. *: $p < 0.05$, **: $p < 0.01$, ***: $p<0.001$ and ****: $p < 0.0001$.

Fig. S5

(relates to Figure 7)

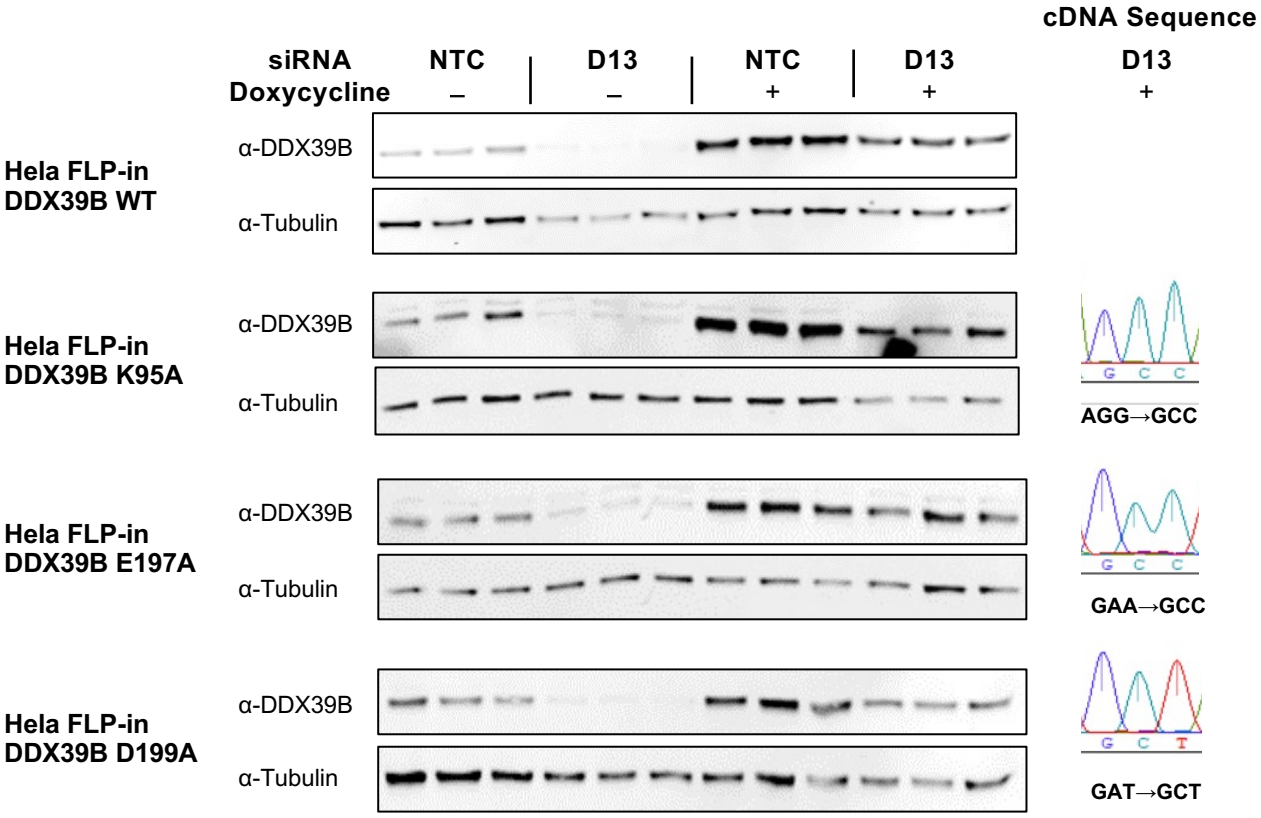


Fig. S5. Expression of WT or mutant DDX39B in stable cell lines. Stable HeLa cell lines with inducible expression of WT or mutant DDX39B *trans*-genes were transfected with control (NTC) or DDX39B (D13) siRNAs, and expression of a siRNA-resistant DDX39B transgene was induced with Doxycycline. DDX39B protein expression was quantified by western blot (left). Sanger sequencing of DDX39B RT-PCR amplicons from rescue with mutant DDX39B *trans*-genes (D13 + Doxycycline) shows preferential expression of DDX39B *trans*-genes under rescue conditions (right).