RNA-binding motifs of hnRNP K are critical for induction of antibody diversification by activation-induced cytidine deaminase

Ziwei Yin[a,1], Maki Kobayashi[a,1], Wenjun Hu[a], Koichi Higashi[b], Nasim A. Begum[a], Ken Kurokawa[b], and Tasuku Honjo[a,2]

[a]Department of Immunology and Genomic Medicine, Graduate School of Medicine, Kyoto University, 606-8501 Kyoto, Japan; and [b]Center for Information Biology, National Institute of Genetics, Mishima, 411-8540 Shizuoka, Japan

Contributed by Tasuku Honjo, March 14, 2020 (sent for review December 2, 2019; reviewed by Yukio Kawahara and Sean M. Post)

Activation-induced cytidine deaminase (AID) is the key enzyme for class switch recombination (CSR) and somatic hypermutation (SHM) to generate antibody memory. Previously, heterogeneous nuclear ribonucleoprotein K (hnRNP K) was shown to be required for AID-dependent DNA breaks. Here, we defined the function of major RNA-binding motifs of hnRNP K, GXXGs and RGGs in the K-homology (KH) and the K-protein-interaction (KI) domains, respectively. Mutation of GXXG, RGG, or both impaired CSR, SHM, and *cMyc/IgH* translocation equally, showing that these motifs were necessary for AID-dependent DNA breaks. AID–hnRNP K interaction is dependent on RNA; hence, mutation of these RNA-binding motifs abolished the interaction with AID, as expected. Some of the polypyrimidine sequence-carrying prototypical hnRNP K-binding RNAs, which participate in DNA breaks or repair bound to hnRNP K in a GXXG and RGG motif-dependent manner. Mutation of the GXXG and RGG motifs decreased nuclear retention of hnRNP K. Together with the previous finding that nuclear localization of AID is necessary for its function, lower nuclear retention of these mutants may worsen their functional deficiency, which is also caused by their decreased RNA-binding capacity. In summary, hnRNP K contributed to AID-dependent DNA breaks with all of its major RNA-binding motifs.

activation-induced cytidine deaminase | heterogeneous nuclear ribonucleoprotein K | RNA-binding motifs | DNA breaks | IgH

Activation-induced cytidine deaminase (AID) is specifically expressed in activated B lymphocytes and is responsible for class switch recombination (CSR) and somatic hypermutation (SHM) in the adaptive immune system (1). AID is a 198-amino-acid protein consisting of an N-terminal domain necessary for the induction of single strand breaks (SSBs) of DNA, a cytidine-deaminase catalytic domain in the central region and a C-terminal domain required for the DNA repair steps of CSR (1–3). After AID activation, DNA breaks occur in both switch (S) and variable (V) regions of immunoglobulin heavy chain (IgH) genes followed by the different repair steps for SHM and CSR. The error-prone polymerases repair the DNA break sites in V regions for SHM (4), and for CSR the nonhomologous end-joining repair pathway mainly works in two distant S regions. CSR consists of a more complex combination of several steps, including the processing of SSBs into double strand breaks (DSBs) by several DNA end-processing enzymes, including APE1 and the MRN complex (5), followed by AID-dependent DNA synapsis formation and recombination to complete CSR (6).

However, there has been a long-standing debate regarding the molecular mechanism of AID in SSBs in the V and S regions and repair in the S regions (6). Because AID is the cytidine (C)-to-uracil (U) converting enzyme, the question of which is the target of AID—C in RNA or C in DNA—has not been resolved yet. "DNA deamination by AID" hypothesis proposes that base excision repair or mismatch repair mechanism produces DNA breaks (7). However, various mutants of AID showed that level of in vitro DNA deamination does not always correlate with the frequencies of SHM and CSR in vivo, questioning the plausibility of DNA deamination by AID (8). Alternatively, the "RNA editing" hypothesis proposes that AID edits some putative RNAs for DNA breaks and the other RNAs for DNA repair with the help of the several cofactors (6). Our previous studies showed that heterogeneous nuclear ribonucleoprotein (hnRNP) K is necessary for both SHM and CSR, while hnRNP L, U, and SERBP1 are specifically required for CSR (9, 10). This is further supported by the evidence that AID distributes in two different complexes in "light" and "heavy" fractions separated by ultracentrifuge (10). The "light" fraction contains hnRNP K and wild-type (WT) or C-terminally mutated AID which can induce DNA breaks. In contrast, the "heavy" fraction includes hnRNP L, U, and SERBP1 functioning in DNA repair and wild-type AID which can support DNA repair. Furthermore, C-terminus mutants of AID do not dimerize and only localize to the "light" fraction while wild-type AID dimerizes and localizes to both "light" and "heavy" fractions, indicating that the two different AID–cofactor complexes support the two distinct AID's functions. Actually AID has been proved to edit RNA when it is

Significance

Heterogeneous nuclear ribonucleoprotein K (hnRNP K), an RNA-binding protein, is the cofactor of activation-induced cytidine deaminase (AID) that induces DNA breaks in immunoglobulin (Ig) genes. Here, we elucidated that the GXXG and RGG RNA-binding motifs were critically necessary for class switch recombination and somatic hypermutation. Nuclear localization of hnRNP K and interaction with AID were also dependent on all of the RNA-binding motifs. This study clearly demonstrated that hnRNP K contributed to AID-dependent DNA breaks through its RNA-binding capacity, suggesting the possibility that hnRNP K holds and presents some editing target RNAs to AID, which eventually induces DNA breaks in Ig genes.

Author contributions: Z.Y., M.K., and T.H. designed research; Z.Y. and M.K. performed research; W.H., K.H., N.A.B., and K.K. contributed new reagents/analytic tools; Z.Y., M.K., K.H., K.K., and T.H. analyzed data; and Z.Y., M.K., and T.H. wrote the paper.

Reviewers: Y.K., Osaka University; and S.M.P., MD Anderson Cancer Center.

The authors declare no competing interest.

Data deposition: The row data of somatic hypermutation analysis by next-generation sequencing was uploaded to https://www.ncbi.nlm.nih.gov/bioproject/612426. Python-formatted files used in informatics analysis have been deposited at https://github.com/makikbys/somatichypermutation.

[1]Z.Y. and M.K. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: honjo@mfour.med.kyoto-u.ac.jp.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1921115117/-/DCSupplemental.

First published May 8, 2020.

encapsulated in the hepatitis B virus (11). Additionally, we reported the other mechanism of DNA breaks, in which topoisomerase I (Top1) decrease by AID alters DNA helical structure to the non-B form in both the repeat-rich S and V regions and promotes DNA cleavage by Top1 (12, 13). Therefore, the function of AID will be more than DNA deamination.

hnRNP K is a member of the poly (C)-binding protein (PCBP) family (14), which was originally purified from a hnRNP complex. hnRNP K has been shown to be necessary for precursor messenger RNA (pre-mRNA) metabolism (15–17). Additionally, hnRNP K functions as the scaffold to organize the interaction between nucleic acids and other protein partners, regulating gene expression in different phases. For example, through binding to the 3′ untranslated region (3′ UTR) with other RNA-binding proteins, hnRNP K participates in stabilizing mRNA (18) or controlling mRNA translation (19). It also associates with noncoding RNAs (ncRNAs) involved in gene translation (20) or transcription (21). These revealed molecular functions give the basis for that hnRNP K has several functions in cell proliferation (22), DNA repair (23, 24), neuronal cell development (25), and cancer progression (26). A recent study revealed that diffuse large B-cell lymphoma patients with high levels of hnRNP K expression show a poor outcome due to the activation of the oncogene cMyc by hnRNP K (27), whereas haploinsufficiency of hnRNP K also results in myeloid malignancy caused by a decrease of C/EBPα (28), showing that both lower and higher hnRNP K expression results in the deregulation of the cell cycle by regulating different targets.

Because hnRNP K is specifically necessary for AID-dependent DNA breaks, to answer the question of which domain of hnRNP K is responsible for the association with AID and DNA breaks we focused on the molecular dissection of hnRNP K using murine lymphoma-derived CH12F3-2A (CH12) cells, which enable monitoring of CSR from IgM to IgA by cytokine mixture (cluster of differentiation [CD]40L, interleukin 4 [IL-4], and transforming growth factor type β [TGF-β] [CIT]), as well as AID-dependent DNA breaks and other IgH gene recombination events. Two possible types of RNA-binding domains are found in hnRNP K (29). The first type is the three K-homology (KH) domain, which is highly conserved in other PCBPs (30), and the other is the K-protein-interaction (KI) domain harboring RGG motifs. Every KH domain contains a GXXG motif, which favorably binds to poly(C) sequences in both RNA and single-stranded DNA (ssDNA) (31, 32). However, the KI domain is very unique to hnRNP K (29, 30) and the RNA-binding capacity of this KI domain is not sufficiently defined, although it encodes multiple RGG motifs, which has the potential to bind both proteins and RNAs (33).

This study showed that both the GXXG and RGG motifs played an important role in CSR and SHM, because they were necessary for AID-dependent DNA breaks. Moreover, CSR- and SHM-deficient hnRNP K mutants almost lost the RNA-dependent interaction with AID, as well as the ability to bind with the typical hnRNP K-binding RNAs. It suggested that specific RNA(s), binding of which was abolished by the mutation of GXXG or RGG motifs, might be responsible for AID-dependent DNA breaks. Additionally, both GXXG and RGG motifs were required for nuclear localization of hnRNP K. Because it was previously shown that nuclear localization signal (NLS) mutants of AID are defective in CSR and SHM (3, 34), lower nuclear localization in these RNA-binding motif mutants could partially contribute to their malfunction in AID-dependent DNA breaks.

Results

All Three KH Domains of hnRNP K Contributed to Efficient AID-Dependent CSR. To evaluate the relative importance of the three KH domains in hnRNP K, we evaluated the mutants of one KH domain deletion ($\Delta KH1^R \sim \Delta KH3^R$) and deletion of two KH domains ($KH1^R \sim KH3^R$) by "CSR rescue" assay (*SI Appendix*, Fig. S1 *A* and *B*). These mutants were tagged by c-Myc-FLAG at the C terminus and had silent mutations enabling resistance to small interfering RNA (siRNA)-mediated knockdown (indicated by superscript "R"). To compare their CSR rescue ability with that of the wild-type hnRNP K (WT K^R), we used a hnRNP K-mutated cell line, clone K2-20, generated from CH12 cells (9). Because this K2-20 cell line had residual endogenous hnRNP K, IgA% of siControl (~24%, average) was further decreased to ~4% by sihnRNP K (sihnRNP K without any plasmid, "no plasmid") (*SI Appendix*, Fig. S1 *B–E*) and was restored by WT K^R at a similar level to the siControl with CIT (CD40L, IL-4, and TGF-β) ($P = 0.943$). WT K^R was used as the positive control for comparing CSR rescue efficiency with that of the mutants, because WT K^R and the mutants were prepared by the same procedure combining sihnRNP K and rescue by WT or the mutant of hnRNP K. All of the KH domain mutants rescued CSR but almost at half the level of WT K^R (IgA% = 10 ~15%, $P < 0.05$). Because all of the KH domain mutants equally rescued CSR, every KH domain of hnRNP K additively contributed to AID-dependent CSR (*SI Appendix*, Fig. S1 *D* and *E*). "GFP" rescued by the cMyc-FLAG-tagged green fluorescent protein (GFP) showed a level similar to IgA% with no plasmid.

All Three GXXG Motifs in the KH Domains of hnRNP K Were Necessary for CSR. To identify the essential RNA-binding motifs of hnRNP K, which were specifically required for AID-dependent CSR, the reported RNA-binding motifs in the KH domain of hnRNP K were mutated. Instead of a large domain deletion that may possibly alter the protein structure, the mutation of a few amino acids that did not affect the whole protein structure was used for functional molecular dissection. A few amino acids in the KH3 domain were mutated based on the precise structural data by NMR and X-ray analysis (32, 35) and an in vitro study of RNA-binding ability (36). These reports revealed the necessity of the GXXG motif and the adjacent isoleucine ^{403}I for ssDNA binding capacity. Especially, the GDDG mutation from GXXG did not change the structure and stability of the KH domain. The CSR rescue ability of these three KH3 mutants, GXXG to GDDG ($KH3.1^R$), ^{403}I to ^{403}D ($KH3.2^R$), and DGDDG ($KH3.3^R$), were compared to that of $KH3^R$ (*SI Appendix*, Fig. S2 *A* and *B*). As the result, all three mutants of a few amino acids tended to decrease IgA%, although the difference between $KH3^R$ and these mutants was very small and statistical significance was found only between $KH3^R$ and $KH3.2^R$ (*SI Appendix*, Fig. S2 *C–E*).

Because GDDG mutation keeps the structure of the KH domain (36), we introduced the GDDG mutation into every GXXG motif in each KH domain to construct the GXXG mutants, $Mu1^R \sim Mu5^R$ and $3GDDG^R$, to determine their CSR rescue efficiency (Fig. 1*A* and *B*). Expression of all these GXXG motif mutants were not lower than that of WT K^R (Fig. 1*C*), because hnRNP K versus GAPDH ratio measured by ImageJ is 2.11, 1.75, 2.02, and 1.46 in $Mu4^R$, $Mu5^R$, $3GDDG^R$, and WT K^R, respectively. CSR rescue by these GXXG mutants became less efficient when more GXXG motifs were mutated in hnRNP K (Fig. 1*C–F*). Actually, CSR efficiency of the single GDDG mutant $Mu1^R \sim Mu3^R$ was significantly lower than WT K^R except for $Mu3^R$. The double GXXG mutant $Mu4^R$ and $Mu5^R$ showed much lower CSR rescue ability (IgA switch = 6.5 and 5.5% [average]) than $Mu1^R \sim Mu3^R$ (Fig. 1*F*, $P < 0.05$). Moreover, the triple GXXG mutant $3GDDG^R$ restored CSR up to ~5%, which was calculated to be 12% of the WT K^R (Fig. 1*E* and Table 1), although the statistical significance between $3GDDG^R$ and the double mutants $Mu4^R$ and $Mu5^R$ was not found. These results suggested that the three GXXG motifs of hnRNP K acted additively to support AID-dependent CSR.

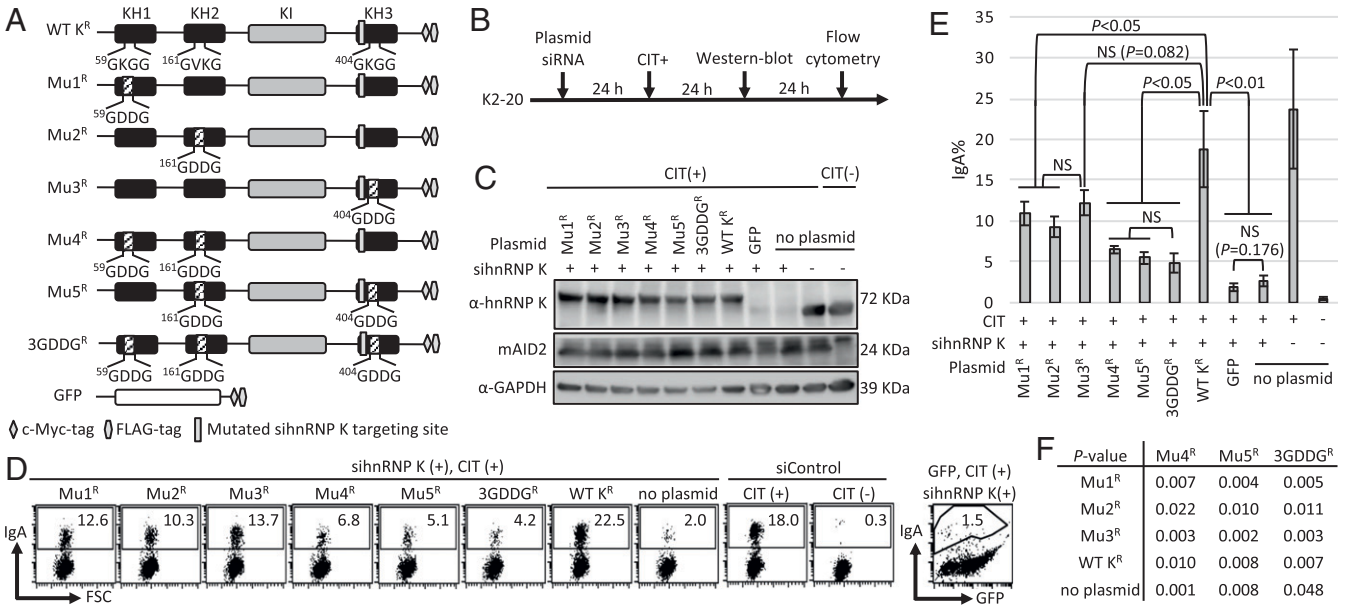

Fig. 1. CSR efficiency by the hnRNP K GXXG motif mutants. (*A*) The constructs of sihnRNP K-resistant hnRNP K tagged with c-Myc-FLAG, with or without the GXXG motif mutation in the KH domains. The GXXG motifs were mutated to GDDG as illustrated. (*B*) The time course for the CSR rescue assay. hnRNP K-depleted K2-20 cells were transfected with sihnRNP K and rescued by the constructs shown in *A*. (*C*) Protein expression of each mutant, as confirmed by Western blot analysis. (*D*) A representative profile of the CSR restoration by the WT K^R and the GXXG mutants. FSC, forward scattered. (*E*) Averaged IgA% of the rescued samples calculated from three independent experiments. Data are the mean ± SD. Statistical significance was calculated by Student's *t* test. NS, not significant. The "no plasmid" sample was the negative control in these experiments. (*F*) *P* values between GXXG mutants.

Both the GXXG and RGG Motifs of hnRNP K Were Required for CSR. Because the $3GDDG^R$ still had very low residual CSR rescue ability when compared to the negative controls (no plasmid, *P* value is 0.048 in comparison between $3GDDG^R$ and no plasmid), the other potential RNA-binding motifs, RGGs in the KI domain were tested for CSR rescue function. To test whether the enrollment of RGG motifs overlapped with the GXXG motifs or not, the RGG motif-only mutants and the mutants of both the RGG and GXXG motifs were constructed and compared (Fig. 2*A*). The structure of the KI domain was not reported previously; therefore, ADD mutations in the RGG motif were designed based on the study of human fragile mental retardation protein (FMRP) (37).

tFig. 2*A* shows the RGG motif mutants derived from the construct of WT K^R, $Mu6^R$ ~ $5ADD^R$. The same ADD mutations were introduced into the $3GDDG^R$ ($Mu9^R$ ~ 3GDDG + $5ADD^R$). The time course of the experiment was the same as the previous experiment (Fig. 1*B*). All of the RGG motif mutants expressed equally to WT K^R and $3GDDG^R$ (Fig. 2*B*). Among them, $Mu6^R$ and $Mu9^R$ retained the same size as WT K^R while others showed a higher speed of electrophoretic mobility. The averaged IgA% of the mutants $Mu6^R$ ~ $Mu8^R$ decreased less than half of the WT K^R ($P < 0.05$) and $5ADD^R$ showed much less average (~5%) of IgA%, although *P* values between the intermediate mutants ($Mu6^R$ ~ $Mu8^R$) and $5ADD^R$ were not significant (Fig. 2*C*–*E*). IgA% of $3GDDG^R$ and 3GDDG + $5ADD^R$ were not different, meaning that the RGG motifs did not support remaining CSR rescue activity of $3GDDG^R$. Normalized CSR rescue (percent) of $3GDDG^R$, $5ADD^R$, and 3GDDG + $5ADD^R$ were 12%, 18%, and 16%, respectively (Table 1). This small residual CSR rescue activity in these mutants might be contributed by the potential unknown RNA-binding motifs in hnRNP K or the alternative RNA-binding proteins, although their compensation is incomplete.

RGG and GXXG Motifs of hnRNP K Supported AID-Dependent SHM. To test whether SHM, the other AID-dependent IgH diversification, was also reduced by mutation of hnRNP K, we analyzed SHM of the Sμ region in K2-20 cells using the rescue assay (Fig. 3*A*). Fragments in the 5′-Sμ region were sequenced by the Sanger method (676 bp; *SI Appendix*, Fig. S3*A*) and next-generation sequencing (NGS) (447 bp; Fig. 3*B*).

Table 1. CSR rescue ability of hnRNP K mutants relative to WT K^R

	Construct	CSR rescue, %
Type of mutant	WT K^R	100
KH-deleted mutant	$\Delta KH1^R$	45 ± 5.9
	$\Delta KH2^R$	37 ± 5.8
	$\Delta KH3^R$	48 ± 10.5
	$KH1^R$	43 ± 4.4
	$KH2^R$	57 ± 5.8
	$KH3^R$	32 ± 7.0
KH3 mutant	$KH3.1^R$	19 ± 7.8
	$KH3.2^R$	18 ± 1.7
	$KH3.3^R$	25 ± 12.4
GXXG mutant	$Mu1^R$	53 ± 13.0
	$Mu2^R$	43 ± 17.4
	$Mu3^R$	62 ± 17.4
	$Mu4^R$	25 ± 5.7
	$Mu5^R$	18 ± 3.2
	$3GDDG^R$	12 ± 1.0
RGG mutant	$Mu6^R$	33 ± 4.3
	$Mu7^R$	41 ± 15.0
	$Mu8^R$	25 ± 5.8
	$5ADD^R$	18 ± 1.4
GXXG + RGG mutant	$Mu9^R$	25 ± 5.4
	$Mu10^R$	21 ± 2.6
	$Mu11^R$	23 ± 3.2
	$3GDDG+5ADD^R$	16 ± 1.2
NLS mutant	$Mu12^R$	90 ± 14.1

The percent values are relative IgA% calculated by (the mutant – no plasmid)/(WT K^R – no plasmid). All CSR rescue assays were performed three independent times and the results are presented as mean ± SD.

By the Sanger method, SHM of WT K^R and 3GDDG^R transfectants were analyzed (*SI Appendix*, Fig. S3 *A–C*). By comparing the CIT (+) samples of WT K^R and 3GDDG^R, it seems that 3GDDG^R induces less SHM than WT K^R. However, the *P* values are not statistically significant and it was difficult to conclude the SHM level of 3GDDG^R, suggesting the necessity of NGS for SHM. K2-20 cells were originally monoclonal; however, they were diversified into a heterogeneous population with the accumulation of many SNPs that occurred during long-term culture because at least five different typical sequences were revealed in the nonstimulated cells, as shown by Sanger sequencing (Variants 1 through 5 in *SI Appendix*, Fig. S3*D*).

In SHM analysis by NGS, protein expression of AID and the transfected hnRNP K molecules were not affected and levels of IgA switching were similar to the other experiments (Fig. 3*C* and Table 1). Variant 1 (*SI Appendix*, Fig. S3*D*) was arbitrarily selected as the reference in the NGS analysis. NGS provided about 370,000 to 510,000 reads per sample (*SI Appendix*, Table S1). Mutation frequency of each nucleotide position was calculated as shown in *SI Appendix*, Table S2. The comparison between CIT (+) and (–) or WT K^R and the other mutants revealed statistically significant differences (Fig. 3*D*, *Top*). To evaluate AID-dependent SHM, the difference of mutation frequency between CIT (+) and (–) samples (ΔMut. Freq.) was calculated (Fig. 3*D*, *Middle*). The 3GDDG^R, 5ADD^R, and 3GDDG + 5ADD^R samples showed a reduction of the total mutation frequency, as well as the ΔMut. Freq. when compared with the WT K^R sample. The ΔMut. Freq. ratio of these mutants to WT K^R, normalized by the value of the WT K^R sample (= 100%) and that of the no plasmid sample (= 0%), showed the SHM rescue abilities of these mutants (Fig. 3*D*, *Bottom*). The 3GDDG^R, 5ADD^R, and 3GDDG + 5ADD^R were 30%, 31%, and 24% of the WT K^R, respectively (see Fig. 8). Equally deficient SHM rescue efficiency in these mutants was consistent with the results of the CSR experiments. The mutational pattern of all the samples does not seem to be different, as shown in *SI Appendix*, Table S3.

GXXG and RGG Motifs of hnRNP K Were Required for AID-Induced DNA Breaks. As a hallmark of lymphomagenesis (38), aberrant chromosomal translocation between the oncogene cMyc and the IgH locus is due to the off-target DNA breaks induced by AID (39). To test the function of the GXXG and RGG motifs of hnRNP K in AID-dependent DNA breaks, we compared the *cMyc*/*IgH* translocation levels (Fig. 4*A*) of K2-20 cells with sihnRNP K rescued by WT K^R, 3GDDG^R, 5ADD^R, and 3GDDG + 5ADD^R (Fig. 4*B*). Here, we used the same rescue assay shown in Fig. 1*B* in combination with knockdown of Top1 to improve the translocation efficiency. In conditions without siTop1, the *cMyc*/*IgH* translocation efficiency of WT K^R-rescued samples under CIT was too low to evaluate the difference statistically between them (*SI Appendix*, Fig. S4*A*). The knockdown of Top1 was confirmed by detecting the protein amount and IgA switch enhancement (*SI Appendix*, Fig. S4 *B* and *C*). DNA purified from the cells after 48 h of CIT stimulation was used for this assay (Fig. 4*B*). Consistent with the results of CSR restoration,

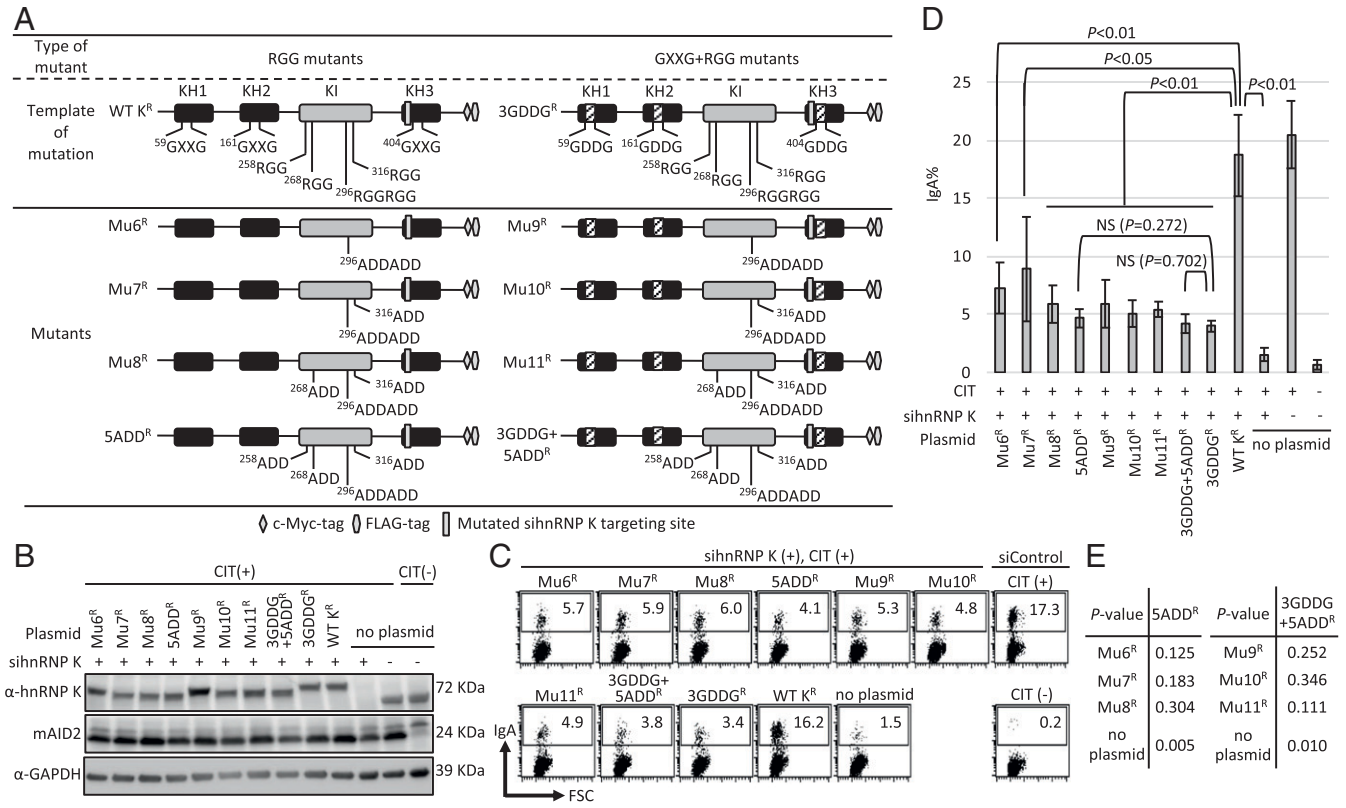

Fig. 2. Evaluation of CSR rescue by RGG mutants and RGG + GXXG mutants. (*A*) The position of the RGG motifs in the KI domain of hnRNP K are marked in the template construct, WT K^R. Mu6^R-Mu8^R and 5ADD^R mutants harbor intact GXXG motif (*Left*), Mu9^R-Mu11^R and 3GDDG + 5ADD^R carry the GDDG mutation in all GXXG motifs (*Right*). (*B*) Protein expression of each mutant, as confirmed by Western blot analysis. (*C*) A representative profile of the CSR restoration by the WT K^R and its mutants. FSC, forward scattered. (*D*) Averaged IgA% of the rescued samples calculated from three independent experiments. Data are the mean ± SD. Statistical significance was calculated by Student's *t* test. NS, not significant. The "no plasmid" sample was the negative control in these experiments. Time course for the CSR rescue assay in B–D is the same as shown in Fig. 1*B*. (*E*) *P* values between RGG mutants (*Left*) or GXXG+RGG mutants (*Right*).

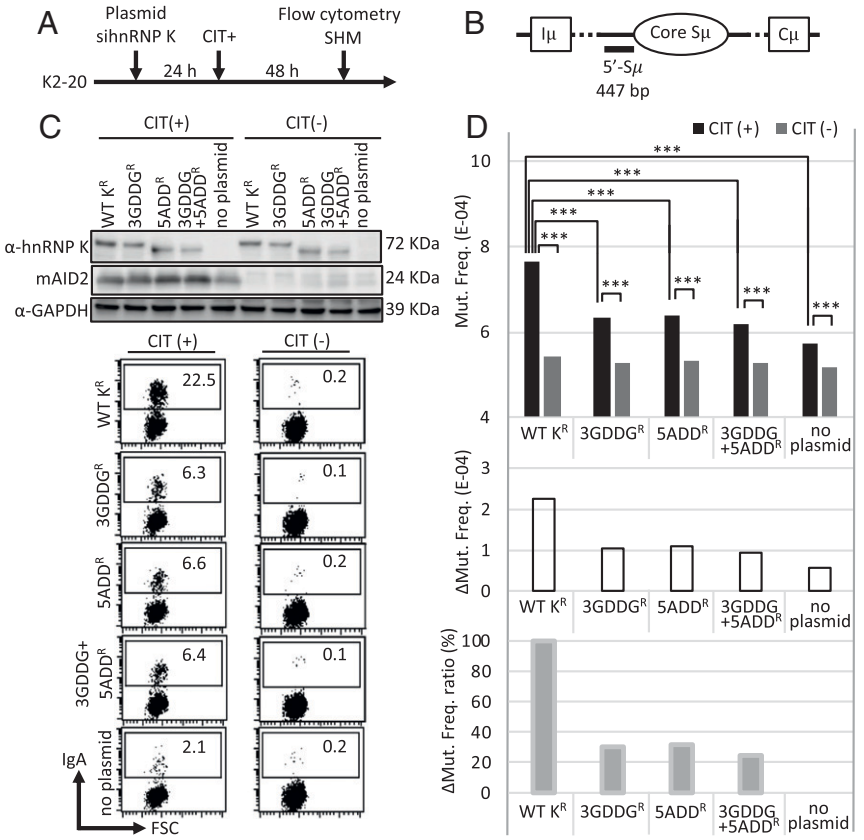

Fig. 3. AID-dependent SHM analysis for the 5′-Sμ region in K2-20 cells rescued by the hnRNP K GXXG or RGG mutants using NGS. (*A*) Experimental design for the SHM analysis. Cells transfected only with sihnRNP K (no plasmid) were used as the negative control. (*B*) The position of the 447-bp regions in the 5′-Sμ analyzed by NGS. (*C*) Confirmation of protein expression (*Top*) and IgA switching (*Bottom*) for the samples used in SHM analysis by NGS. FSC, forward scattered. (*D*, *Top*) The mutation frequency detected by NGS. Statistical significance was calculated by χ^2 test, ***$P < 0.001$. (*D*, *Middle*) The difference in mutation frequency (ΔMut.Freq.). (*D*, *Bottom*) Calculated ΔMut.Freq. ratio using the results of the WT K^R and no plasmid samples.

3GDDG^R, 5ADD^R, or 3GDDG + 5ADD^{R}transfectants showed significantly less translocation frequency of *cMyc*/*IgH* than that of the WT K^R (Fig. 4*C* and *D*), further demonstrating that both the RGG and GXXG motifs of hnRNP K were required for AID-dependent DNA breaks.

Because hnRNP K was considered to be an AID cofactor to induce SSBs of DNA, the RNA-binding motif mutants of hnRNP K were supposed to be deficient in SSBs and the processed form, DSBs. To examine SSB and DSB levels by these motif mutants, we performed linker ligation-mediated PCR (LM-PCR) amplifying DNA break ends formed in the 5′-Sμ region, which is one of the targets of AID (Fig. 4*E*). The DSB signals were detected by 5′-Sμ probes (Fig. 4*F*). Overexpressed hnRNP K molecules were expressed equally and CSR efficiency matched the average data shown in Table 1 (*SI Appendix*, Fig. S4*D*). In contrast to the WT K^R sample with abundant DSB signals, 3GDDG^R, 5ADD^R, and 3GDDG + 5ADD^{R}transfectants showed extremely low DSB levels, which were almost the same as the negative control. With this finding, we hypothesized that the CSR, SHM, and *cMyc*/*IgH* translocation in the 3GDDG^R, 5ADD^R, and 3GDDG + 5ADD^R mutants may be blocked at the DNA break step.

GXXG and RGG Motif Mutations Abolished the AID–hnRNP K Interaction. Interaction of AID with hnRNP K was tested by coimmunoprecipitation (co-IP) using stimulated K2-20 cells (Fig. 5*A*) because the AID–hnRNP K interaction was supposed to be essential for AID-dependent DNA breaks, as shown in our previous study (9). For the GXXG mutants, the more GXXG motifs were mutated, the less AID interacted with them (Fig. 5*B*, *Top*). Among them, the 3GDDG^R mutant showed very low levels of AID association (8% of WT K^R; Fig. 5*D*), which was consistent with the results of CSR rescue (Fig. 5*B*, *Bottom* and *SI Appendix*, Fig. S5*A*). However, mutating three of the RGG motifs (Mu7^R) was sufficient to decrease the AID–hnRNP K interaction (Fig. 5*B*, *Top*, 4% of WT K^R, Fig. 5*D*), probably because of the conformational change revealed by the faster electrophoretic mobility of the RGG mutants, Mu7^R, Mu8^R, and 5ADD^R compared to the WT K^R; 3GDDG + 5ADD^R also did not interact with AID, in accordance with the result of IgA% (Fig. 5*C* and *D* and *SI Appendix*, Fig. S5*B*). These results indicated that the AID–hnRNP K interaction was mediated by both the GXXG and RGG motifs.

GXXG and RGG Motifs Were Required for the RNA-Binding Capacity of hnRNP K. To confirm that RNA binding by hnRNP K was really impaired by mutation of the GXXG and RGG motifs, RNA-IP using K2-20 cells transfected with sihnRNP K and variously mutated FLAG-tagged hnRNP K plasmids was performed (Fig. 6*A*). IP efficiency evaluated by Western blot analysis showed the same level of IP efficiency in the rescued cells by WT K^R and the motif mutants (*SI Appendix*, Fig. S6 *A* and *B*). Initially, the dependency of the RNA-binding capacity of hnRNP K on the GXXG and RGG motifs was tested by single, double, or triple GXXG mutants (Mu1^R, Mu4^R, and 3GDDG^R) or triple, quadruple, or quintuple RGG mutants (Mu7^R, Mu8^R, or 5ADD^R) (Fig. 6*B*). cMyc mRNA, CARM1 mRNA, and lincRNA-p21, which are the typical hnRNP K-binding RNAs (21, 22, 40), were enriched by WT K^R. The IP efficiency (%input) of all of the tested transcripts was reduced to one-half to one-fourth the level in the

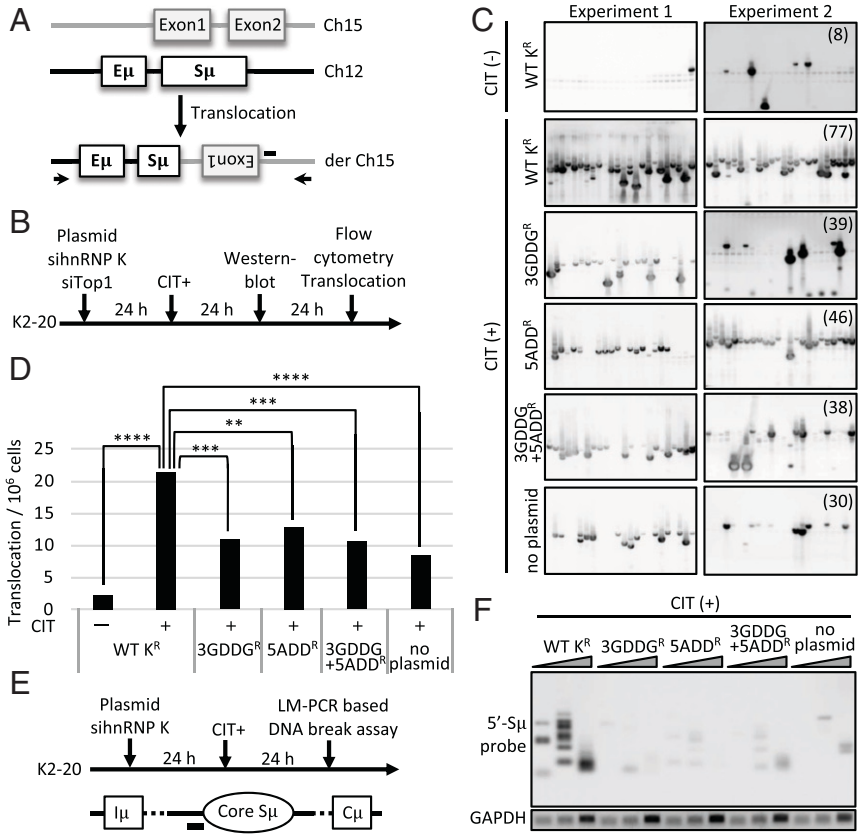

Fig. 4. Analysis of AID-dependent DNA breaks in hnRNP K mutant cells. (*A*) Illustration of PCR amplification for detecting *cMyc/IgH* translocations. Primer positions for amplifying the recombined chromosome are represented as the black arrows. The horizontal black bar shows the position of the *Myc*-specific probe used for the Southern blot hybridization. (*B*) Time course of *cMyc/IgH* translocation assay. sihnRNP K and siTop1 oligos were cotransfected with the WT K^R and its mutant's plasmids into K2-20 cells. (*C*) Southern blot analysis of the indicated samples. The numbers shown in the top right corners describe the number of the translocated cells among the loaded cells in the total 48 lanes (1.5×10^5 cells per lane) of two independent experiments. (*D*) *cMyc/IgH* translocation frequency calculated from the Southern blot analysis. *P* values were calculated using Fisher's exact test. $^{**}P < 0.01$. $^{***}P < 0.001$. $^{****}P < 0001$. (*E*, *Top*) Experimental design for the LM-PCR. K2-20 cells were cotransfected with sihnRNP K and plasmids. (*E*, *Bottom*) The position of the 5′-Sμ probe is marked by a horizontal black bar. (*F*) LM-PCR amplifying the 5′-Sμ region followed by Southern blot with the 5′-Sμ specific probe. The triangles indicate threefold increased amounts of the template. PCR analysis of GAPDH was used as a loading control. F is a representative of three independently repeated experiments.

cells of the single GXXG mutant $Mu1^R$ and further decreased to background levels in the mutants with more mutated GXXG motifs. All of the RGG motif mutants did not enrich these target RNAs, similar to their abolished interaction with AID, shown in Fig. 5*B*. In order to assess whether hnRNP K directly binds to the transcripts derived from AID-dependent DNA break sites, binding of hnRNP K to μ-germline transcript (GLT) and α-GLT including Sμ and Sα regions, respectively, were examined. Transcription of Sμ and Sα regions is known to be critically necessary for CSR and SHM (41–43) and μ- and α-GLT are reported to associate with AID (44); however, these transcripts were not enriched even with the WT K^R. Additionally, Top1 mRNA was not bound by hnRNP K, whereas the AID–hnRNP K complex is supposed to regulate the translation of Top1 (9).

To identify the RNAs which are bound by hnRNP K and functioning for AID-dependent DNA breaks, binding affinities of the candidate RNAs to WT hnRNP K and the functionally deficient mutants, $3GDDG^R$, $5ADD^R$, and 3GDDG + $5ADD^R$ were compared (Fig. 6*C*). Candidate RNAs were selected from the study analyzing hnRNP K-binding RNAs in hippocampal neurons (45). From their list, mRNAs of the proteins related to DNA breaks and/or DNA repair (Sets 1 through 3) and mRNAs having a positive effect on CSR/SHM (Set 3) were selected as the candidate RNAs. Additionally, some long noncoding RNAs (lncRNAs) reported to be bound by hnRNP K and conserved in both human and mouse, such as Lncenc1, Pvt1, SCAT7, and NEAT1 (46, 47), were also screened for their binding to WT K^R and the mutants (Set 4, Fig. 6*C*), although it was not sure whether they were involved in the AID-dependent DNA break step.

Some of the screened mRNAs showed an enrichment signal higher than actin in WT hnRNP K cells (Fig. 6*C*). All of the tested candidate RNAs except for NEAT1 were not efficiently enriched to the RNA-binding motif mutants of hnRNP K, showing the importance of these motifs in hnRNP K's RNA-binding property. However, it is not clear yet whether the direct interaction of RNA to these motifs was abolished in these mutants or structural alteration by these mutations decreased the RNA-binding efficiency. Some candidate mRNAs which showed equivalent or more enrichment than cMyc were tested furthermore for their regulation by hnRNP K (*SI Appendix*, Fig. S6 *C* and *D*). As the result, their protein expression level was comparable between hnRNP K-proficient and -deficient cells, showing that these binding RNAs to hnRNP K were not actually regulated by hnRNP K. Because SMARCA4 and BRD4 showed especially higher enrichment than the positive controls (Fig. 6*C*, Sets 1 and 2), their expression profile was precisely evaluated in the K2-20 cells rescued by WT K^R and the mutants (*SI Appendix*, Fig. S6*D*). However, we could not find any RNA or protein regulated by hnRNP K with these experiments.

NEAT1 expresses two isoforms of the lncRNA NEAT1, longer and shorter (48) (*SI Appendix*, Fig. S7*A*). Primer set NEAT1_2,

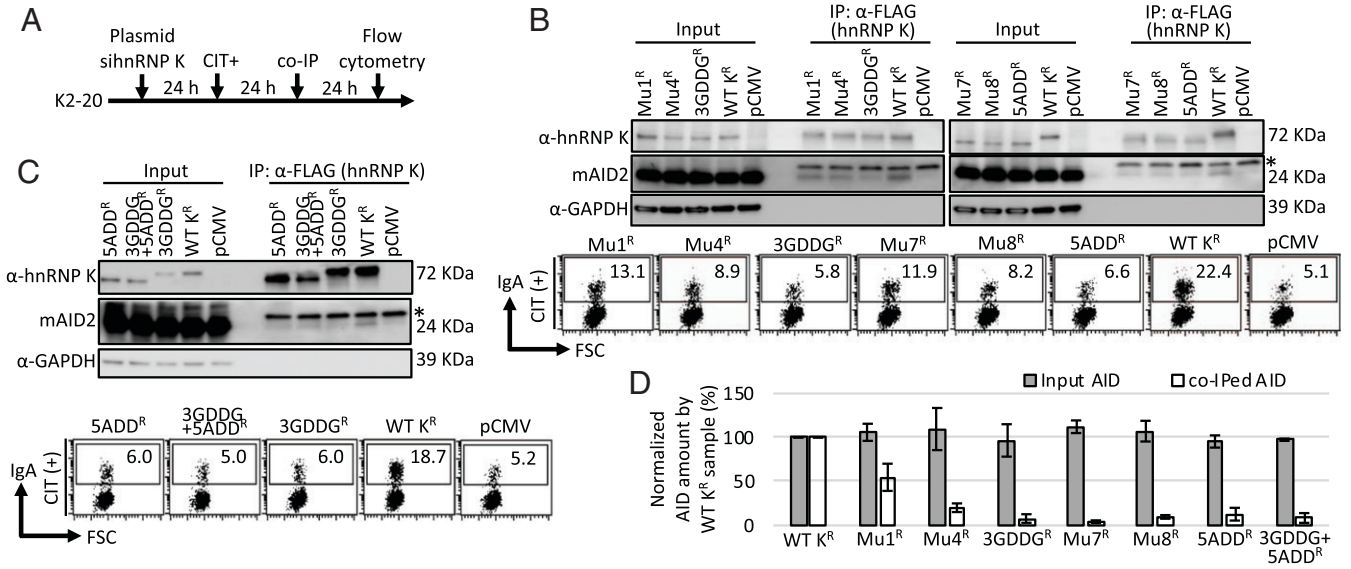

Fig. 5. Comparison of the AID–hnRNP K interaction between WT K^R and the functionally deficient mutants. (*A*) Time course for detecting the interaction between endogenous AID and FLAG-tagged hnRNP K. (*B*, *Top*) Western blot analysis of the co-IP to compare the AID–hnRNP K interaction between the WT K^R and the indicated GXXG mutants (*Left*) or RGG mutants (*Right*). The cell sample transfected with an empty vector (pCMV-3Tag-1A) was used as the negative control (pCMV) and GAPDH was used as the loading control. The asterisk indicates a nonspecific band around 28 kDa. (*B*, *Bottom*) A representative profile of the CSR rescue analysis. FSC, forward scattered. (*C*, *Top*) Western blot analysis of the co-IP to compare the AID–hnRNP K interaction between WT K^R, $3GDDG^R$, $5ADD^R$, and $3GDDG+5ADD^R$. (*C*, *Bottom*) A representative profile of the CSR rescue analysis. (*D*) The averaged signals of coimmunoprecipitated (co-IPed) AID with $3ADD^R$ and $5GDDG^R$ are from three independent experiments, while the signals from other mutants are from two independent experiments. Data are the mean ± SD. The percent values are relative signal calculated by [Input AID(mutant)]/[Input AID(WT K^R)] for Input AID, [co-IPed AID (mutant−pCMV)]/[co-IPed AID (WT K^R−pCMV)] for co-IPed AID. The Western blot signals were quantified by ImageJ software.

which amplified only the longer isoform, showed its higher enrichment to both WT K^R and ADD^R (Fig. 6*C*, Set 4). Although expression level of longer isoform was modestly stabilized by hnRNP K (*SI Appendix*, Fig. S7*B*), knockdown of longer isoform increased IgA switching (*SI Appendix*, Fig. S7 *C* and *D*) by both endogenous AID and exogenous AID-ER (estrogen receptor ligand binding domain). This indicated that regulation of NEAT1 by hnRNP K did not support the function of AID. SHM of the Sμ region in NEAT1 knockdown cells was 75% of the control, but there was no significant difference with this change (*SI Appendix*, Fig. S7 *E* and *F*). Collectively, NEAT1 might independently suppress CSR.

GXXG and RGG Motifs Were Required for Nuclear Localization of hnRNP K. It was reported that the nuclear–cytoplasmic shuttling of AID is essential to induce CSR and SHM (3, 34). Because hnRNP K is also reported to shuttle between the cytoplasm and nucleus (49), functionally deficient mutants of hnRNP K were examined for a change in subcellular localization. Initially, the subcellular localization of WT K^R and $3GDDG^R$ was compared between CIT (+) and CIT (−) conditions (Fig. 7*A* and *SI Appendix*, Fig. S8*A*). The amount of nuclear WT K^R was not affected by CIT stimulation, and the nuclear $3GDDG^R$ was much less than that of the WT K^R. To examine whether the RNA motif mutants generally changed their subcellular localization, the nuclear localization of the other hnRNP K mutants harboring different CSR rescue ability was examined (Fig. 7*B*). For comparison of the relative amount of nuclear hnRNP K, nuclear fraction% (nuclear fraction/[cytoplasmic + nuclear fractions]) was calculated from the intensity data obtained by ImageJ (Fig. 7*E*). Nuclear localization of $3GDDG^R$ was lowest (4% of nuclear fraction%) among the other GXXG motif mutants, while nuclear $Mu1^R$-$Mu3^R$, whose CSR rescue efficiencies were almost half of the WT K^R, were at the middle level (13 to 22%) between $3GDDG^R$ and WT K^R (41%) (Fig. 7*B* and *E*). The RGG mutants $Mu6^R$ and ADD^R localized in the nucleus at the low (9%) and unrecognized (6%) level, respectively. These results of nuclear localization correspond to the CSR rescue ability shown in Table 1. Therefore, the more RNA-binding motifs of hnRNP K that were mutated the less nuclear accumulation of hnRNP K.

To elucidate the requirement of nuclear hnRNP K for CSR, the NLS mutant was tested for its CSR rescue ability. Surprisingly, the nuclear localization of $Mu12^R$, carrying a mutation of the classical NLS (49) (Fig. 7*C*), decreased to 23% (Fig. 7*D* and *E*); however, CSR rescue of $Mu12^R$ was almost normal (90% of the WT K^R, $P = 0.460$; Fig. 7*E* and *SI Appendix*, Fig. S8 *B* and *C*). This result indicated that a nuclear retention of 23% was enough to induce the full strength of CSR. At the same time, $Mu12^R$'s nuclear retention was not as low as that of $3GDDG^R$ or $5ADD^R$ so the possibility still remained that much less (~5% level) nuclear retention of these mutants was disadvantageous for the function of hnRNP K. Because the NLS mutants of AID were functionally dead, the AID–hnRNP K complex was expected to localize in the nucleus to cause AID-dependent DNA breaks (3, 9, 34). Taken together, mutating the RNA-binding motifs decreased the nuclear localization of hnRNP K, which might contribute to the functional deficiency, at least in part, in addition to the abolished RNA-binding capacity and interaction with AID.

Discussion

Our study elucidated the essential RNA-binding motifs in hnRNP K for AID-dependent CSR, SHM, *cMyc*/*IgH* translocation, and DNA breaks (Figs. 1–4 and 8). The function of the GXXG and RGG motifs in AID-dependent DNA breaks was revealed by this study. Especially, the importance of RGG motifs in KI domain for RNA-binding property of hnRNP K was shown by RNA-IP experiments. A previous study found that the loss of function of hnRNP K promotes genomic instability by reducing

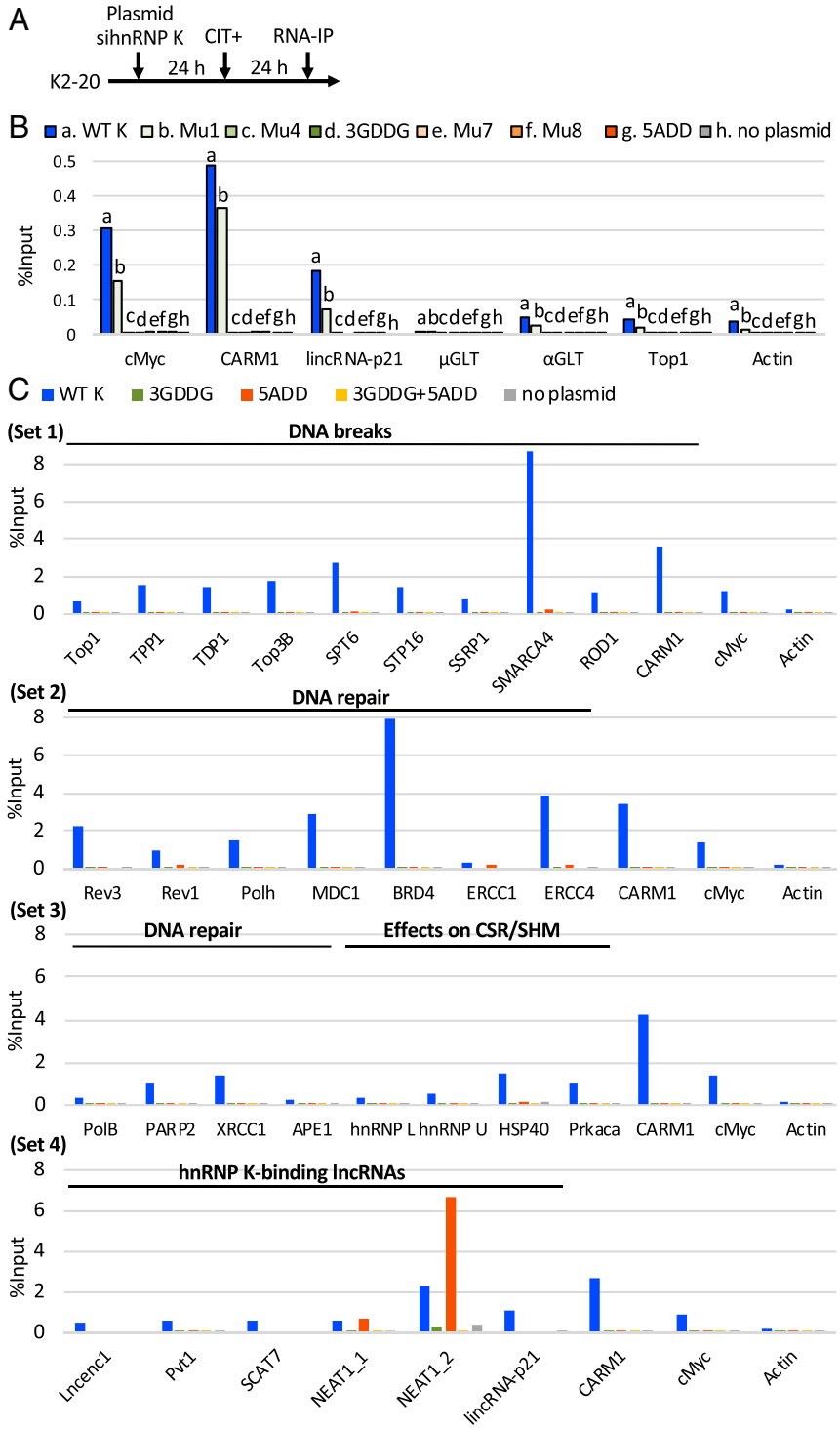

Fig. 6. RNA-binding capacity of the hnRNP K functionally deficient mutants. (*A*) Experimental design for RNA-IP. K2-20 cells with knockdown and the rescue of hnRNP K proteins were collected after CIT stimulation for 24 h, followed by RNA immunoprecipitation (RNA-IP) with an α-FLAG antibody. (*B*) q-PCR analysis for comparing the RNA-binding capacity of WT K^R and its mutants shown on the top of the bar graph. %Input indicates the enrichment of the immunoprecipitated RNAs normalized by the input signal. Actin mRNA was used as the negative control. (*C*) q-PCR analysis for detecting the enrichment of the reported hnRNP K-binding mRNAs related to AID function (mRNAs encoding proteins required for inducing DNA breaks, DNA repair, or AID-interaction, Sets 1 through 3), or the reported hnRNP K-binding lncRNAs (Set 4), using WT K^R, 3GDDG^R, 5ADD^R and 3GDDG+5ADD^R. LincRNA-p21, mRNA of CARM1, and cMyc were positive controls; actin mRNA was the negative control. The results in *B* and *C* were confirmed by three independent experiments.

the genome maintenance activity of p53 (24), which is the opposite effect on genome instability to our study. Also, hnRNP K is considered to be important for the DNA repair step after DNA damage (23, 24). However, our study clearly showed that the RNA-binding motifs of hnRNP K were essential in the DNA break step, but not the repair step, in AID-dependent Ig gene

recombination, unveiling the RNA-mediated function of hnRNP K in the formation of programmed endogenous DNA breaks.

We aimed to identify the binding RNA which is regulated by hnRNP K and contributes to AID-dependent DNA breaks. We screened binding of hnRNP K to candidate RNAs including μ- and α-GLTs transcribed from AID-dependent break sites, Top1 mRNA, and mRNAs encoding DNA break- or repair-related protein and lncRNAs reported to bind to hnRNP K. Unfortunately, we could not identify any RNA which could explain the AID-dependent DNA break mechanism (Fig. 6*B* and *C* and *SI Appendix*, Figs. S6 and S7). This result showing that μ- and α-GLTs are not bound by hnRNP K was reasonable because our previous study showed that knockdown of hnRNP K did not affect the expression of μ- and α-GLTs (9). Additionally, the result in which hnRNP K did not bind to Top1 mRNA was supported by the effect of siTop1 on CSR (*SI Appendix*, Fig. S4*B* and *C*). In the RNA-binding motif mutants of hnRNP K, siTop1 up-regulates IgA%; however, the up-regulated IgA% does not reach to the standard level of WT K^R (without siTop1, ~24%). If low levels of DNA breaks in these RNA-binding motif mutants were due to the disturbance of the Top1 reduction, the artificial reduction of Top1 by siRNA would overcome the defect and recover CSR efficiency, but it did not. This result suggested that lower IgA% values in these mutants were not related to Top1 reduction by AID (12, 13).

In our study, it was not solved that the loss of RNA-binding is due to decrease of direct RNA-binding capacity or the changes of whole structural and/or posttranslational modification (PTM) (*SI Appendix*, Fig. S9). RNA binding by GXXG motifs is well-defined by the many studies revealing that GXXG favored pyrimidine-rich sequences (32). Meanwhile arginine of RGG generally binds to some RNAs if the arginine can make π stack and hydrogen bond with the bases (33), suggesting that RNA binding by RGG does not show any sequence specificity. Actually there is very little direct evidence reported on RNA-binding capacity of the RGG motifs in KI domain. In vitro transcribed 5′ UTR of enterovirus 71 binds to the hnRNP K mutant in KI domain-dependent manner, although the effect of conformation change is not excluded (50). In the other study of RNA-binding ability of RGG motifs, the ^{296}RGGR-deleted mutant in KI domain failed to bind to the c-FOS mRNA-CAT mRNA fusion RNA probe in the RNA-electrophoretic mobility assay (51). Because this mutant did not show the large difference of electrophoretic mobility with WT hnRNP K, possibly this result suggests some direct RNA binding. In contrast, structural change in the RGG mutants would give multiple effects on the function of hnRNP K. Furthermore, RGG to ADD mutations of ^{258}R, ^{268}R, ^{296}R, and ^{299}R in hnRNP K in our study presumably abolished arginine methylation which might affect both RNA binding directly or through the acceleration of phosphorylation by c-Src (52–55).

Interestingly, the $5ADD^R$ mutant was functionally defective in DNA breaks but did not totally lose binding capacity to some RNA, such as NEAT1, showing that intact GXXG motifs still support the binding of some of the target RNAs. The evidence for the direct RNA-binding ability of RGG motifs was not sufficient, because mutation of RGG motifs may initially change the hnRNP K's structure and this causes the defective RNA-binding ability and less interaction with AID. At the same time, remaining RNA-binding capacity of $5ADD^R$ suggests that the subtraction of RNAs binding to $5ADD^R$ from RNAs binding to WT K^R may narrow the candidate RNAs, which will provide the basis of AID-dependent DNA breaks.

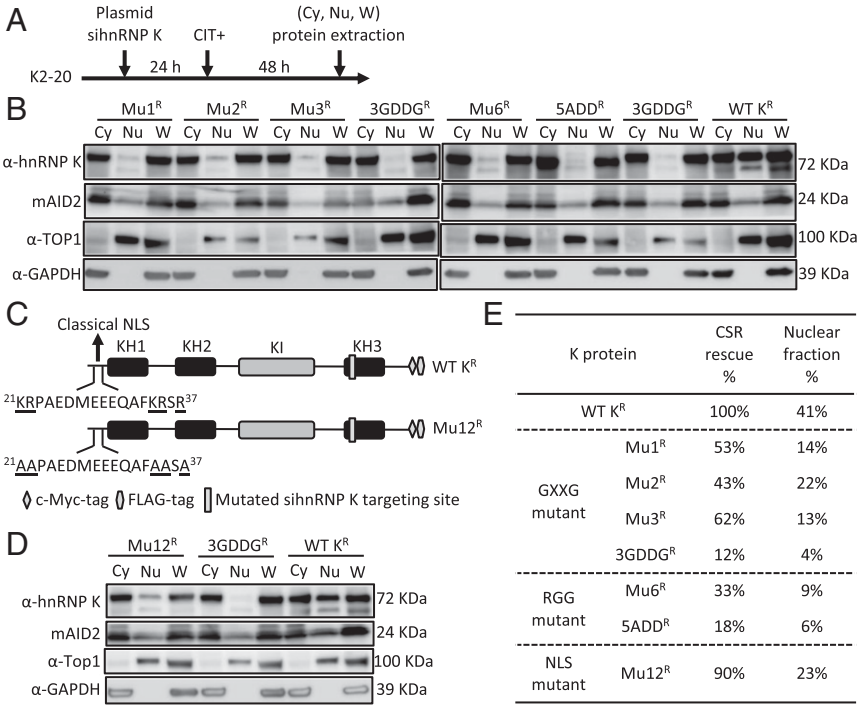

Fig. 7. Subcellular localization of the GXXG mutants, RGG mutants, and NLS mutant. (*A*) Time course for the detection of subcellular localization. A similar CSR restoration experiment was performed using K2-20 cells. Cytoplasmic (Cy), nuclear (Nu), and whole-cell (W) proteins were separately extracted. (*B*) Western blot analysis comparing the subcellular localization of WT K^R and its mutants. Top1 was used as a nuclear marker; GAPDH was used as a cytoplasmic marker. (*C*) The mutated amino acid residues in the NLS are positioned in the scheme of the WT K^R construct (*Top*). Amino acid change in the NLS mutant $Mu12^R$ (*Bottom*). (*D*) Western blot analysis detecting the subcellular localization of $3GDDG^R$, $Mu12^R$, and WT K^R. The result was confirmed by two independent experiments. (*E*) Comparison of the CSR rescue abilities and subcellular distribution. Normalized CSR rescue was calculated as shown in Table 1. Nuclear fraction% was calculated using the values quantified by the ImageJ software.

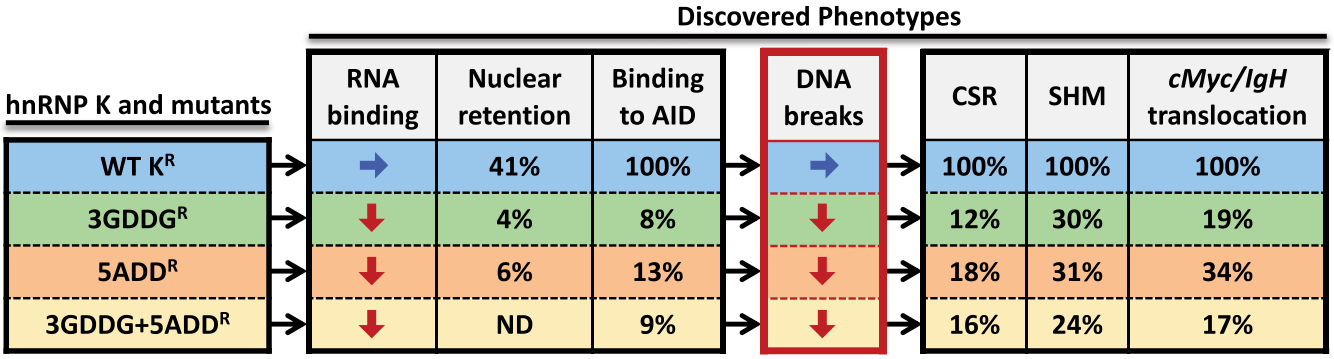

Fig. 8. Summary for the requirement of RNA-binding motifs in hnRNP K for AID-dependent DNA breaks. The RNA-binding motifs (3GXXG and 5RGG) in hnRNP K are required for its RNA-binding capacity and nuclear retention, as well as the interaction with AID, which are supposed to be necessary for inducing DNA breaks in IgH locus, followed by CSR, SHM, and *cMyc/IgH* translocation. Blue arrow, WT level of hnRNP K. Red arrow, significantly decreased. The percent values in the parentheses are relative values of WT K^R normalized by no plasmid samples, except for the analysis of AID-binding (by pCMV sample) and nuclear retention (by Cy + Nu fractions). ND, not done.

In our study, GXXG and/or RGG motif mutations in hnRNP K affected not only binding to RNAs and interaction with AID but also its nuclear accumulation, unexpectedly. Because hnRNP K has the NLS motif, the reason why loss of RNA-binding capacity leads to decreased nuclear retention remains elusive. However, this decrease might be influenced by lower RNA binding in both the GXXG and RGG mutants and structural change and PTM particularly in the RGG mutants (*SI Appendix*, Fig. S9). Notably, recruitment of hnRNP K to the specific genomic loci is guided by the binding of lncRNAs (21, 56), suggesting that the RNA-binding property of hnRNP K determines its fate. Additionally, nuclear retention is possibly affected by arginine methylation (57). Taken together, nuclear accumulation was affected by RNA-binding capacity at least in part and both may collaboratively promote the complex formation of hnRNP K with AID and, further, AID-dependent DNA breaks.

Recently, the overexpression of hnRNP K in B lymphoma cells was reported to be a poor-prognosis marker because it enhances cMyc expression (27). However, a loss-of-function mutation in hnRNP K is found to be one of the responsible genes of Kabuki-like syndrome. In general, Kabuki-like syndrome patients present multiple congenital anomalies, various levels of hypogammaglobulinemia (42 to 79%), and repeated infection (26 to 42%) (58, 59). Because Kabuki-like syndrome is caused not only by hnRNP K mutation but also by MLL2 (lysine methyltransferase 2D) and KDM6A (lysine demethylase 6A) mutation (58–60), the hypogammaglobulinemia symptoms are not solely attributed to hnRNP K mutation. Actually, a double mutation of MLL3 and MLL4 (MLL2 in humans) in mice decreases switching of IgG1 and IgG3 in vitro to almost half the level of WT mice (61). At the time of preparation of this paper, 10 patients with Kabuki-like syndrome caused by hnRNP K mutations were reported (62, 63) and only one of them presented repeated respiratory infection (59). Almost all their mutations happened in KI or KH3 domain. These mutations might be enough to disturb the function of hnRNP K in the other organs, but not enough to affect CSR and SHM in the activated B cells. Probably some population in the previous Kabuki-like syndrome patients who showed hypogammaglobulinemia might be possibly caused by a mutation of hnRNP K. Our study will provide the clue for understanding the immunodeficiency in the hnRNP K-mutated patients.

Technically NGS was adopted to evaluate SHM, as the other group reported (64, 65). Actually, the background SHM frequency was higher (5.1 to 5.3 $\times$ 10^{-4}; Fig. 3*D*) than that of the Sanger method (1.6 to 2.7 $\times$ 10^{-4}; *SI Appendix*, Fig. S3 *B* and *C*) because NGS elucidated all of the "mutations," changed nucleotides from the reference including allelic mutations, single-nucleotide polymorphisms (SNPs) developed during culture, spontaneous AID-independent mutations, and technical sequencing errors in addition to AID-dependent mutation. We tried to minimize PCR bias by increasing the starting DNA amount. Because 3 pg of DNA was supposed to be derived from a single cell, ~320,000 cells, contained in 960 ng DNA per sample, were used in the initial PCR step. From the total read number of the samples (370,000 to 510,000 reads per sample; *SI Appendix*, Table S1), one cell was probably sequenced 1.2 to 1.6 times. Considering all these conditions, the total SHM was calculated by excluding 1) only one mutation event out of all reads and 2) the specific mutations with a frequency of more than 10%, which are allelic mutations and major SNPs. The final SHM results comparing WT and the RNA-binding motif mutants obtained by this protocol reasonably correlated with the level of CSR, DNA breaks, and *cMyc/IgH* translocation and were considered to represent AID-dependent mutations.

In summary, this study suggested that the RNA-binding motifs of hnRNP K, GXXG and RGG, were identified as the necessary motifs for RNA binding, nuclear retention, and the interaction with AID. These results supported our hypothesis that hnRNP K presents some RNAs to AID for editing, and the edited RNAs provoke DNA breaks in IgH locus, following on several AID-mediated processes—CSR, SHM, and *cMyc/IgH* translocation (Fig. 8). Still, the molecular mechanism of AID-dependent DNA breaks is highly enigmatic—Which RNA is edited by AID, what is the function of the edited RNA, and how is the DNA cut? However, the remaining questions will be answered when RNAs associated with hnRNP K are analyzed. In the future, comparing the trapped RNAs between the motif mutants and the WT hnRNP K will give an important clue to the molecular mechanism of AID-induced DNA breaks.

Materials and Methods

Detailed materials and methods for mutational constructs, cell culture, CSR rescue assay, Western blot analysis, SHM analysis, *cMyc/IgH* translocation assay, LM-PCR–based DNA break assay, co-IP, RNA-IP, NEAT1-related experiments, fractionation of cytoplasmic and nuclear proteins, and statistical analyses are described in *SI Appendix*.

Data and Material Availability. The row data of SHM analysis by NGS was uploaded to https://www.ncbi.nlm.nih.gov/bioproject/612426 (66). Python-formatted files used in informatics analysis have been deposited at https://github.com/makikbys/somatichypermutation (67). Other data, as well as associated-protocols, are available in the main manuscript and *SI Appendix*. Cell lines and plasmids are available from the corresponding author upon request.

ACKNOWLEDGMENTS. We thank Dr. Misao Takemoto, Ms. Maki Sasamuma, and Ms. Nakata Mikiyo for their kind technical assistance and members of the T.H. laboratory for sharing reagents and for meaningful daily discussions. This research was supported by Japan Society for the Promotion of Science (JSPS) KAKENHI Grant-in-Aid for Scientific Research (A) 19H01027 to T.H. and Grant-in-Aid for Scientific Research (C) 19K06485 to M.K. This research was also supported by JSPS KAKENHI Grant 16H06279 (Platform for Advanced Genome Science). Z.Y. received support for her PhD scholarship from the China Scholarship Council and the Japanese Government (Ministry of Education, Culture, Sports, Science and Technology).

1. M. Muramatsu *et al*., Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* **102**, 553–563 (2000).
2. V.-T. Ta *et al*., AID mutant analyses indicate requirement for class-switch-specific cofactors. *Nat. Immunol.* **4**, 843–848 (2003).
3. R. Shinkura *et al*., Separate domains of AID are required for somatic hypermutation and class-switch recombination. *Nat. Immunol.* **5**, 707–712 (2004).
4. J. Daly *et al*., Altered Ig hypermutation pattern and frequency in complementary mouse models of DNA polymerase ζ activity. *J. Immunol.* **188**, 5528–5537 (2012).
5. J. Xu, A. Husain, W. Hu, T. Honjo, M. Kobayashi, APE1 is dispensable for S-region cleavage but required for its repair in class switch recombination. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 17242–17247 (2014).
6. N. A. Begum, H. Nagaoka, M. Kobayashi, T. Honjo, *Molecular Mechanisms of AID Function* (Academic Press, 2014).
7. M. S. Neuberger, R. S. Harris, J. Di Noia, S. K. Petersen-Mahrt, Immunity through DNA deamination. *Trends Biochem.Sci.* **28**, 305–312 (2003).
8. V. Shivarov, R. Shinkura, T. Honjo, Dissociation of in vitro DNA deamination activity and physiological functions of AID mutants. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 15866–15871 (2008).
9. W. Hu, N. A. Begum, S. Mondal, A. Stanlie, T. Honjo, Identification of DNA cleavage- and recombination-specific hnRNP cofactors for activation-induced cytidine deaminase. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 5791–5796 (2015).
10. S. Mondal, N. A. Begum, W. Hu, T. Honjo, Functional requirements of AID's higher order structures and their interaction with RNA-binding proteins. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E1545–E1554 (2016).
11. G. Liang *et al*., RNA editing of hepatitis B virus transcripts by activation-induced cytidine deaminase. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 2246–2251 (2013).
12. M. Kobayashi *et al*., AID-induced decrease in topoisomerase 1 induces DNA structural alteration and DNA cleavage for class switch recombination. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 22375–22380 (2009).
13. M. Kobayashi *et al*., Decrease in topoisomerase I is responsible for activation-induced cytidine deaminase (AID)-dependent somatic hypermutation. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 19305–19310 (2011).
14. M. S. Swanson, G. Dreyfuss, Classification and purification of proteins of heterogeneous nuclear ribonucleoprotein particles by RNA-binding specificities. *Mol. Cell. Biol.* **8**, 2237–2241 (1988).
15. S. Piñol-Roma, Y. D. Choi, M. J. Matunis, G. Dreyfuss, Immunopurification of heterogeneous nuclear ribonucleoprotein particles reveals an assortment of RNA-binding proteins. *Genes Dev.* **2**, 215–227 (1988).
16. M. J. Matunis, W. M. Michael, G. Dreyfuss, Characterization and primary structure of the poly(C)-binding heterogeneous nuclear ribonucleoprotein complex K protein. *Mol. Cell. Biol.* **12**, 164–171 (1992).
17. W. Cao, A. Razanau, D. Feng, V. G. Lobo, J. Xie, Control of alternative splicing by forskolin through hnRNP K during neuronal differentiation. *Nucleic Acids Res.* **40**, 8059–8071 (2012).
18. A. Skalweit *et al*., Posttranscriptional control of renin synthesis: Identification of proteins interacting with renin mRNA 3′-untranslated region. *Circ. Res.* **92**, 419–427 (2003).
19. M. Yano, H. J. Okano, H. Okano, Involvement of Hu and heterogeneous nuclear ribonucleoprotein K in neuronal differentiation through p21 mRNA post-transcriptional regulation. *J. Biol. Chem.* **280**, 12690–12699 (2005).
20. Y. Kawasaki *et al*., MYU, a target lncRNA for Wnt/c-Myc signaling, mediates induction of CDK6 to promote cell cycle progression. *Cell Rep.* **16**, 2554–2564 (2016).
21. M. Huarte *et al*., A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* **142**, 409–419 (2010).
22. M. Notari *et al*., A MAPK/HNRPK pathway controls BCR/ABL oncogenic potential by regulating MYC mRNA translation. *Blood* **107**, 2507–2516 (2006).
23. A. Moumen, P. Masterson, M. J. O'Connor, S. P. Jackson, hnRNP K: An HDM2 target and transcriptional coactivator of p53 in response to DNA damage. *Cell***123**, 1065–1078 (2005).
24. N. Wiesmann *et al*., Knockdown of hnRNPKleads to increased DNA damage after irradiation and reduces survival of tumor cells. *Carcinogenesis* **38**, 321–328 (2017).
25. A. Folci *et al*., Loss of hnRNP K impairs synaptic plasticity in hippocampal neurons. *J. Neurosci.* **34**, 9088–9095 (2014).
26. M. Gallardo *et al*., Aberrant hnRNP K expression: All roads lead to cancer. *Cell Cycle* **15**, 1552–1557 (2016).
27. M. Gallardo *et al*., Uncovering the role of RNA-binding protein hnRNP K in B-cell lymphomas. *J. Natl. Cancer Inst.* **112**, 95–106 (2020).
28. M. Gallardo *et al*., HnRNP K is a haploinsufficient tumor suppressor that regulates proliferation and differentiation programs in hematologic malignancies. *Cancer Cell* **28**, 486–499 (2015).
29. K. Bomsztyk, O. Denisenko, J. Ostrowski, hnRNP K: One protein multiple processes. *BioEssays* **26**, 629–638 (2004).
30. A. V. Makeyev, S. A. Liebhaber, The poly(C)-binding proteins: A multiplicity of functions and a search for mechanisms. *RNA* **8**, 265–278 (2002).
31. H. Siomi, M. Choi, M. C. Siomi, R. L. Nussbaum, G. Dreyfuss, Essential role for KH domains in RNA binding: Impaired RNA binding by a mutation in the KH domain of FMR1 that causes fragile X syndrome. *Cell* **77**, 33–39 (1994).
32. D. T. Braddock, J. L. Baber, D. Levens, G. M. Clore, Molecular basis of sequence-specific single-stranded DNA recognition by KH domains: Solution structure of a complex between hnRNP K KH3 and single-stranded DNA. *EMBO J.* **21**, 3476–3485 (2002).
33. P. A. Chong, R. M. Vernon, J. D. Forman-Kay, RGG/RG motif regions in RNA binding and phase separation. *J. Mol. Biol.* **430**, 4650–4665 (2018).
34. S. Ito *et al*., Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 1975–1980 (2004).
35. P. H. Backe, A. C. Messias, R. B. G. Ravelli, M. Sattler, S. Cusack, X-ray crystallographic and NMR studies of the third KH domain of hnRNP K in complex with single-stranded nucleic acids. *Structure* **13**, 1055–1067 (2005).
36. D. Hollingworth *et al*., KH domains with impaired nucleic acid binding as a tool for functional analysis. *Nucleic Acids Res.* **40**, 6873–6886 (2012).
37. A. T. Phan *et al*., Structure-function studies of FMRP RGG peptide recognition of an RNA duplex-quadruplex junction. *Nat. Struct. Mol. Biol.* **18**, 796–804 (2011).
38. R. Küppers, R. Dalla-Favera, Mechanisms of chromosomal translocations in B cell lymphomas. *Oncogene* **20**, 5580–5594 (2001).
39. A. R. Ramiro *et al*., Role of genomic instability and p53 in AID-induced c-myc-Igh translocations. *Nature* **440**, 105–109 (2006).
40. A. Liepelt *et al*., Translation control of TAK1 mRNA by hnRNP K modulates LPS-induced macrophage activation. *RNA* **20**, 899–911 (2014).
41. S. Jung, K. Rajewsky, A. Radbruch, Shutdown of class switch recombination by deletion of a switch region control element. *Science* **259**, 984–987 (1993).
42. A. Peters, U. Storb, Somatic hypermutation of immunoglobulin genes is linked to transcription initiation. *Immunity* **4**, 57–65 (1996).
43. J. Zhang, A. Bottaro, S. Li, V. Stewart, F. W. Alt, A selective defect in IgG2b switching as a result of targeted mutation of the I gamma 2b promoter and exon. *EMBO J.* **12**, 3529–3537 (1993).
44. S. Zheng *et al*., Non-coding RNA generated following lariat debranching mediates targeting of AID to DNA. *Cell* **161**, 762–773 (2015).
45. G. Leal *et al*., The RNA-binding protein hnRNP K mediates the effect of BDNF on dendritic mRNA metabolism and regulates synaptic NMDA receptors in hippocampal neurons. *eNeuro* **4**, ENEURO.0268-17 (2017).
46. Y. Lubelsky, I. Ulitsky, Sequences enriched in Alu repeats drive nuclear localization of long RNAs in human cells. *Nature* **555**, 107–111 (2018).
47. Y. Xu *et al*., New insights into the interplay between non-coding RNAs and RNA-binding protein hnRNPK in regulating cellular functions. *Cells* **8**, 62 (2019).
48. T. Naganuma *et al*., Alternative 3′-end processing of long noncoding RNA initiates construction of nuclear paraspeckles. *EMBO J.* **31**, 4020–4034 (2012).
49. W. M. Michael, P. S. Eder, G. Dreyfuss, The K nuclear shuttling domain: A novel signal for nuclear import and nuclear export in the hnRNP K protein. *EMBO J.* **16**, 3587–3598 (1997).
50. J. Y. Lin *et al*., Heterogeneous nuclear ribonuclear protein K interacts with the enterovirus 71 5′ untranslated region and participates in virus replication. *J. Gen. Virol.***89**, 2540–2549 (2008).
51. M. H. Lee, S. Mori, P. Raychaudhuri, trans-Activation by the hnRNP K protein involves an increase in RNA synthesis from the reporter genes. *J. Biol. Chem.* **271**, 3420–3427 (1996).
52. A. Ostareck-Lederer *et al*., Asymmetric arginine dimethylation of heterogeneous nuclear ribonucleoprotein K by protein-arginine methyltransferase 1 inhibits its interaction with c-Src. *J. Biol. Chem.* **281**, 11115–11125 (2006).
53. H. Gross *et al*., Binding of the heterogeneous ribonucleoprotein K (hnRNP K) to the Epstein-Barr virus nuclear antigen 2 (EBNA2) enhances viral LMP2A expression. *PLoS One* **7**, e42106 (2012).
54. A. C. Messias, C. Harnisch, A. Ostareck-Lederer, M. Sattler, D. H. Ostareck, The DICE-binding activity of KH domain 3 of hnRNP K is affected by c-Src-mediated tyrosine phosphorylation. *J. Mol. Biol.* **361**, 470–481 (2006).
55. T. A. Tahir, H. Singh, N. P. J. Brindle, The RNA binding protein hnRNP-K mediates post-transcriptional regulation of uncoupling protein-2 by angiopoietin-1. *Cell. Signal.* **26**, 1379–1384 (2014).
56. Z. Sun *et al*., The long noncoding RNA Lncenc1 maintains naive states of mouse ESCs by promoting the glycolysis pathway. *Stem Cell Reports* **11**, 741–755 (2018).
57. Y. I. Chang *et al*., Identification of the methylation preference region in heterogeneous nuclear ribonucleoprotein K by protein arginine methyltransferase 1 and its implication in regulating nuclear/cytoplasmic distribution. *Biochem.Biophys. Res. Commun.* **404**, 865–869 (2011).
58. J. D. Hoffman *et al*., Immune abnormalities are a frequent manifestation of Kabuki syndrome. *Am. J. Med. Genet. A.* **135**, 278–281 (2005).
59. M. L. Dentici *et al*., Clinical spectrum of Kabuki-like syndrome caused by HNRNPK haploinsufficiency. *Clin.Genet.* **93**, 401–407 (2018).

60. S. Banka *et al*., How genetically heterogeneous is Kabuki syndrome?: MLL2 testing in 116 patients, review and analyses of mutation and phenotypic spectrum. *Eur. J. Hum. Genet.* **20**, 381–388 (2012).
61. L. M. Starnes *et al*., A PTIP-PA1 subcomplex promotes transcription for IgH class switching independently from the associated MLL3/MLL4 methyltransferase complex. *Genes Dev.* **30**, 149–163 (2016).
62. P. Y. B. Au *et al*.; Care for Rare Canada Consortium, Phenotypic spectrum of Au-Kline syndrome: A report of six new cases and review of the literature. *Eur. J. Hum. Genet.* **26**, 1272–1281 (2018).
63. N. Miyake *et al*., A case of atypical Kabuki syndrome arising from a novel missense variant in HNRNPK. *Clin.Genet.* **92**, 554–555 (2017).
64. Q. Wang *et al*., The cell cycle restricts activation-induced cytidine deaminase activity to early G1. *J. Exp. Med.* **214**, 49–58 (2017).
65. V. Delgado-Benito *et al*., The chromatin reader ZMYND8 regulates Igh enhancers to promote immunoglobulin class switch recombination. *Mol. Cell* **72**, 636–649.e8 (2018).
66. M. Kobayashi, Raw data of next-generation sequencing (NGS) in [RNA-binding motifs of hnRNP K are critical for induction of antibody diversification by activation-induced cytidine deaminase. Yin & Kobayashi et al]. BioProject in NCBI. https://www.ncbi.nlm.nih.gov/bioproject/612426. Deposited 13 March 2020.
67. M. Kobayashi, SHM-Commond for analyzing NGS data. GitHub. https://github.com/makikbys/somatichypermutation. Deposited 13 March 2020.