Article:
Thwaites, DT orcid.org/0000-0002-1504-7712, Carter, C, Lawless, D orcid.org/0000-0001-8496-3725 et al. (2 more authors) (2019) A novel RAG1 mutation reveals a critical in vivo role for HMGB1/2 during V(D)J recombination. Blood, 133 (8). pp. 820-829. ISSN 0006-4971

https://doi.org/10.1182/blood-2018-07-866939

A novel *RAG1* mutation reveals a critical *in vivo* role for HMGB1/2 during V(D)J recombination

Daniel T. Thwaites[1], Clive Carter[2], Dylan Lawless[2], Sinisa Savic[2,3] and Joan M. Boyes[1,*]

[1]School of Molecular and Cellular Biology, University of Leeds, Leeds, LS2 9JT, UK
[2]Department of Clinical Immunology and Allergy, St James's University Hospital, Leeds, UK
[3]National Institute for Health Research - Leeds Biomedical Research Centre and Leeds Institute of Rheumatic and Musculoskeletal Medicine, Wellcome Trust Brenner Building, St James's University Hospital, Beckett Street, Leeds, LS9 7TF, U.K

[*]To whom correspondence should be addressed:
School of Molecular and Cellular Biology,
University of Leeds,
Leeds,
LS2 9JT,
UK
Tel: +44(0)113 343 3147
Email: j.m.boyes@leeds.ac.uk

Short title: **HMGB1 is essential for V(D)J recombination *in vivo***

Word count, text: 4000
Word count, abstract: 191
Figure count: 6
Table count: 1 Supplemental
Reference count: 36

Scientific category: Immunobiology and Immunotherapy

Key Points:

- A novel *RAG1* mutation significantly reduces binding of the accessory protein, HMGB1, to diminish V(D)J recombination *in vitro* and *in vivo*
- Complementary knock-down studies of HMGB1 provide the first compelling evidence that HMGB1 is essential for V(D)J recombination *in vivo*

Abstract

The Recombination Activating Genes, *RAG1* and *RAG2*, are essential for V(D)J recombination and adaptive immunity. Mutations in these genes often cause immunodeficiency, the severity of which reflects the importance of the altered residue(s) during recombination. Here, we describe a novel *RAG1* mutation that causes immunodeficiency in an unexpected way: The mutated protein severely disrupts binding of the accessory protein, HMGB1. Although HMGB1 enhances RAG cutting *in vitro*, its role *in vivo* was controversial. We show here that reduced HMGB1 binding by the mutant protein dramatically reduces RAG cutting *in vitro* and almost completely eliminates recombination *in vivo*. The RAG1 mutation, R401W, places a bulky tryptophan opposite the binding site for HMG Box A at both 12- and 23-spacer recombination signal sequences, disrupting stable binding of HMGB1. Replacement of R401W with leucine and then lysine, progressively restores HMGB1 binding, correlating with increased RAG cutting and recombination *in vivo*. We show further that knock-down of HMGB1 significantly reduces recombination by wild type RAG1 whilst its re-addition restores recombination with wild-type, but not the mutant RAG1 protein. Together, these data provide compelling evidence that HMGB1 plays a critical role during V(D)J recombination *in vivo*.

Introduction

The ability of vertebrates to combat a vast range of potential pathogens critically relies on V(D)J recombination. This stochastically mixes and matches individual V, D and J gene segments in the immunoglobulin and T cell receptor loci to generate a huge array of variable exons, that collectively encode more than 10^7 different antigen binding sites[1]. The importance of V(D)J recombination is demonstrated by mutations in the proteins involved in the cutting or joining steps of the reaction: the outcome is invariably combined immunodeficiency (CID)[2].

Only two proteins are essential for initiation of V(D)J recombination: RAG1 and RAG2. These lymphoid-specific proteins bind to recombination signal sequences (RSSs) that lie adjacent to V, D and J gene segments and consist of conserved heptamer and nonamer sequences, separated by a relatively non-conserved 12±1 or 23±1 bp spacer[3]. Recombination proceeds by RAG proteins bringing two complementary RSSs (i.e. 12-RSS + 23-RSS) into a synaptic complex[4]. Coupled cleavage of the RSSs is then achieved by RAG1 binding to the nonamer of one RSS whilst RAG2 directs the RAG1 catalytic DDE motif to nick the partner RSS precisely at the heptamer/coding junction[5]; this ultimately generates blunt double strand breaks at the two RSSs and hairpin structures at the two coding ends. Subsequent processing and joining of the broken DNA ends is normally carried out by the classical non-homologous end joining (NHEJ) machinery[6].

Complete loss of function of either RAG protein leads to T-, B- severe combined immunodeficiency (SCID)[2]. Similarly, inactivating mutations in the NHEJ proteins also result in SCID but in this case, the immunodeficiency is accompanied by increased cellular radio-sensitivity due to defects in DNA repair. Hypomorphic RAG mutations that retain some residual recombination activity result in Omenn syndrome which, like SCID, generally manifests soon after birth. More recently, milder forms of CID have been reported in older children with less inactivating *RAG* mutations. By studying the effects of individual *RAG* mutations, a strong correlation has been noted between the

severity of the immunodeficiency and the importance of the mutated amino acids to RAG function[2].

Although RAG1 and RAG2 are sufficient for cleavage of an RSS substrate *in vitro,*[7] the high mobility group box (HMGB) proteins, HMGB1 or HMGB2, were found to increase cutting by 7- to 100-fold[8] and to decrease the *K*d of the RAG complex for both 12- and 23-RSSs[9]. HMGB proteins are an abundant group of non-specific DNA binding proteins that bind in the minor groove to facilitate DNA bending[10]. Band shift assays showed that HMGB1/2 predominantly increases RAG binding to 23-RSSs, where it was proposed to bend the 23 bp spacer to bring the nonamer and heptamer sequences closer together[8], a prediction that was confirmed by recent single molecule fluorescence experiments[11] and structural studies[12,13]. HMGB1 also stabilises the bend at the 12-RSS to help to position the RAG complex for correct cleavage of its target site[13,14]. HMGB1 has two DNA binding regions, HMG boxes A and B and high resolution structures of the RAG/RSS complex showed that a single HMG box (likely Box A) binds to the 12-RSS spacer, whereas two HMG boxes bind to the 23-RSS spacer, where the additional binding stabilises a 150° DNA bend[12,13].

Whilst the role of HMGB1/2 in V(D)J recombination has been thoroughly investigated *in vitro*, there is less evidence that HMGB1/2 is required *in vivo*. HMGB1 stimulates RAG cutting of nucleosome templates *in vitro*, implying that HMGB1/2 potentially increases the accessibility of RSSs to RAGs *in vivo*[15]. Subsequently, recombination of an extrachromosomal substrate was found to increase by >3 fold when additional HMGB1/2 was provided by transfection[16]. In contrast, HMGB1 knockout mice displayed normal levels of serum immunoglobulins and a complete T cell receptor repertoire after birth, raising doubts about the requirement of HMGB1 for V(D)J recombination *in vivo*[17]. It is possible, however, that HMGB2 which has 79% amino acid identity and is expressed in lymphoid tissues, substitutes for HMGB1 to promote recombination.

Here, we analyse the effects of compound heterozygous RAG1 mutations from an adult patient with primary immunodeficiency. One of these mutations permits residual RSS binding and recombination, accounting for the patient's relatively mild phenotype. The other mutation, however, has profound effects on HMGB1 binding *in vitro* and on V(D)J recombination *in vivo*, providing strong evidence that HMGB1/2 is essential for efficient V(D)J recombination *in vivo*.

Methods

RAG constructs

To express RAG proteins in 293T cells, maltose binding protein (MBP)-tagged core RAG1 and core RAG2 cDNAs were cloned into pEF-XC[18]. Full length RAG1 cDNA was cloned into pCS2MT (Clontech, Mountain View, CA, USA) and used in transfection-based recombination assays. RAG1 mutations were generated using the NEB Q5® Site-Directed Mutagenesis Kit, using primers 8-15 (Table S1).

Extrachromosomal V(D)J recombination assay

NIH3T3 cells were seeded at 2 x 10^5 cells per well in a 6-well plate and transfected with 1 µg of the recombination substrate pJH299[19], 160 ng of wild type (WT) or mutant pCS2MT-RAG1 and 320 ng pEFXC-RAG2 using polyethylenimine (PEI). Transfected cells were cultured for 48 hours; plasmid DNA was recovered by Hirt extraction.

Recombination levels were determined by nested qPCR. First round amplification used Taq DNA Polymerase with primers DR55 and 1233 (1+2; Table S1). Thermocycling comprised 19 cycles of 15 seconds at 95℃, 15 seconds at 55℃ and 30 seconds at 68℃. qPCR reactions (10 µl) contained Luna Universal Probe qPCR Master Mix, (NEB, Ipswich, MA, USA), 1 µl first round PCR product, 4 pmol SJ-F and SJ-R primers (3+4; Table S1), 10 pmol of JH299 SJ probe (7; Table S1) and were performed in a RotorGene 6000 cycler. Recombination was normalised to the total amount of recovered recombination substrate using primers CAT-F and CAT-R (5+6; Table S1). qPCR reactions contained SYBR green master mix (Sensifast No ROX BIO-98005; Bioline, London, UK), 1 µl of 1:100 dilution of first round PCR product and 4 pmol of each primer. Analysis was with RotorGene-Q software, v 2.3.1.

Western blotting

NIH3T3 cells (1 x 10^6) were transfected with 0.8 µg WT or mutant pCS2MT-RAG1, 1.6 µg pEFXC-RAG2 and 500 ng pTkβ-galactosidase (Clontech). After 48 hours, extracts were made and western blotting performed as described[20].

Incubations were as follows: 1:2000 dilution of anti-c-Myc antibody (9E10; Abcam, Cambridge, UK), 1:10000 dilution of anti-goat antibody (5210-0170; Sera Care, MA, USA) and 1:20000 dilution of HRP conjugated anti-rabbit antibody (A120-101P, Bethyl, TX, USA). Blots were stripped and incubated with 1:2000 dilution of anti-β-galactosidase antibody (Z3781; Promega, Madison, WI, USA) and 1:20000 dilution of HRP conjugated anti-mouse antibody (715-035-151, Jackson ImmunoResearch, Cambridgeshire, UK). Signal detection was via SuperSignal™ WestPico Chemiluminescent Substrate (34080; ThermoFisher, Waltham MA, USA) and a Sysgene G:Box ChemiXRQ Imaging System.

Purification of RAG and HMGB1 proteins

MBP-tagged wildtype or mutant core RAG1 and core RAG2 were purified from HEK293T cells as described[21].

The full-length rat HMGB1 expression plasmid, pETM11-HMGB1, has a N-terminal 6x His tag, followed by a Tobacco Etch Virus (TEV) protease cleavage site. The 6xHis-HMGB1 fusion protein was expressed in *E. coli* and purified by affinity chromatography using Ni-NTA, followed by removal of the His tag (except Figure 4A) by TEV cleavage and purification via ion exchange chromatography using a HiTrap™ Q HP column.

In vitro RAG cleavage and binding assays

RAG cleavage assays were performed using fluorescently labelled RSS probes (Supplemental methods) whilst binding assays were performed using radioactively labelled oligonucleotides to improve sensitivity. Oligonucleotide sequences are given in Table S1.

Generation of HMGB1-knock-down cells

CRISPRi was used to knock-down HMGB1 expression using a catalytically dead Cas9 nuclease (dCas9) fused to a KRAB transcriptional repressor domain[22]. Detailed descriptions of the plasmids and production of lentiviral particles are in Supplemental methods.

NIH3T3 cells were seeded at 1 x 10^5 cells per well of a 6-well plate. After 24 hours, the media was replaced with 1 ml of each lentiviral supernatant containing 6 μg/ml polybrene (TR-1003-G, Merck Kenilworth, NJ, USA). Plates were centrifuged at 800 *g* for 30 minutes at 32℃. 24 hours post-transduction the media was replaced with media containing 10 μg/ml blasticidin S HCl (sc-495389, Santa Cruz Biotechnology Heidelberg, Germany) and 400 μg/ml zeocin (J67140.XF, Alfa Aesar, Haverhill, MA, USA). HMGB1 knockdown was confirmed by western blotting using a mouse monoclonal anti-HMGB1 antibody (sc-56698, Santa Cruz Biotechnology) and a mouse monoclonal anti-β-Tubulin antibody as a loading control, both at a 1:1000 (TA347064; Origene, Rockville, MD, USA). Antibiotics were withdrawn 24 hours prior to transfection for the recombination assay to minimise their effects.

Expression of HMGB1/2

HMGB1 and HMGB2 were re-introduced using lentiviruses encoding full length HMGB2 and the truncated proteins, HMGB1ΔC and HMGB2ΔC, fused to puromycin *N*-acetyltransferase separated by a self-cleaving P2A peptide sequence. Both HMGB1ΔC and HMGB2ΔC (aa 1-172) lack the C-terminal acidic tail, express at much higher levels than their full-length counterparts and HMGB1ΔC has higher DNA and RAG binding activity[23]. Detailed descriptions of plasmids and lentiviral transductions are in Supplemental methods.

Data sharing statement

For original data, please contact j.m.boyes@leeds.ac.uk. One Table is in the supplemental text, available with the online version of this article.

Results

A detailed clinical description of the patient is provided in a recent publication[24] (patient 14). Briefly, the patient presented at age 34 with established bronchiectasis, recurrent chest infections, normal immunoglobulin levels and mild T cell lymphopenia. Following genetic testing using a gene panel based approach, he was found to have biallelic mutations in RAG1: c.1210C>T p.Arg404Trp, and c.1520G>A p.Arg507Gln.

To dissect how these RAG1 mutations (Figure 1A) impact on V(D)J recombination, recombination levels were assayed by transfecting a reporter plasmid and RAG expression vectors into non-lymphoid cells. In this reporter plasmid, the 12- and 23-RSSs are initially in the same orientation; recombination causes the intervening sequence to be inverted, allowing quantification of its levels using PCR primers that become convergent only after recombination (Figure 1B).

Having confirmed that recombination increases quantitatively according to RAG1 activity (Figure 1C), we introduced individual mutations into mouse RAG1 cDNA to alter the equivalent amino acids (R401W and R504Q) to those in the patient. Expression vectors for these RAG1 proteins were then transfected into NIH3T3 cells, together with the recombination reporter plasmid and a RAG2 expression vector. Whilst mutation R504Q gave detectable levels of recombination, equivalent to ~2-5% of wild type (WT; Figure 1D & 1E), recombination with R401W was negligible.

Effects of mutation R504Q on RAG activity *in vitro*

To investigate the molecular basis of the reduced recombination, RAG binding and cutting assays were carried out *in vitro*. Core RAG proteins (amino acids 384-1008 of RAG1 and 1-382 of RAG2) are fully active in binding and cutting but more amenable to purification than full length RAG proteins[21]. Therefore, to enable our data to be compared with available high resolution structures, the relevant mutations were introduced into mouse core RAG1, (which is 96.3% identical and 97.6% similar to human core RAG1) and was expressed

together with core RAG2 in 293T cells. These were then purified and used in RAG cutting assays where an end labelled oligonucleotide carrying a 12- or 23-RSS was incubated with RAG proteins and full length HMGB1, with or without the relevant 12- or 23- partner RSS. With R504Q, cutting is reduced by ~15-fold compared to WT RAG1 at both 12- and 23-RSSs (Figure 2A), which mirrors the reduction in RSS binding (Figure 2B) and the observed decrease in recombination *in vivo* (Figure 1E).

Mutation R401W significantly alters RAG activity

With R401W, RAG cutting is almost completely eliminated at both types of RSS (Figure 2A). Moreover, DNA binding was significantly altered. RAGs form two complexes with DNA, a faster migrating complex, SC1, which consists of two molecules of RAG1 and one of RAG2 and SC2 that contains two molecules of each protein. In the presence of HMGB1, these complexes are shifted to slower mobility complexes, HSC1 and HSC2[25] where HSC2 is thought to be active in RAG cleavage. At a 12-RSS, although RSS binding by R401W appears normal, forming SC1/SC2 complexes like WT RAG1 (Figure 2D), in the presence of HMGB1, HSC1 is substantially reduced and HSC2 is reduced by more than 50% (Figure 2B). Indeed, HSC1 is almost completely replaced by a faster migrating complex that has the same mobility as SC1. This implies that HMGB1 does not bind efficiently to the R401W/12-RSS complex. With a 23-RSS, R401W forms similar complexes to WT RAG1, although HSC1 and HSC2 have slightly faster mobilities (which are more fully resolved in Figure 4D), suggesting altered conformations.

To better understand the molecular basis of this altered binding, the recent cryo-electron microscopy (cryo-EM) and X-ray crystal structures of the mouse and zebrafish RAG/RSS complexes[12,13] were examined. Two areas of HMGB1 binding, representing Box A and Box B, were mapped to the 23-RSS spacer and correspond to 60° and 90° DNA bends[13]. Although unresolved, a single region of HMGB1 density, likely Box A, was observed in the 12-RSS spacer, also at a 60° bend. Notably, the location of Box A at both RSSs is directly opposite the region occupied by R401 (Figure 3). It is likely that the

nonamer binding domain of RAG1 (where R401 is located) establishes the 60° bend [8,26] since HMGB1 is recruited after RAG[27] and Box A recognises pre-bent DNA[28]. Indeed, this bend may be a requirement for Box A binding and it is possible that the bulky tryptophan in R401W disrupts this bending to reduce HMGB1 association.

At the 23-RSS, Box B lies in the middle of the 23-RSS spacer, where it stabilises a 90° bend. With this arrangement, it is possible Box B stabilises HMGB1 binding to the RAG complex whilst the R401W mutation again lies in the region that is bound by Box A (Figure 3, right) and where the bulky tryptophan could again destabilise Box A binding. Consistent with this, HSC2 and HSC1 formed with R401W at the 23-RSS have slightly faster mobilities (Figure 2B). Moreover, since bending of the 12- and 23-RSSs by 60° and 150°, respectively, is required to position the RSSs for efficient cleavage, the altered binding can explain R401W's drastically reduced activity (Figure 2A).

R401W decreases HMGB1 binding

To test if the altered RAG/RSS complexes indeed result from changes in HMGB1 binding, His-tagged HMGB1 was employed in RAG binding assays. This forms very similar complexes to those with untagged HMGB1 (Figure 4A); importantly, addition of an anti-His antibody super-shifts the complexes formed with HMGB1 whilst the faster migrating complex formed with R401W, is unchanged (Figure 4A). This therefore supports the idea that HMGB1 binds significantly less well to RAG/DNA complexes formed with R401W at 12-RSSs. These data further suggest that the addition of a second molecule of RAG2 to form HSC2 may stabilise HMGB1 binding but since recombination is negligible, this HSC2-like complex appears to be non-functional.

We next performed RAG cutting assays in the presence of increasing amounts of HMGB1 to test if the weak HMGB1 binding is overcome at higher amounts of HMGB1 and if this can restore activity of R401W. Whilst cutting by WT RAG1 gradually increases as more HMGB1 is added, the activity of R401W was negligible, irrespective of the amount of HMGB1 (Figure 4B).

HMGB1 is required for RAG activity under our reaction conditions (Figure 4B), but it is also possible that the inactivity of R401W is independent of reduced HMGB1 binding. To test this idea, RAG cutting assays were performed in the presence of manganese, which enables cutting of single RSSs[29]. This contrasts to standard RAG cutting conditions where the divalent cation is magnesium and cutting requires synaptic complex formation between complementary 12- and 23-RSSs. Using these less stringent manganese conditions, cutting by R401W is readily observed at similar levels to WT protein (Figure 4C). Thus, the R401W mutant protein has catalytic activity but this is not manifested in the absence of stable HMGB1 binding. This is also consistent with recent structural studies, outlined above, that show that HMGB1 is required to position RSSs correctly in the synaptic complex for RAG cutting[13].

We next investigated how R401W reduces HMGB1 binding and reasoned that this could be due to two non-mutually exclusive effects: Firstly, the bulky tryptophan of R401W could sterically hinder HMGB1 binding and/or DNA bending within the RAG/DNA complex. Alternatively, the loss of arginine's positive charge could disrupt RAG1 interactions with phosphate groups in the DNA spacer to reduce HMGB1 binding to the RAG/DNA complex.

To differentiate these possibilities, R401 was mutated to leucine, which has a similar length to arginine, but which lacks a positive charge. This new mutation, R401L, partially restores binding to generate HSC1 with lower mobility and increased amounts of HSC2 (Figure 4D); however, no cutting is observed under standard conditions (Figure 4E). Moreover, (H)SC1 is more smeared at a 12-RSS and HSC2 levels are below those of WT RAG1.

To test if a positive charge is required, R401 was substituted with lysine. This resulted in almost WT levels of cutting at both 12- and 23-RSSs (Figure 4E). Moreover, R401K formed very similar amounts of HSC1 and HSC2 to WT RAG1 at a 23-RSS. At a 12-RSS, however, HSC1 formed with R401K is more smeared than with WT RAG1 (Figure 4D), suggesting that the longer arginine

side chain and/or its guanidinium group, that can form contacts in three directions compared to just one for lysine[30], might be required for RSS bending and stable HMGB1 binding.

To assess the effects of these new mutations on V(D)J recombination *in vivo,* we employed the transfection-based recombination assay. Whilst R401L catalyses only negligible levels of recombination, consistent with lower/less stable formation of HSC1 and HSC2 complexes *in vitro*, R401K restores recombination to almost WT levels, again correlating with *in vitro* cutting and formation of HSC1 and HSC2 (Figure 4F). Together, these data strongly support the idea that R401W disrupts association of HMGB1 with the RAG/RSS complex and highlight the critical role of HMGB1 binding for RAG cutting and recombination *in vivo*.

Knock-down of HMGB1 significantly decreases recombination

If HMGB1 is indeed essential for V(D)J recombination *in vivo*, we reasoned that in the absence of HMGB1, recombination will be negligible. We further predicted that if HMGB1 is restored, V(D)J recombination will be re-established in the presence of WT RAG1 but not R401W. To test these ideas, we employed a fusion protein between catalytically dead Cas9 (dCas9) and the KRAB repressor (dCas9-KRAB)[22]. This was targeted to the start site of the HMGB1 promoter using a CRISPR guide RNA (Figure 5A), resulting in knock-down of HMGB1 by up to 90% (Figure 5B). V(D)J recombination is also reduced by an average of 80% with WT RAG1 whereas recombination with R401W remains negligible (Figure 5C), exactly as predicted by our model. Upon transduction of these cells with a lentivirus that expresses HMGB1 under the control of the EF1α promoter, HMGB1 levels are restored to 250-400% of WT cells. Also as predicted, V(D)J recombination activity is rescued with WT RAG1 and R401K but remains negligible with R401W (Figure 5C). These studies therefore strongly suggest that HMGB1 is required for V(D)J recombination *in vivo*.

Paradoxically, HMGB1 knock-out mice develop a normal immune system[17]. Since HMGB2, which is 79% identical, substitutes for HMGB1 in RAG cutting assays *in vitro*[8], it seemed plausible that HMGB2 substitutes for HMGB1 *in vivo*. To test if this is the case, we expressed HMGB2 in our HMBG1 knock-down cells and find indeed that HMGB2 restores recombination activity with wild type RAG1 but not R401W (Figure 5C). Together, these studies provide compelling evidence that at least one of the highly related bending proteins, HMGB1/2, is required for V(D)J recombination *in vivo*.

Discussion

Previous studies of RAG mutations highlighted the amino acids that are important for RAG function *per se*, for example, that mediate RAG/DNA or RAG/RAG contacts[2,31]. Here, a RAG1 mutation instead demonstrates the critical role of an accessory protein during V(D)J recombination. Indeed, a long-standing question is whether HMGB1/2 is required for V(D)J recombination *in vivo*. Whilst there is very good evidence that HMGB1 stimulates RAG cutting *in vitro*[8] and recent structural studies explained the mechanistic basis, relatively little evidence existed that HMGB1/2 plays this role *in vivo,* primarily due to inconclusive mouse knock-out studies[17]. Here, we show that the RAG1 mutation, R401W, which places a bulky tryptophan group exactly at the position of a critical 60° bend at both 12- and 23-RSSs, dramatically alters subsequent HMGB1 binding *in vitro* and almost eliminates V(D)J recombination *in vivo*. Moreover, when this tryptophan is replaced with lysine, HMGB1 binding and RAG cutting are restored to close to WT levels *in vitro*, correlating with an equivalent increase in recombination *in vivo*. These data, together with the significant decrease in recombination upon knock-down of HMGB1, provide strong evidence that HMGB1 is required for V(D)J recombination *in vivo*. Moreover, by showing that HMGB2 can substitute for HMGB1 *in vivo*, our data solve a long-standing discrepancy by demonstrating that either of the DNA bending proteins, HMGB1 or HMGB2, is required for V(D)J recombination.

By analysing point mutations in each RAG1 allele individually from a patient with primary immunodeficiency, the relative contribution of each mutation to V(D)J recombination could be dissected. R401W has negligible RAG cutting and recombination activity. Whilst it binds DNA, and forms the expected complexes with RAG2, it does not form the expected complexes with HMGB1. This is particularly apparent at a 12-RSS where HSC1 is substantially reduced and HSC2 is reduced by >50%; at a 23-RSS, HSC1 and HSC2 are formed, albeit with increased mobility. Importantly, changing R401 to leucine and then lysine causes HMGB1 binding to become progressively more like the WT pattern. These changes strongly correlate with changes in RAG activity both

in vitro and *in vivo*. Previous studies showed that HMGB1 preferentially binds to a pre-formed RAG1/RSS complex where RAG1 causes partial RSS bending[32]. The observed alterations in HMGB1 binding to the RAG1/RSS complex could be explained since R401 binds a phosphate group in the DNA backbone directly opposite the region that is bound in the minor groove by the HMG boxes[13]. R401 thus stabilises a bend at the nonamer/RSS spacer junction at both RSSs and consequently, when R401 is mutated to tryptophan, the bulky side chain likely disrupts this R401-phosphate interaction, to destabilise HMGB1 Box A binding at both RSSs (Figure 6). Although a second region of HMGB1 binding (Box B) maps to the middle of the 23-RSS spacer to maintain HMGB1 association, binding to just this single site would not induce bending[13], correlating with reduced RAG cutting.

From Figures 2B and 4D, it is apparent that R401W binds to RSSs less efficiently than WT RAG1 in the presence of HMGB1. Recent studies of human RAG1 mutations suggested that this may be due to this arginine (R404 in human RAG1) stacking with R443 in the other RAG1 monomer, to alter the conformation of the nonamer binding domain and consequently RAG1/RSS binding[33]. Such interactions however, do not necessarily exclude the effects described above on HMGB1 binding when this arginine is replaced by tryptophan.

The second RAG1 mutation, R504Q, retains residual recombination activity although RSS binding and RAG cutting are reduced (Figure 2A and 2B). This is consistent with structural studies that show R504 contacts the first two bases within the spacer after the heptamer[12] and it is possible that the loss of positive charge in R504Q destabilises RAG binding. Notably, when we examined the effect of the RAG1 mutations on the patient's TCRβ repertoire, Vβ14 and Vβ12 are over-represented whilst Vβ2 and Vβ16 are under-represented (data not shown). It is possible that weaker RAG binding to less conserved RSSs is exaggerated with R504Q.

Although the R504Q mutation reduces recombination activity, the patient nonetheless has a partially functional immune system, consistent with their

late presentation. Notably, when RAG1 proteins with the R401W and R504Q mutations are mixed, intermediate levels of RAG cutting and recombination are observed (Figures 1E & 2A). This is possibly because R401W inhibits the formation of (partially) active RAG complexes by R504Q but complexes with two R504Q proteins can still form, giving residual levels of recombination and the mild phenotype of the patient.

Reduced binding of HMGB1 to RAG/RSS complexes formed with the RAG1 R401W mutation strongly correlates with negligible activity in RAG cutting assays *in vitro* and undetectable recombination *in vivo*. This, coupled with the fact that knock-down of HMGB1 greatly reduces V(D)J recombination by WT RAG1, strongly implies that HMGB1 plays an essential role in V(D)J recombination *in vivo*.

Acknowledgements

We thank Professor Marco Bianchi (San Raffaele University, Milan, Italy) for his kind gift of the HMGB1 expression plasmid, Richard Bayliss for helpful comments on the manuscript and Alastair Smith for help with the statistics. This work is supported by a University of Leeds Research Scholarship to DTT and Bloodwise grant 15042 to JMB. SS is supported by the National Institute for Health Research (NIHR) Leeds Biomedical Research Centre. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Authorship contributions

DT, SS and JB conceived the project, designed the experiments and wrote the paper. DT, CC and DL performed the experiments.

Conflict of Interest Statement

The authors declare no competing interests.

References

1. Gellert M. V(D)J recombination: RAG proteins, repair factors, and regulation. *Annu. Rev. Biochem.* 2002;71:101–132.
2. Notarangelo LD, Kim M-S, Walter JE, Lee YN. Human RAG mutations: biochemistry and clinical implications. *Nature Reviews Immunology* 2016;16:234-46.
3. Schatz DG, Swanson PC. V(D)J Recombination: Mechanisms of Initiation. *Annu. Rev. Genet.* 2011;45(1):167–202.
4. Tonegawa S. Somatic generation of antibody diversity. *Nature*. 1983;302(5909):575–581.
5. Yin FF, Bailey S, Innis CA, et al. Structure of the RAG1 nonamer binding domain with DNA reveals a dimer that mediates DNA synapsis. *Nat Struct Mol Biol*. 2009;16(5):499–508.
6. Helmink BA, Sleckman BP. The Response to and Repair of RAG-Mediated DNA Double-Strand Breaks. *Annu. Rev. Immunol.* 2012;30(1):175–202.
7. McBlane JF, van Gent DC, Ramsden DA, et al. Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. *Cell*. 1995;83(3):387–395.
8. van Gent DC, Hiom K, Paull TT, Gellert M. Stimulation of V(D)J cleavage by high mobility group proteins. *The EMBO journal*. 1997;16(10):2665–2670.
9. Lovely GA, Brewster RC, Schatz DG, Baltimore D, Phillips R. Single-molecule analysis of RAG-mediated V(D)J DNA cleavage. *Proc Natl Acad Sci USA*. 2015;112(14):E1715–E1723.
10. Malarkey CS, Churchill MEA. The high mobility group box: the ultimate utility player of a cell. *Trends in biochemical sciences*. 2012;37(12):553–562.
11. Ciubotaru M, Trexler AJ, Spiridon LN, et al. RAG and HMGB1 create a large bend in the 23RSS in the V(D)J recombination synaptic complexes. *Nucleic acids research*. 2013;41(4):2437–2454.
12. Ru H, Chambers MG, Fu T-M, et al. Molecular Mechanism of V(D)J Recombination from Synaptic RAG1-RAG2 Complex Structures. *Cell*. 2015;163:1138-1152.
13. Kim M-S, Chuenchor W, Chen X, et al. Cracking the DNA Code for V(D)J Recombination. *Molecular Cell*. 2018;70(2):358–370.
14. Zagelbaum J, Shimazaki N, Esguerra ZA, et al. Real-time analysis of RAG complex activity in V(D)J recombination. *Proc Natl Acad Sci USA*. 2016;113(42):11853–11858.
15. Kwon J, Imbalzano AN, Matthews A, Oettinger MA. Accessibility of nucleosomal DNA to V(D)J cleavage is modulated by RSS positioning and HMG1. *Mol. Cell*. 1998;2(6):829–839.
16. Aidinis V, Bonaldi T, Beltrame M, et al. The RAG1 homeodomain recruits HMG1 and HMG2 to facilitate recombination signal sequence binding and to enhance the intrinsic DNA-bending activity of RAG1-RAG2. *Mol Cell Biol*. 1999;19(10):6532–6542.
17. Calogero S, Grassi F, Aguzzi A, et al. The lack of chromosomal protein Hmg1 does not disrupt cell growth but causes lethal hypoglycaemia in newborn mice. *Nat Genet*. 1999;22(3):276–280.

18. Mizushima S, Nagata S. pEF-BOS, a powerful mammalian expression vector. *Nucleic acids research*. 1990;18(17):5322.
19. Hesse JE, Lieber MR, Mizuuchi K, Gellert M. V(D)J recombination: a functional definition of the joining signals. *Genes & Development*. 1989;3(7):1053–1061.
20. Bevington S, Boyes J. Transcription-coupled eviction of histones H2A/H2B governs V(D)J recombination. *The EMBO journal*. 2013;32(10):1381–1392.
21. Bergeron S, Anderson DK, Swanson PC. RAG and HMGB1 Proteins: Purification and Biochemical Analysis of Recombination Signal Complexes. *DNA Repair, Part A*. 2006;408:511–528.
22. Gilbert LA, Larson MH, Morsut L, et al. CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell*. 2013;154(2):442–51.
23. Bergeron S, Madathiparambil T, Swanson PC. Both high mobility group (HMG)-boxes and the acidic tail of HMGB1 regulate recombination-activating gene (RAG)-mediated recombination signal synapsis and cleavage in vitro. *Journal of Biological Chemistry*. 2005;280(35):31314–31324.
24. Lawless D, Geier CB, Farmer JR, et al. Prevalence and clinical challenges among adults with primary immunodeficiency and recombination-activating gene deficiency. *J. Allergy Clin. Immunol.* 2018;141(6):2303–2306.
25. Swanson, P. C. Fine structure and activity of discrete RAG-HMG complexes on V(D)J recombination signals. *Molecular and Cellular Biology*. 2002;22(5):1340–51.
26. Grundy GJ, Ramon-Maiques S, Dimitriadis EK, et al. Initial Stages of V(D)J Recombination: The Organization of RAG1/2 and RSS DNA in the Postcleavage Complex. *Molecular Cell*. 2009;35(2):217–227.
27. Tsai CL, Drejer AH, Schatz DG. Evidence of a critical architectural function for the RAG proteins in end processing, protection, and joining in V(D)J recombination. *Genes & Development*. 2002;16(15):1934–1949.
28. Webb M, Payet D, Lee KB, Travers AA, Thomas JO. Structural requirements for cooperative binding of HMG1 to DNA minicircles. *Journal of molecular biology.* 2001; 309:79-88.
29. van Gent DC, Ramsden DA, Gellert M. The RAG1 and RAG2 proteins establish the 12/23 rule in V(D)J recombination. *Cell*. 1996;85(1):107–13.
30. Sokalingam S, Raghunathan G, Soundrarajan N. A study on the effect of surface lysine to arginine mutagenesis on protein stability and structure using green fluorescent protein. *PLoS ONE* 2012;7:e40410.
31. Matthews AG, Kuo AJ, Ramon-Maiques S, et al. RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. *Nature*. 2007;450(7172):1106–1110.
32. Little AJ, Corbett E, Ortega F, Schatz DG. Cooperative recruitment of HMGB1 during V(D)J recombination through interactions with RAG1 and DNA. *Nucleic acids research*. 2013;41:3289-301.

33. Lee YN, Frugoni F, Dobbs K, et al. Characterization of T and B cell repertoire diversity in patients with RAG deficiency. *Science Immunology*. 2016;1(6):eaah6109–eaah6109.
34. Steen SB, Gomelsky L, Speidel SL, Roth DB. Initiation of V(D)J recombination in vivo: role of recombination signal sequences in formation of single and paired double-strand breaks. *The EMBO journal*. 1997;16(10):2656–2664.
35. Taccioli GE, Rathbun G, Oltz E, et al. Impairment of V(D)J recombination in double-strand break repair mutants. *Science*. 1993;260(5105):207–210.
36. Sekiguchi JA, Whitlow S, Alt FW. Increased accumulation of hybrid V(D)J joins in cells expressing truncated versus full-length RAGs. *Mol. Cell*. 2001;8(6):1383–90.

Figure legends

Figure 1: RAG1 mutations affect V(D)J recombination *in vivo* (A) Schematic of RAG1 showing the main core RAG1 sub-domains (NBD: nonamer binding domain; DDBD: dimerisation and DNA binding domain; preR: pre-RNaseH domain; ZBD: zinc binding domain; CTD: carboxy-terminal domain). The positions of the mutations in the patient are shown with the equivalent mouse amino acids in brackets beneath (B) Plasmid pJH299 used to quantify V(D)J recombination *in vivo*. 12- and 23-RSSs are indicated as triangles. Following inversional recombination, primers 1+2 and 3+4 are used in a qPCR nested assay, with a hydrolysis probe across the RSS junction. Total plasmid amounts are measured using primers 5 and 6. (C) Increasing amounts of WT RAG1 expression plasmid were transfected into NIH3T3 cells and the levels of recombination determined. A direct correlation with the amount of RAG1 activity is observed. *N*=3; error bars show SEM. (D) Western blot showing expression levels of the various RAG1 proteins in transfected NIH3T3 cells; a β-galactosidase expression vector was co-transfected as a loading control. (E) Recombination levels using WT and mutated RAG1 proteins, relative to WT RAG1. Recombination was normalised according to the total amount of recombination plasmid recovered; similar levels of RAG1 protein expression were verified by western blotting (Figure 1D). *N*=3; error bars show SEM.

Figure 2: RAG1 mutations affect cutting and binding *in vitro*

(A) Cutting at a 12-RSS (left) or 23-RSS (right) is shown. An oligonucleotide carrying a consensus 12- or 23-RSS, as indicated, was incubated with equivalent amounts of core RAG2 (cRAG2) plus the core RAG1 (cRAG1) proteins shown (Figure 2C) in the presence of HMGB1, with a 10-fold excess of unlabelled 23- or 12- partner RSS. For cutting reactions on the far right of each gel, equal amounts of R401W and R504Q were mixed to give the same total amount of RAG1 as the other lanes. The per cent cutting is indicated beneath the gels; an asterisk indicates the labelled oligonucleotide. (B)

Binding to a labelled 12-RSS (left) and 23-RSS (right). Binding reactions were performed with cRAG2 and the cRAG1 proteins indicated. Complexes SC1 and SC2 as well as HSC1 and HSC2 are shown. (C) The levels of purified RAG proteins are equivalent. Purified, MBP-tagged cRAG1 and cRAG2 proteins were separated by electrophoresis and the gel stained with Coomassie blue. The various RAG1 proteins are indicated above each lane; wt = wild type. (D) Comparison of R401W and WT RAG1 binding to a 12-RSS. RAG binding reactions were performed in the presence or absence of HMGB1 as indicated. Whilst R401W forms equivalent complexes to WT with the 12-RSS in the absence of HMGB1, R401W does not form HSC1.

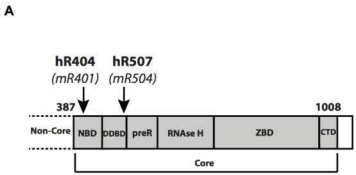
Figure 3: HMGB1 binding to the RAG/DNA complex. HMGB1 binding was determined from the published cryo-EM and X-ray crystal structures of mouse RAG/RSS complexes[12,13]. Blue and green cartoons represent individual RAG1 monomers. Left: HMGB1 Box A from the 23-RSS structure was superimposed over the density that likely corresponds to HMGB1 Box A at the 12-RSS, denoted by the asterisk[13]. This falls directly opposite R401 of mouse RAG1, indicated in red. R401 contacts DNA in the 12-RSS spacer, close to the RSS nonamer. Right: HMGB1 Box B binding is observed in the middle of the 23-RSS spacer whilst an additional area of HMGB1 binding (Box A) lies opposite R401 (in red) at the 23-RSS. Data from PDB 6CG0[13].

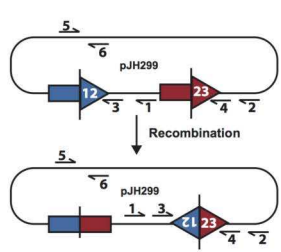
Figure 4: Amino acid substitutions progressively restore HMGB1 binding and RAG1 activity (A) HMGB1 binds less well to R401W/12-RSS complexes. WT or R401W mutant RAG1 complexes were formed with a labelled 12-RSS in the presence or absence of His-tagged HMGB1. A 23-RSS oligonucleotide partner was present in all cases. Supershifted complexes formed upon addition of an anti-His tag antibody, are indicated. The mobilities of SC1, SC2, HSC1 and HSC2 are shown. (B) Increasing amounts of HMGB1 do not restore cutting with R401W. RAG cutting reactions were performed with WT RAG1 or R401W in the presence of increasing amounts of HMGB1 (0, 25, 50, 100, 200, 400, 800 nmoles). A 23-RSS substrate was used as HMGB1 has the greatest effect at this type of RSS. It is notable that in the absence of HMGB1, cutting by WT RAG1 is not detectable (lane 2). (C)

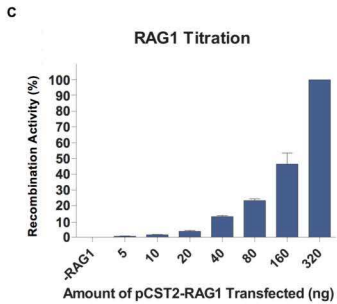
R401W has catalytic activity. RAG cutting reactions were performed using a single labelled 12-RSS in the presence of manganese. (D) HSC1 and HSC2 complexes become increasingly like WT RAG1 complexes with R401L and R401K. RAG binding to a labelled 12- or 23-RSS in the presence of an unlabelled partner RSS and HMGB1, as indicated. The gel was run for longer than Figure 2B to better separate the complexes; a cropped image is shown to highlight the mobility differences. (E) RAG cutting with R401W, R401L and R401K. RAG cutting was performed with a labelled 12-RSS or 23-RSS, indicated by the asterisk, in the presence of unlabelled partner and the RAG1 proteins indicated, with or without HMGB1. The per cent cutting is given beneath the gel. (F) R401K recombination activity is close to WT RAG1 levels *in vivo*. Recombination was monitored using pJH299 and equivalent levels of RAG1 proteins (Figures 1B & 1D). Recombination levels are shown relative to WT RAG1. *N*=3; error bars show SEM.

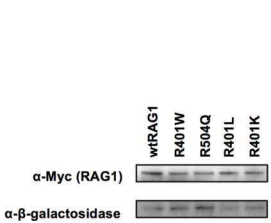
Figure 5: Knock-down of HMGB1 significantly reduces V(D)J recombination *in vivo.* A) Schematic of the steps used in CRISPR/Cas9-mediated knock-down of HMGB1. Transduction with lentiviruses and antibiotic selection steps are shown. B) Representative western blots showing the levels of HMGB1 and full length (fl) and C-terminal deletion (ΔC) HMGB2 relative to a β-tubulin control. Antibodies that specifically recognise HMGB2ΔC are unavailable as the epitope lies at the C-terminus. However, the core regions of HMGB1 and HMGB2 are 85.5% identical and the anti-HMGB1 antibody cross reacts with HMGB2 (indicated by the asterisk). C) V(D)J recombination levels with the RAG1 proteins shown in WT NIH3T3 cells (blue bars), when the KRAB repressor as introduced in the absence of guide RNAs (purple bars) when HMGB1 has been knocked down (red bars) or when HMGB1 (dark green bars) or HMGB2 full length (light green bars) or HMGB2ΔC (turquoise bars) have been reintroduced *N* = 3; error bars show SEM. The changes in recombination activity are statistically significant (Student's t-test; ns = non-specific, ** p = 0.01 and *** p = 0.001). The endogenous HMGB2 in NIH3T3 cells could account for residual recombination activity following HMGB1 knock-down.

Figure 6: Schematic showing altered HMG binding to wild type RAG1 and R401W. The DNA strands are shown with the heptamer, spacer and nonamer indicated. A and B represent the likely positions of HMG Box A and B, respectively and R1 and R2 represent RAG1 and RAG2. The bend angles within each RSS are shown.


A
hR404
(mR401)
hR507
(mR504)
387
1008
Non-Core
NBD
DDBD
preR
RNAse H
ZBD
CTD
Core


B
5
6
pJH299
12
23
3
1
4
2
Recombination
5
6
pJH299
1
3
12
23
4
2


C
RAG1 Titration
Recombination Activity (%)
100
90
80
70
60
50
40
30
20
10
0
-RAG1
5
10
20
40
80
160
320
Amount of pCST2-RAG1 Transfected (ng)


D
wtRAG1
R401W
R504Q
R401L
R401K
α-Myc (RAG1)
α-β-galactosidase

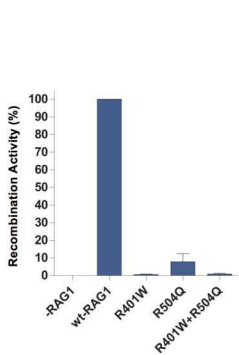

E
Recombination Activity (%)
100
90
80
70
60
50
40
30
20
10
0
-RAG1
wt-RAG1
R401W
R504Q
R401W+R504Q

Figure 1

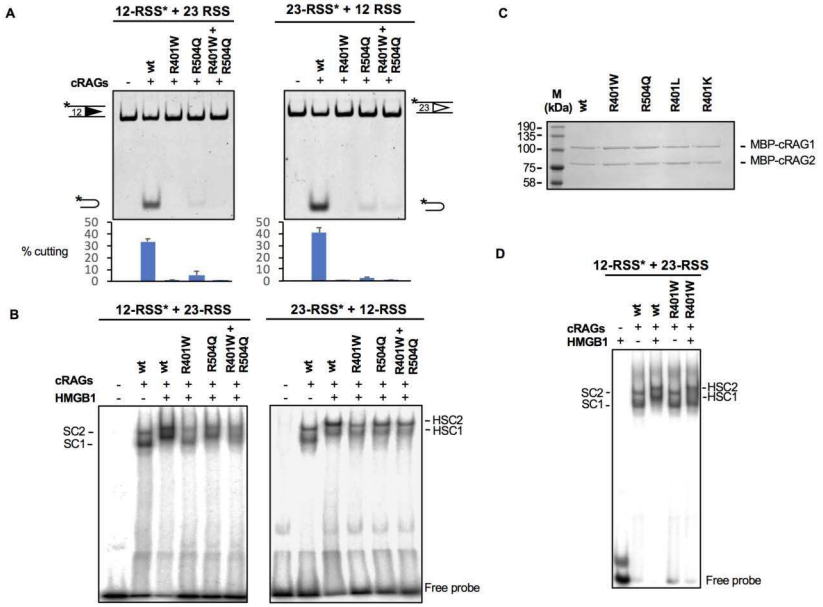

A
12-RSS* + 23 RSS
23-RSS* + 12 RSS
cRAGs
wt
R401W
R504Q
R401W + R504Q
% cutting
B
12-RSS* + 23-RSS
23-RSS* + 12-RSS
HMGB1
SC2
SC1
HSC2
HSC1
Free probe
C
M (kDa)
R401L
R401K
MBP-cRAG1
MBP-cRAG2
D

Figure 2

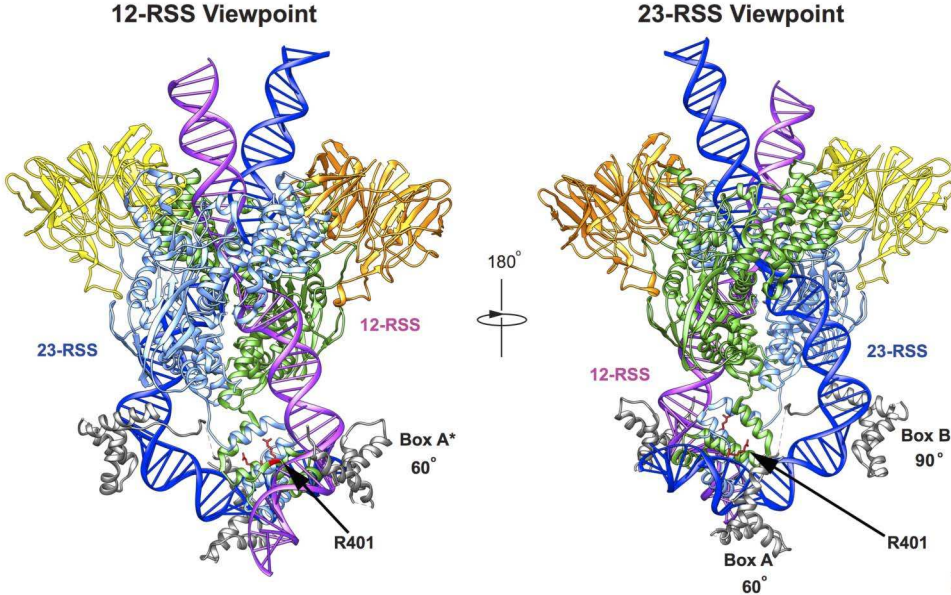

12-RSS Viewpoint
23-RSS Viewpoint
180°
23-RSS
12-RSS
Box A*
60°
R401
12-RSS
23-RSS
Box B
90°
R401
Box A
60°

Figure 3

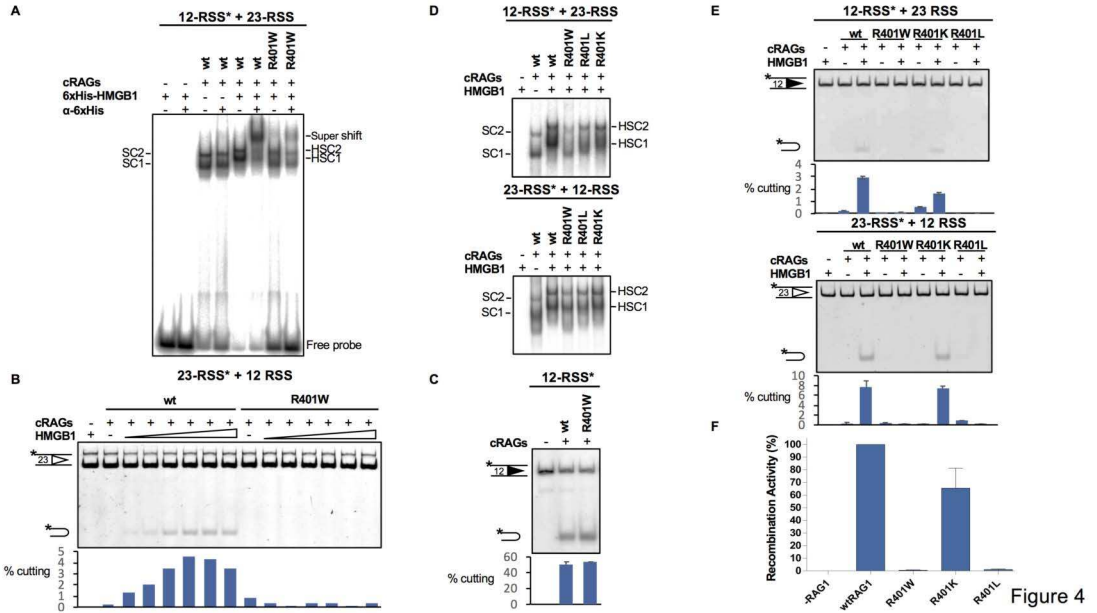

A
12-RSS* + 23-RSS
wt wt wt R401W R401W
cRAGs - - + + + + +
6xHis-HMGB1 + + - - + + +
α-6xHis - + - + + - +
Super shift
HSC2
HSC1
SC2
SC1
Free probe
B
23-RSS* + 12 RSS
wt
R401W
cRAGs - + + + + + + + + + + +
HMGB1 + - - - - - - - - - - -
23
% cutting
C
12-RSS*
wt R401W
cRAGs - + +
12
% cutting
D
12-RSS* + 23-RSS
wt wt R401W R401L R401K
cRAGs - + + + + +
HMGB1 + - + + + +
SC2
SC1
HSC2
HSC1
23-RSS* + 12-RSS
wt wt R401W R401L R401K
cRAGs - + + + + +
HMGB1 + - + + + +
SC2
SC1
HSC2
HSC1
E
12-RSS* + 23 RSS
wt R401W R401K R401L
cRAGs - + + + + + + + +
HMGB1 + - + - + - + - +
12
% cutting
23-RSS* + 12 RSS
wt R401W R401K R401L
cRAGs - + + + + + + + +
HMGB1 + - + - + - + - +
23
% cutting
F
Recombination Activity (%)
-RAG1
wtRAG1
R401W
R401K
R401L

Figure 4

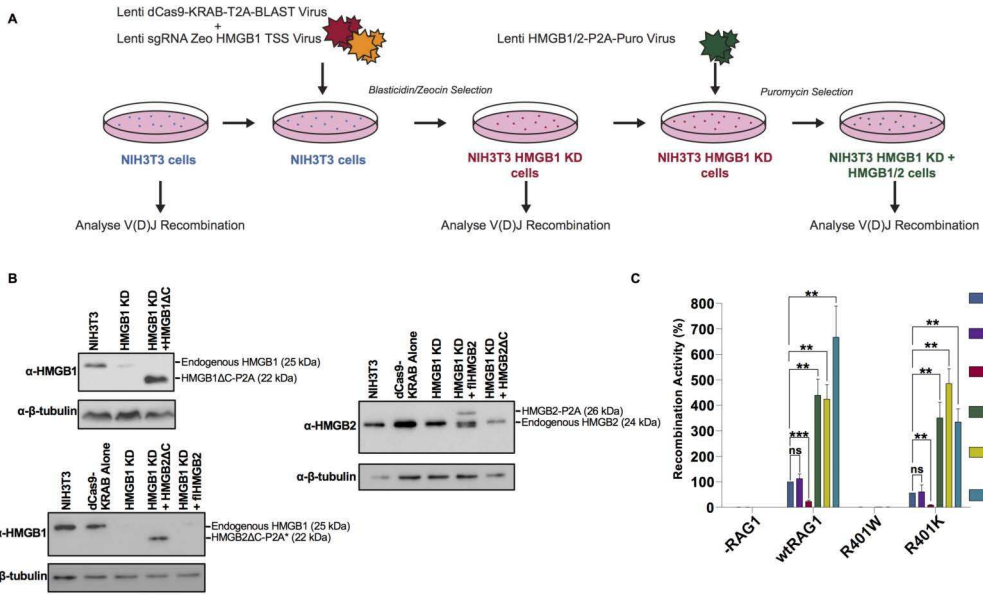
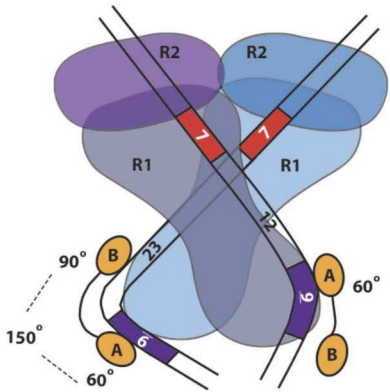

A
Lenti dCas9-KRAB-T2A-BLAST Virus
+
Lenti sgRNA Zeo HMGB1 TSS Virus
Lenti HMGB1/2-P2A-Puro Virus
Blasticidin/Zeocin Selection
Puromycin Selection
NIH3T3 cells
NIH3T3 cells
NIH3T3 HMGB1 KD cells
NIH3T3 HMGB1 KD cells
NIH3T3 HMGB1 KD + HMGB1/2 cells
Analyse V(D)J Recombination
Analyse V(D)J Recombination
Analyse V(D)J Recombination
B
NIH3T3
HMGB1 KD
HMGB1 KD +HMGB1ΔC
α-HMGB1
Endogenous HMGB1 (25 kDa)
HMGB1ΔC-P2A (22 kDa)
α-β-tubulin
NIH3T3
dCas9-KRAB Alone
HMGB1 KD
HMGB1 KD + flHMGB2
HMGB1 KD + HMGB2ΔC
α-HMGB2
HMGB2-P2A (26 kDa)
Endogenous HMGB2 (24 kDa)
α-β-tubulin
NIH3T3
dCas9-KRAB Alone
HMGB1 KD
HMGB1 KD + HMGB2ΔC
HMGB1 KD + flHMGB2
α-HMGB1
Endogenous HMGB1 (25 kDa)
HMGB2ΔC-P2A* (22 kDa)
α-β-tubulin
C
Recombination Activity (%)
800
700
600
500
400
300
200
100
0
**
**
**
***
ns
**
**
**
**
ns
-RAG1
wtRAG1
R401W
R401K
NIH3T3
dCas9-KRAB Alone
HMGB1 KD
HMGB1 KD + HMGB1ΔC
HMGB1 KD + flHMGB2
HMGB1 KD + HMGB2ΔC

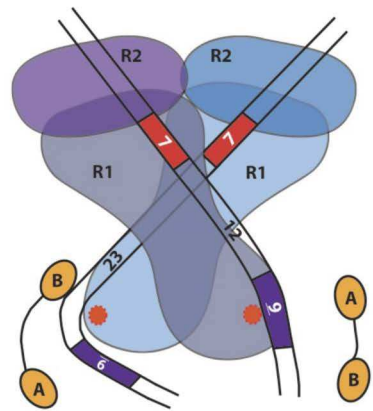
Figure 5

wtRAG1



Correct bend formation and HMGB1 recruitment
Complex active

R401W



Aberrant bend formation and HMGB1 recruitment
Complex inactive

Figure 6