

Gain-of-function analysis of cis-acting diversification elements in DT40 cells

Randolph B Caldwell¹ , Herbert Braselmann¹, Steffen Heuer¹, Ulrike Schötz² & Horst Zitzelsberger^{1,3,4}

1 Department of Radiation Sciences - Research Unit Radiation Cytogenetics, Helmholtz Center Munich - German Research Center for Environmental Health (GmbH), Neuherberg 85764, Germany

2 Department of Radiotherapy and Radiooncology, Philipps-University Marburg, University Hospital Gießen and Marburg, Marburg 35043, Germany

3 Helmholtz Center Munich, Clinical Cooperation Group 'Personalized Radiotherapy of Head and Neck Cancer', Neuherberg 85764, Germany

4 Department of Radiotherapy and Radiation Oncology, Ludwig-Maximilians-Universitaet, University Hospital Munich, Munich 81377, Germany

Keywords

Activation-induced cytosine deaminase (AID), cis-targeting elements, DNA regulatory motifs, immunoglobulin diversification, somatic hypermutation, transgene diversification.

Correspondence

Randolph B Caldwell, Department of Radiation Sciences - Research Unit Radiation Cytogenetics, Helmholtz Center Munich - German Research Center for Environmental Health (GmbH), Neuherberg 85764, Germany.
E-mail: randolph.caldwell@helmholtz-muenchen.de

Received 2 January 2018; Revised 9 April 2018; Accepted 9 April 2018

doi: 10.1111/imcb.12158

Immunology & Cell Biology 2018; 1–10

Abstract

Activation-induced cytidine deaminase (AID) is required for the immunoglobulin diversification processes of somatic hypermutation, gene conversion and class-switch recombination. The targeting of AID's deamination activity is thought to be a combination of cis- and trans-acting elements, but has not been fully elucidated. Deletion analysis of putative proximal cis-regulatory motifs, while helpful, fails to identify additive *versus* cumulative effects, redundancy, and may create new motifs where none previously existed. In contrast, gain-of-function analysis can be more insightful with fewer of the same drawbacks and the output is a positive result. Here, we show five defined DNA regions of the avian *Igλ* locus that are sufficient to confer events of hypermutation to a target gene. In our analysis, the essential cis-targeting elements fully reconstituted diversification of a transgene under heterologous promotion in the avian B-cell line DT40. Furthermore, to the best of our knowledge two of the five regions we report on here have not previously been described as individually having an influence on somatic hypermutation.

INTRODUCTION

In the jawed vertebrate's adaptive immune response, the first stage of immunoglobulin (Ig) combinatorial repertoire development is based on antigen-independent V(D)J recombination. In the second stage of B-cell maturation, stimulation induces gene conversion (GC) and/or somatic hypermutation (SHM) of the *Ig* gene, and class-switch recombination (CSR).^{1,2} All three processes, GC/SHM/CSR, strongly depend on the function of Activation-induced cytidine deaminase (AID).^{3,4} AID initiates the process by the deamination of deoxycytidine thereby

generating uracil residues in the DNA.⁵ The resulting U:G mismatches in the context of immunoglobulin SHM have not been shown to be significantly resolved via mismatch-mediated repair.⁶ Rather, the uracils are further processed by uracil-DNA glycosylase giving rise to abasic sites.⁷ An abasic site can be repaired either in a relatively error-free process such as base excision repair or re-routed into an error-prone process such as trans-lesion synthesis.^{8,9} In the case of multiple AID-induced lesions in close proximity, even error-free repair can lead to repair-induced staggered double-strand breaks, thus complicating the process and leading to repair path re-routing.¹⁰

As deamination activity has intrinsic mutagenic potential and erroneous targeting of AID can lead to B-cell lymphoma, AID's activity needs to be tightly regulated.^{11,12} Although several groups have shown that AID is not completely restricted to the *Ig* locus, its mutational activity at other loci is lower by several orders of magnitude.¹³⁻¹⁶ The mechanism of AID's locus specific enhanced targeting is not fully understood, but currently thought to include the interplay of trans-acting factors and cis-targeting elements. Some of the trans-acting factors that have been shown to exert an effect are E2A, NF- κ B, Nemo, PU.1, Pax5, PARP1 and Positive Cofactor 4 (PC4).^{10,17-21} However, it is not clear in regard to some identified trans-acting factors if the observed effect is a direct one or due to signaling and/or transcriptional activities.

In support of cis-acting elements, several studies have shown various proximal segments of DNA sequence having influence to differing degrees on SHM, GC or CSR outcomes.²²⁻²⁶ Many identified sequences, e.g. E-boxes, enhancers and silencers, point further to transcriptional influences. Deletion analysis, while helpful to identify new regions of interest, may be thwarted by redundant elements or spatio-temporal effects.²⁷ In a gain-of-function analysis, regions of DNA analyzed can be better controlled for confounding factors, more flexibly tested and activity leads to a positive result.

In this report, we show in a gain-of-function analysis the activity of five putative motif containing cis-targeting regions of DNA in DT40 Chicken B-cells that were identified in a supervised search. Using transfection constructs having constitutive expression of a fluorescent transgene target to separate out transcriptional influences, we tested the selected segments of proximal DNA individually, directionally and in various combinations. Our results clearly suggest that while each element led to differing measured levels of activity individually and in simple combinations, it is the cumulative "super-motif" architecture that defines the process.

RESULTS

Supervised search of putative Cis-targeting motifs

As an AID targeting motif database does not exist, we hypothesized that any such motifs would most likely be located within regions of conserved genome sequence homology, hypersensitive site regions and/or co-located with clusters of suspected protein/transcription-binding motifs. In addition, the pattern of motif clustering itself could be the determining cis-targeting "super-motif" driving function. Therefore, we performed a supervised motif search of DT40's *Ig* λ locus sequence using various

web-based tools, e.g. Cis-element Cluster Finder – Cister,²⁸ hidden Markov modeling – HMMER and Improbizer,^{29,30} Multiple Em for Motif Elicitation – MEME,³¹ Motif Alignment & Search Tool – MAST,³² and BLAT (WUGSC 2.1/galGal3 build).³³ When the tools used required multiple genomic sequence input for comparison, we included available immunoglobulin loci data from *Cairina moschata* (Muscovy duck), *Meleagris gallopavo* (turkey), *Mus musculus* and *Homo sapiens*. However, sequence homology alone of the closely related members of the class Aves was not used as a deciding factor. We merged the results and found five regions most likely to fit our hypothesis (Supplementary figure 1a–d). Based on the International Chicken Genome Sequencing Consortium's Gallus_gallus-4.0/galGal4 build, the five regions (designated L, R, S, U and N) identified are as follows: L – chr15: 7933125-7933251, R – chr15:7927271-7928924, S – chr15:7925941-7926238, U – chr15:7924617-7925293 and N – chr15: 7921696-7921859 (Figure 1a).³⁴ The L-region is taken from the intron (leader sequence) of *V* λ and is CpG island rich. The R-region goes from the end of the constant domain, ends prior to a putative silencer and encompasses conserved non-coding sequences to that of *Homo sapiens*, *Danio rerio* and *Xenopus laevis*. The S-region encompasses part of the putative 3'-Enhancer and an expanded E-box (CANNTG) type Basic Helix Loop Helix motif situated adjacent to an NF- κ B (GGRRNNYYCC) site (CGCAGCTGTGCGCCGGGGC ATCCC). The U-region is part of a 3' hypersensitive region and contains duplicates of 2 putative motifs found multiple times in the R-region. The N-region encompasses part of a second 3' hypersensitive region, contains a putative zinc finger motif that is also found on the reverse strand in the R-region, and conserved non-coding sequences to that of *Didelphis marsupialis*. The top three putative motif associations found in the supervised search were paired-homeo (e.g. Pax4), paired (e.g. Pax5, Pax6) and C2H2-type zinc finger (e.g. NEMO, ZAS3).

Testing of the L, R, S, U and N regions for gain-of-function

In testing for gain-of-function, we included the cell lines Aidr1 *Ig* λ - CisKoGFP2 as negative control and for positive control Aidr1 *Ig* λ - CisKoGFP2-W, which contains the "W" fragment (the complete region encompassing the V(D)J rearranged ~9.8 kb from the *Ig* λ transcription start to carbonic anhydrase 15-like, Figure 1a) as previously described.²⁵ As the CpG island rich L-region is situated in the beginning of the *V* λ segment, we wanted to test it in a more native position.

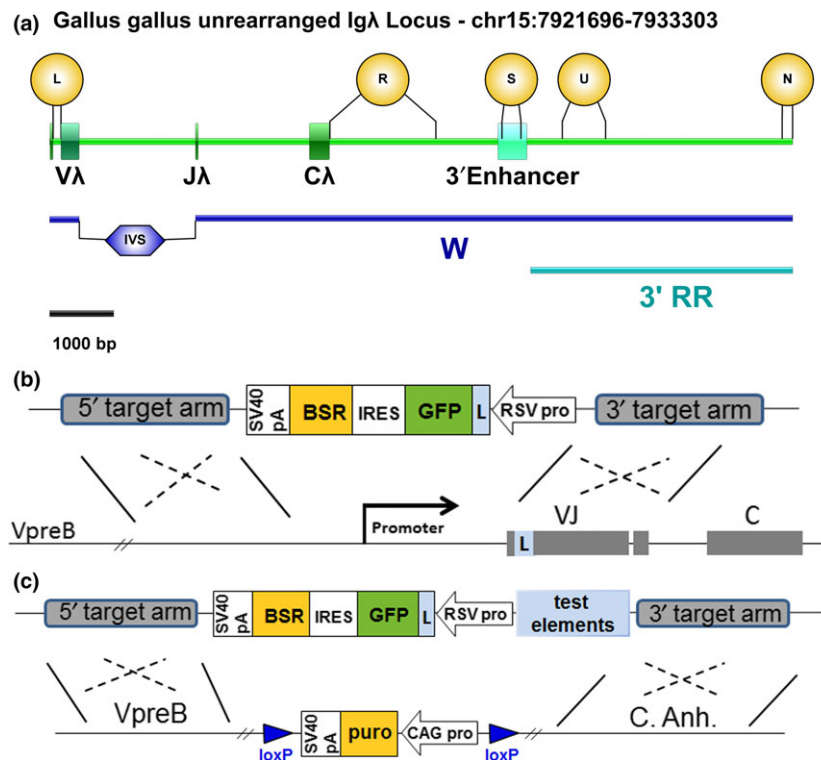


Figure 1. Mapping the five supervised search resultant regions of interest and overview of targeting constructs for their testing. **(a)** the positions of the L, R, S, U and N regions along the *Igλ* locus are shown. Chromosomal coordinates are based on the International Chicken Genome Sequencing Consortium's *Gallus_gallus*-4.0/galGal4 build. For comparison, the rearranged W-region²⁵ (absent the intervening sequence – IVS) is shown in blue and the 3' RR^{24,26} is shown in turquoise; **(b)** cartoon depicting the targeting of the *Igλ*-promoter for knock-out by pFG2 constructs. Not drawn to scale; **(c)** cartoon depicting the targeting of the *Igλ*-locus for knock-out by pCisKoGFP2 and pFG3 constructs. Not drawn to scale.

To achieve this, we cloned the L-region into the test plasmids between the RSV promoter and the eGFP transgene reporter while the R, S, U and N test elements were cloned into the constructs upstream of the RSV promoter for targeting of the *Igλ*-promoter or the entire *Igλ* locus (Figure 1b, c).

We observed in our fluctuation analysis, 2 weeks (~25 generations) post subcloning, that each region individually induces some level of activity with the N-region and triplicate of the S-region cloned in the reverse direction demonstrating the strongest effect (Figure 2a). While we see some directional variation in regard to the S-region, the more prominent effect comes from element multiplicity. In Figure 2b, we start to observe the combinatorial effects with RU and RSSN regions seemingly having the strongest effect. In Figure 2c, we see the cumulative importance of all the elements together whereby RSSUN appears to induce higher activity than SSUN, RSSU (Wilcoxon rank-sum test: $P = 0.00012$) and even the complete “W” region (Wilcoxon rank-sum test: $P = 0.048$). To rule out

transcription levels playing a role in these results, we tested two primary clones from 18 selected element test clones generated and biological replicates of both the CiscoGFP2 (Cl9) negative control and CisKoGFP2-W (Cl3) positive control using RT-qPCR. As can be seen by the ΔCT in Figure 3, transcription levels were stable in comparison to the BAFF reference gene regardless of selected element tested.

For further testing of the combined regions, we employed our more flexibly and modularly designed pForGene (pFG) plasmid system. The pFG plasmid's multiple multi-cloning-site's design allows for the cloning of test elements both upstream and downstream of the RSV promoter, and also allows for the easy exchange of selectable marker cassettes. Initially, we wanted to test the L-region's activity in *Aidr1* ψ^- cells using pFG2 which targets the *Igλ*-promoter (Figure 1b) while leaving the “W” region in its native position. As can be seen in Figure 4, the L-region has a significant enhancing effect (Wilcoxon rank-sum test: *Aidr1* ψ^- -pFG2GFP-L versus *Aidr1* ψ^- -pFG2GFP $P \leq 0.0001$). The

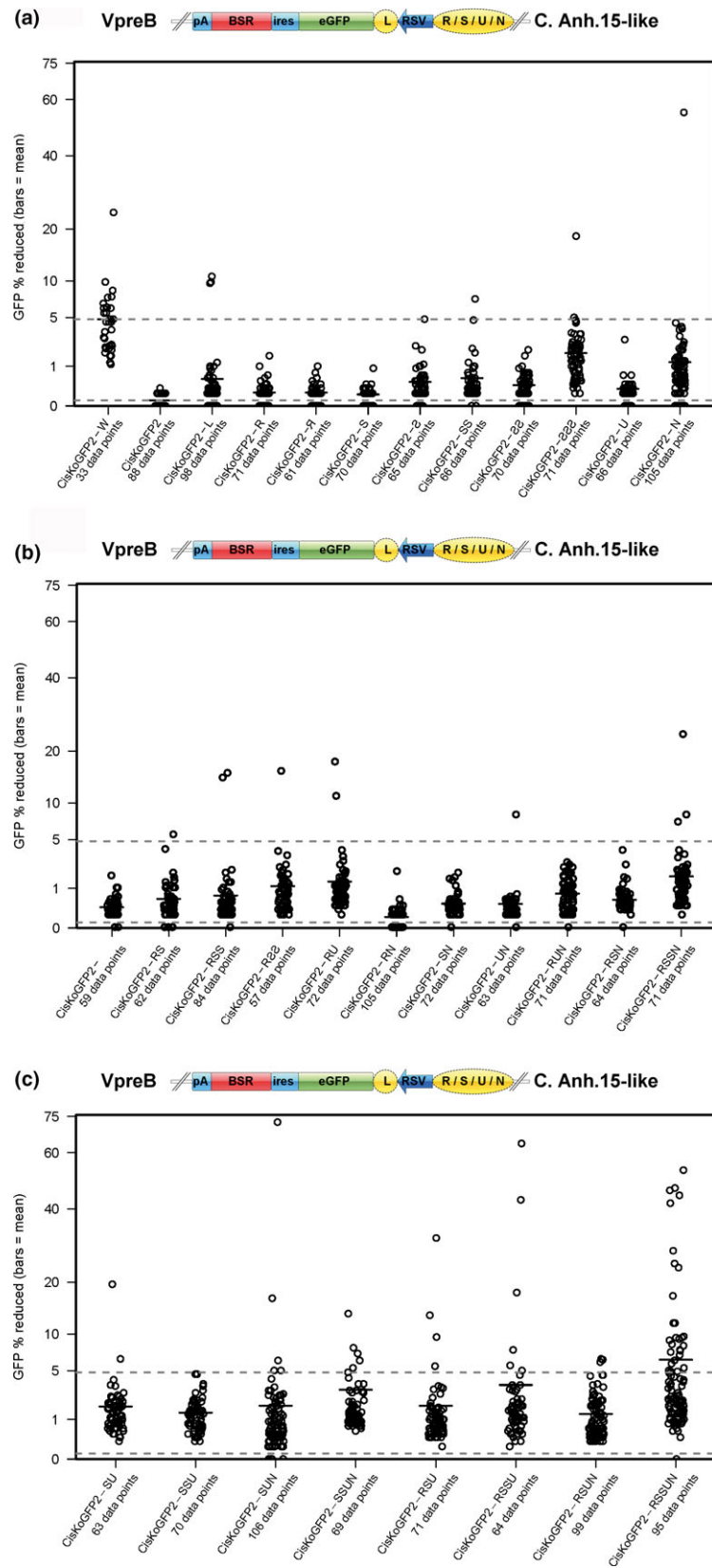


Figure 2. The selected L/R/S/U/N regions individually, multiply and cumulatively induce hypermutation activity in DT40 cells. Fluctuation analysis of fluorescence intensity reduction/loss; **(a)** individual regions tested singly, multiply and in some cases directionally (signified by backwards lettering), **(b)** regions tested in simple combinations including directionally, **(c)** regions further tested in multiple combinations. Each element was tested using 2–4 primary clones, and the CisKoGFP2 negative control was tested as a biological triplicate. The reconstituted positive control, CisKoGFP2-W, was a single primary clone included for reference. Primary clones and biological replicates were subcloned by limiting dilution and up to 36 subclones for each were measured by flow cytometry for GFP fluorescence intensity reduction/loss after 14 days (25 generations). Each data point represents the result of a single subclone and the bar indicates the mean. The dotted reference lines signify the mean values of the *Aid^{r1} Igλ- CisKoGFP2-W* positive control and the *Aid^{r1} Igλ- CisKoGFP2* negative control. Genome representation not drawn to scale. Wilcoxon rank-sum test: RSSUN versus 'W' $P = 0.048$; RSSUN versus SSUN and RSSU $P = 0.00012$.

effect is similar in the $\Delta PC4$ background suggesting a role that is not associated with damage repair as we have previously shown for Positive Cofactor 4.¹⁰ This argues for the L-region's role being most probably associated with the targeting or enhancement of AID activity. This was further confirmed with sequencing 2-week bulk cultures of *Aidr1* ψ - pFG2GFP, *Aidr1* ψ - pFG2GFP-L and *Aidr1* ψ - $\Delta PC4$ pFG2GFP-L (Figure 5a–c). While the L-region alone increased the mutation rate slightly, the mutation spectrum (Figure 5d) results are in line with our previously published results for *Aidr1* ψ - pFG2GFP (5×10^{-5} mut/bp/gen, trs:trv = 1:0.95) (mutations/base pair/generation, transitions:transversions) and *Aidr1* ψ - $\Delta PC4$ pFG2GFP (4.7×10^{-5} mut/bp/gen, trs:trv = 1:3.4).¹⁰ Interestingly, both the L-region alone and the L-region in the $\Delta PC4$ genomic background decreased the number of mutations falling outside the canonical WGCW hotspots³⁵ in comparison to the FG2GFP base-line control, with hotspot:non-hotspot going from 1:1.14 to 1:0.97 and 1:0.78, respectively. This

is even more pronounced if hotspots are expanded to include ± 6 bp with hotspot ± 6 :non-hotspot going from 1:0.67 to 1:0.28 and 1:0.34, respectively, with the ± 4 th bp position showing the largest increase: FG2GFP ± 4 th bp position, 1 mutation; FG2GFP-L ± 4 th bp position, 19 mutations; $\Delta PC4$ FG2GFP-L ± 4 th bp position, 10 mutations. For comparison, from our previous published results of $\Delta PC4$ FG2GFP without the L-region,¹⁰ the hotspot:non-hotspot was 1:0.97, the hotspot ± 6 :non-hotspot was 1:0.53, and the ± 4 th bp position had only 4 mutations.

In Figure 6, we see the results of L and RSSUN in the pFG3 construct which targets the entire *Igλ* locus for knock-out (Figure 1c). We tested the RSSUN in the forward (+) and reverse (-) directions, with and without the L-region, and in some instances with and without AID. As can be readily observed, the L-RSSUN (-) and L-RSSUN(+), as well as RSSUN(+) achieved full reconstitution of activity in comparison to the *Aidr1 Igλ- CisKoGFP2-W* positive control. Interestingly,

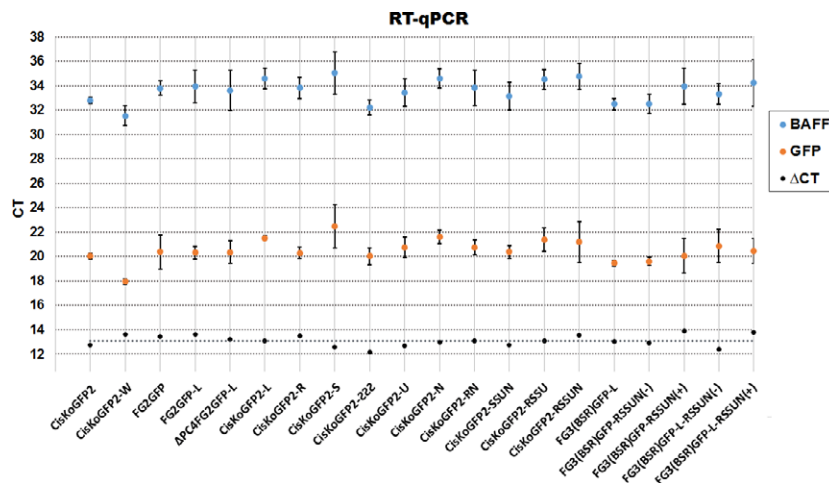


Figure 3. Testing of representative clones to rule out transcriptional influences by the regions of interest. We selected 20 representative clones (two primary clones for each of 18 elements tested and biological duplicates for the negative and reconstituted positive controls) for transcription level comparison by RT-qPCR for expression of GFP in comparison to the BAFF reference gene. Mean CT for BAFF and GFP are shown with standard deviation bars. As the relative ΔCt shows, transcription levels were unaffected by the elements tested. ΔCt TREND = 13.107 ($SD = 0.4651$).

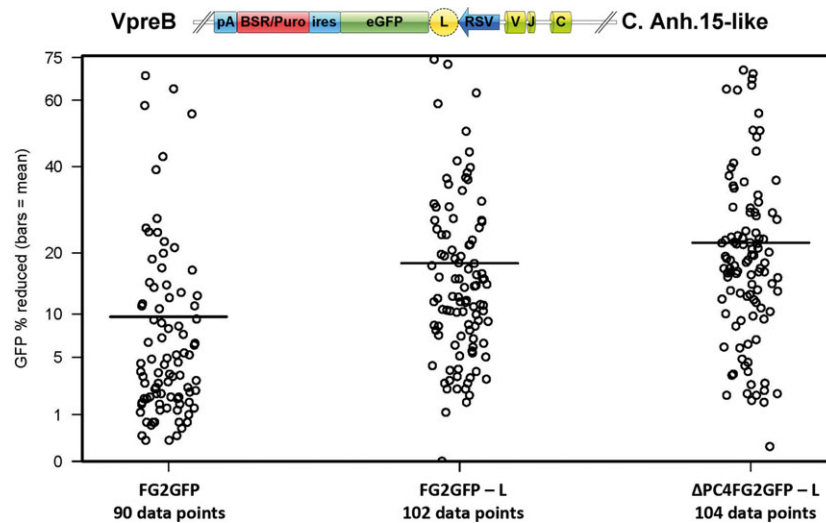


Figure 4. Confirmation of L-region activity in Aid^{r1} ψ⁻ and Aid^{r1} ψ⁻ ΔPC4 cells. Fluctuation analysis of GFP fluorescence intensity reduction/loss. Three primary clones for each were subcloned by limiting dilution and up to 36 subclones of each primary clone were measured by flow cytometry for GFP intensity after 14 days (25 generations). Each data point represents the result of a single subclone and the bar indicates the mean. Genome representation not drawn to scale. Wilcoxon rank-sum test: Aid^{r1} ψ⁻ pFG2GFP-L and Aid^{r1} ψ⁻ ΔPC4 pFG2GFP-L versus Aid^{r1} ψ⁻ pFG2GFP $P \leq 0.0001$; Aid^{r1} ψ⁻ ΔPC4 pFG2GFP-L versus Aid^{r1} ψ⁻ pFG2GFP-L $P = 0.032$.

the RSSUN(-) without the L-region performed less well. However, as expected, the absence of AID reduced the GFP fluorescence intensity reduction rate to effectively 0%.

DISCUSSION

In the age of omics-based Big Data, even small experimentally validated functional datasets are crucial to informing training datasets in order to improve predictive modeling.³⁶ While there exists a substantial body of resources for transcription and transcription-related factors, little is known concerning more complex regulatory systems encoded in the DNA that are not as amenable to current bulk data producing protocols. With limited knowledge of AID activity directing and enabling cis-targeting motifs *a priori*, we approached the supervised motif region search systematically using well-defined parameters regarding conserved sequence homology, motif co-localization and biochemically elucidated hypersensitive sites. We also held the hypothesis that multiple motif-containing regions would most probably be involved since the end effect is the deamination of genomic DNA, an action that needs to be tightly controlled. With our supervised motif search and gain-of-function approach, we identified five distinct cis-regulatory regions that have a role in targeting AID activity that appears in our system, via the use of a constitutive promoter, to be uncoupled from transcription/expression regulation influence.

Although we have identified five distinct functional regions, our experimental system is limited in that it is not geared for high-throughput studies. It is simply not feasibly practical to test each region base by base in living cells to further define the exact DNA motifs and associated contextual nucleotide flanking sequences within the regions identified. For further motif analysis, it would be preferable to use ours and other's published results to aid computer-assisted Big Data modeling to reduce the number of clones needed for further validation studies. We partly did this in regard to the multiplicity of the S-region, as proximal regions upstream and downstream contained additional desired canonical and non-canonical adjacently located E-box and NF-kB sites.³⁷ However, those regions also contained putative motifs that were selected against in our supervised motif search, e.g. silencers. As the *Igλ* locus is highly repetitive and the R, U and N regions already contain multiple repeats of selected putative motifs, we chose to test the S-region multiply. By doing so, we overcame the possible inclusion of activity reducing regulatory elements to achieve a similar "super-motif" architecture.

Our results will complement and inform many of the newer Big Data strategies employed to uncover AID on- and off-target regulation such as epigenetic AID targeting.³⁸ AID activity is a critical biological function that can also be highly detrimental if not properly regulated. AID's off-target aspect is of great importance as besides lymphomagenesis, AID has recently been

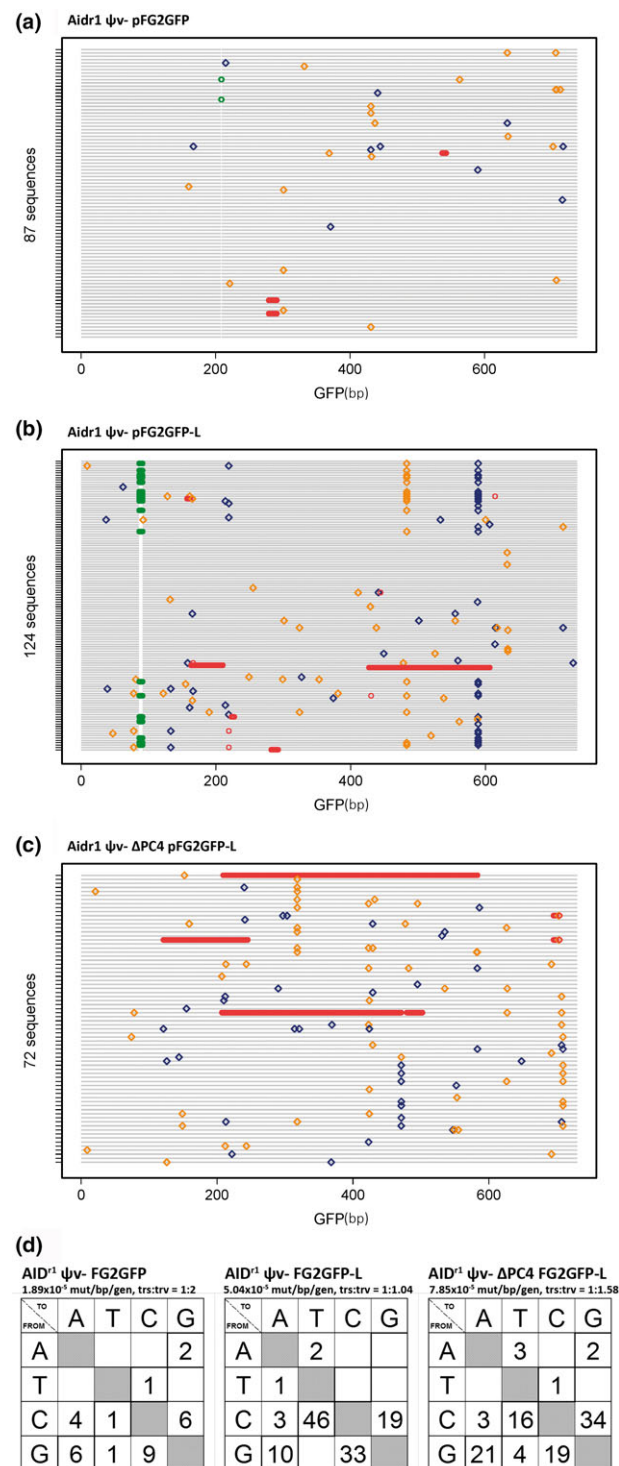


Figure 5. Quick-view and mutational analysis of the L-region effect at the DNA level in DT40 cells. Quick-view analysis of **(a)** Aid^{r1} ψ - pFG2GFP, **(b)** Aid^{r1} ψ - pFG2GFP-L and **(c)** Aid^{r1} ψ - Δ PC4 pFG2GFP-L clones. Following two weeks bulk growth (25 generations), gDNA was recovered from each of the 9 primary clones (three each) and the GFP gene was sequenced for detailed analysis. Transitions and transversions are displayed as blue and gold diamonds, respectively, and insertions and deletions are displayed as green and red circles, respectively. Mutation analysis is shown in **d**: the bold-outlined diagonal boxes from lower left to upper right indicate transition (trs) changes, all others are transversions (trv).

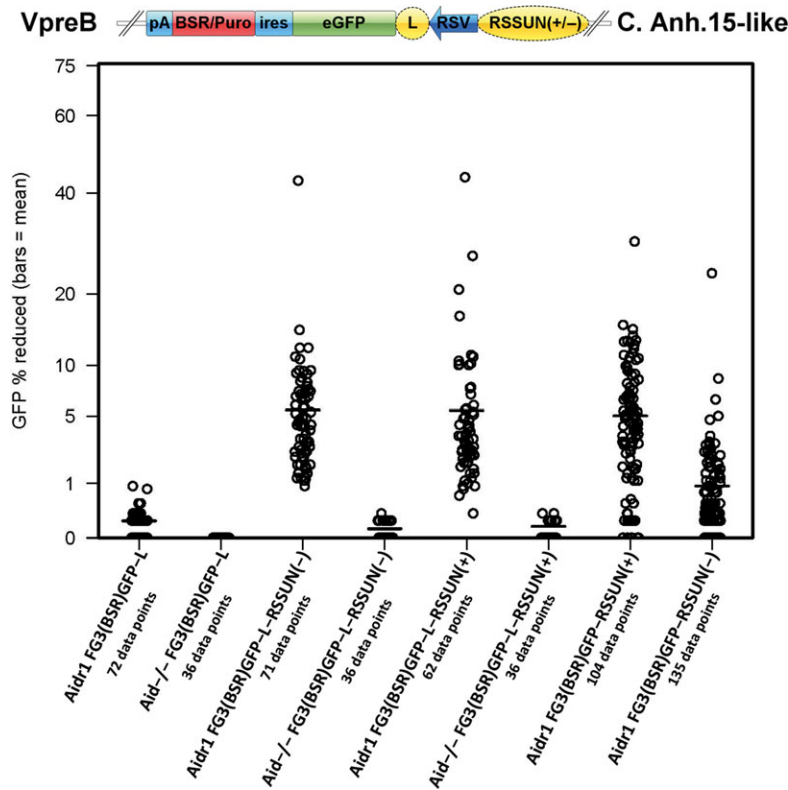


Figure 6. Confirmation and directional testing of ‘RSSUN’ activity with and without the L-region in DT40 Ig-p (cis knockout) cells. Fluctuation analysis of GFP fluorescence intensity reduction/loss. Each element configuration tested had 2–4 primary clones, and a single primary clone for each of the Aid1/- constructs was tested. Cells were subcloned by limiting dilution and up to 36 subclones for each were measured after 14 days (25 generations) by flow cytometry for fluorescence intensity reduction/loss. Each data point represents the result of a single subclone and the bar indicates the mean. Genome representation not drawn to scale.

implicated in active demethylation, kataegis and non-lymphoid cell carcinogenesis.^{39–41}

METHODS

Cell culture, cell transfection and growth curve

DT40 cells were cultured in DMEM/F-12 supplemented with 10% fetal bovine serum, 1% chicken serum, 2 mmol L⁻¹ L-glutamine, 0.1 mmol L⁻¹ β-mercaptoethanol and penicillin/streptomycin (chicken media, CM) at 41°C in a 5% CO₂ environment. Transfections were carried out using the Gene Pulser Xcell System with CE Module (Bio-Rad, Hercules, CA, USA). Briefly, 10⁷ logarithmically growing cells of ≥80% viability were harvested for each transfection. The cells were resuspended in 800 μL ice-cold PBS with 40 μg of linearized plasmid and placed into an electroporation cuvette (0.4 cm gap). For transfecting DT40 cells, the exponential protocol setting with 700 V and 25 μF was used. Upon electroporation, the cells were transferred immediately into 10 mL CM and then aliquoted at 100 μL per well across a 96-well plate and incubated at 41°C. After 24 h (DT40's doubling rate is ~13 h¹⁰), 100 μL of appropriate drug selection media was added per well (final concentration: Blastidicin S 15 μg mL⁻¹, Puromycin

0.5 μg mL⁻¹) and incubated for a further 1–2 weeks for colony picking and testing. Targeting in DT40 was confirmed via PCR using primer pairs PS31-PU5 for Ig-Pro-KO (pFG2 constructs), PS169-PU5 for Igλ-KO (pFG3 and pCisKoGFP2 constructs).

PCR, cloning and sequencing

For the initial cloning and testing of selected cis-regions, we used the pCisKoGFP2 (targeting the entire Igλ locus) construct as previously described.²⁵ The GFP2 cassette has a modified cloning site upstream of the RSV-GFP-ires-BSR-SV40pA consisting of NheI and SpeI. The R,S,U,N fragments have the following PCR introduced modular designed restriction sites for bi-directional, individual, multiple and sequential cloning (and primers used): R – AvrII/Xba-SpeI/NheI/AvrII (rbc557, rbc558), S – SpeI/NdeI-NdeI/NheI (rbc559, rbc561), U – SpeI/KasI-KasI/NheI (rbc563, rbc564), N – SpeI/NsiI-NsiI/NheI/BamHI (rbc565, rbc566). The L-region was assembled for cloning using the overlapping oligos rbc261, rbc262, rbc263 and rbc264 that incorporated a NheI site on the 5' end and a SpeI site on the 3' end. For further cloning of the L-region, rbc600 and rbc601 were used to introduce a SpeI site 5' and an NheI site 3'. For confirmation testing in DT40 cells, we used our pForGene (pFG)

system as previously described.¹⁰ Briefly, the pFG's cassette design is BamHI-SpeI-RSV-NheI/EcoRI/BglII-IRES-NcoI-BSR/Puro-NcoI-SV40pA-BamHI in a multi-cloning site-modified pBluescript KS+ (Agilent Technologies, Santa Clara, CA, USA) backbone. The base construct of pFG2 has homologous arms targeting the knock-out of DT40's rearranged *Igλ*'s promoter (5' arm – chr15: 7933638-7935749, 3' arm (minus intervening sequence) – chr15: 7930437-7961015/7932855-7933303), while the construct of pFG3 has homologous arms targeting the knock-out of DT40's entire *Igλ* locus (5' arm – chr15: 7955324-7957914, 3' arm – chr15: 7920634-7921694) as previously described.^{10,25} For the cloning of the transgene fluorescent marker, the primer pair rbc598-rbc602 was used for eGFP. All cloning was carried out using restriction enzymes and competent cells (New England Biolabs, Ipswich, MA, USA), and the Takahara DNA Ligation Kit (Clontech, Mountain View, CA, USA) per the manufacturers' instructions.

For sequencing of the GFP transgene, genomic DNA (gDNA) was isolated using the DNeasy Blood & Tissue Kit (Qiagen, Venlo, the Netherlands) according to the manufacturer's instructions. The gDNA served as template for the primer pair rbc583-rbc587 and the PCR products were cloned for sequencing as previously described.¹⁰ Bi-directional sequencing was done using the Big Dye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific) with primers rbc776 and rbc585 (LGFP), and rbc734 and rbc585 (GFP) as previously described.^{10,42} Primers used are listed in Supplementary table 1, and Supplementary table 2 gives an overview of clones used and experiments performed.

RT-qPCR

To assess whether RSV-promoter driven transcription was being influenced by the cis-regions tested, cDNA was made from total mRNA using the RNeasy Mini Kit (Qiagen) and SuperScript III First-Strand Synthesis System (Thermo Fisher Scientific) according to the manufacturer's instructions. Quantitative PCR (qPCR) was performed with the LightCycler 1.5 (Roche) using the FastStart DNA Master SYBR Green I kit (Roche) according to the manufacturer's instructions. Primers used were rbc9-rbc10 for the B-cell activating factor (BAFF) control, and rbc843-rbc844 for GFP.

Flow cytometry

For monitoring of cells based on fluorescence status, cells were washed twice and resuspended in sterile PBS. Flow cytometry was performed using the LSR II (BD Biosciences), whereby 5000–20 000 live cells were counted for each subclone using 488 nm for excitation, and measuring GFP at 530 nm. Gating for fluorescence intensity reduction/loss measurement was set at >2-fold below the cloud and subclones demonstrating ≥75% reduction were discounted as likely harboring a mutation prior to subcloning.

Statistical analysis

For comparison of the results generated in the fluctuation analyses, the Wilcoxon rank-sum test (also

called the Mann–Whitney *U*-test) was used to compare the subclone data points, and *P*-values less than 0.05 were considered as significant.

ACKNOWLEDGMENTS

We thank Dr Sabine Possmann for her prior contributions to the cis-project. This work was funded by Helmholtz Center Munich (HMGU)-German Research Center for Environmental Health (GmbH) core funds. HMGU had no influence on the design of the study, or the collection, analysis and interpretation of the data.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

DATA AVAILABILITY

All results generated for this study are included in this article and its supplementary information files.

REFERENCES

- Alt FW, Blackwell TK, Yancopoulos GD. Development of the primary antibody repertoire. *Science* 1987; **238**: 1079–1087.
- Nagaoka H, Yu W, Nussenzweig MC. Regulation of RAG expression in developing lymphocytes. *Curr Opin Immunol* 2000; **12**: 187–190.
- Muramatsu M, Kinoshita K, Fagarasan S, Yamada S, Shinkai Y, Honjo T. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* 2000; **102**: 553–563.
- Barreto V, Reina-San-Martin B, Ramiro AR, McBride KM, Nussenzweig MC. C-terminal deletion of AID uncouples class switch recombination from somatic hypermutation and gene conversion. *Mol Cell* 2003; **12**: 501–508.
- Bransteitter R, Pham P, Scharff MD, Goodman MF. Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proc Natl Acad Sci USA* 2003; **100**: 4102–4107.
- Campo VA, Patenaude AM, Kaden S, *et al.* MSH6- or PMS2-deficiency causes re-replication in DT40 B cells, but it has little effect on immunoglobulin gene conversion or on repair of AID-generated uracils. *Nucleic Acids Res* 2013; **41**: 3032–3046.
- Di Noia JM, Neuberger MS. Immunoglobulin gene conversion in chicken DT40 cells largely proceeds through an abasic site intermediate generated by excision of the uracil produced by AID-mediated deoxycytidine deamination. *Eur J Immunol* 2004; **34**: 504–508.
- Di Noia J, Neuberger MS. Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. *Nature* 2002; **419**: 43–48.
- Nussenzweig MC, Alt FW. Antibody diversity: one enzyme to rule them all. *Nat Med* 2004; **10**: 1304–1305.

10. Caldwell RB, Braselmann H, Schoetz U, Heuer S, Scherthan H, Zitzelsberger H. Positive Cofactor 4 (PC4) is critical for DNA repair pathway re-routing in DT40 cells. *Sci Rep* 2016; **6**: 28890.
11. Okazaki IM, Kotani A, Honjo T. Role of AID in tumorigenesis. *Adv Immunol* 2007; **94**: 245–273.
12. Pérez-Durán P, de Yébenes VG, Ramiro AR. Oncogenic events triggered by AID, the adverse effect of antibody diversification. *Carcinogenesis* 2007; **28**: 2427–2433.
13. Martin A, Scharff MD. Somatic hypermutation of the AID transgene in B and non-B cells. *Proc Natl Acad Sci USA* 2002; **99**: 12304–12308.
14. Kotani A, Okazaki IM, Muramatsu M, et al. A target selection of somatic hypermutations is regulated similarly between T and B cells upon activation-induced cytidine deaminase expression. *Proc Natl Acad Sci USA* 2005; **102**: 4506–4511.
15. Liu M, Duke JL, Richter DJ, et al. Two levels of protection for the B cell genome during somatic hypermutation. *Nature* 2008; **451**: 841–845.
16. Yamane A, Resch W, Kuo N, et al. Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. *Nat Immunol* 2011; **12**: 62–69.
17. Schoetz U, Cervelli M, Wang YD, Fiedler P, Buerstedde JM. E2A expression stimulates Ig hypermutation. *J Immunol* 2006; **177**: 395–400.
18. Jain A, Ma CA, Lopez-Granados E, et al. Specific NEMO mutations impair CD40-mediated c-Rel activation and B cell terminal differentiation. *J Clin Invest* 2004; **114**: 1593–1602.
19. Jundt F, Kley K, Anagnostopoulos I, et al. Loss of PU1 expression is associated with defective immunoglobulin transcription in Hodgkin and Reed-Sternberg cells of classical Hodgkin disease. *Blood* 2002; **99**: 3060–3062.
20. Levin-Klein R, Kirillov A, Rosenbluh C, Cedar H, Bergman Y. A novel pax5-binding regulatory element in the Igk locus. *Front Immunol* 2014; **5**: 240.
21. Paddock MN, Buelow BD, Takeda S, Scharenberg AM. The BRCT domain of PARP-1 is required for immunoglobulin gene conversion. *PLoS Biol* 2010; **8**: e1000428.
22. Michael N, Shen HM, Longerich S, Kim N, Longacre A, Storb U. The E box motif CAGGTG enhances somatic hypermutation without enhancing transcription. *Immunity* 2003; **19**: 235–242.
23. Conlon TM, Meyer KB. The chicken Ig light chain 3'-enhancer is essential for gene expression and regulates gene conversion via the transcription factor E2A. *Eur J Immunol* 2006; **36**: 139–148.
24. Kothapalli N, Norton DD, Fugmann SD. Cutting edge: a cis-acting DNA element targets AID-mediated sequence diversification to the chicken Ig light chain gene locus. *J Immunol* 2008; **180**: 2019–2023.
25. Blagodatski A, Batrak V, Schmidl S, et al. A cis-acting diversification activator both necessary and sufficient for AID-mediated hypermutation. *PLoS Genet* 2009; **5**: e1000332.
26. Kothapalli NR, Collura KM, Norton DD, Fugmann SD. Separation of mutational and transcriptional enhancers in Ig genes. *J Immunol* 2011; **187**: 3247–3255.
27. Bébin AG, Carrion C, Marquet M, et al. In vivo redundant function of the 3' IgH regulatory element HS3b in the mouse. *J Immunol* 2010; **184**: 3710–3717.
28. Frith MC, Hansen U, Weng Z. Detection of cis-element clusters in higher eukaryotic. *DNA Bioinformatics* 2001; **17**: 878–889.
29. Eddy SR. Profile hidden markov models. *Bioinformatics* 1998; **14**: 755–763.
30. Ao W, Gaudet J, Kent WJ, Muttumu S, Mango SE. Environmentally induced foregut remodeling by PHA-4/FoxA and DAF-12/NHR. *Science* 2004; **305**: 1743–1746.
31. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* 1994; **2**: 28–36.
32. Bailey TL, Gribskov M. Score distributions for simultaneous matching to multiple motifs. *J Comput Biol* 1997; **4**: 45–59.
33. Kent WJ. BLAT—the BLAST-like alignment tool. *Genome Res* 2002; **12**: 656–664.
34. Liu W, Xie Y, Ma J, et al. IBS: an illustrator for the presentation and visualization of biological sequences. *Bioinformatics* 2015; **31**: 3359–3361.
35. Wei L, Chahwan R, Wang S, et al. Overlapping hotspots in CDRs are critical sites for V region diversification. *Proc Natl Acad Sci USA* 2015; **112**: E728–E737.
36. Kantorovitz MR, Kazemian M, Kinston S, et al. Motif-blind, genome-wide discovery of cis-regulatory modules in *Drosophila* and mouse. *Dev Cell* 2009; **17**: 568–579.
37. Wong D, Teixeira A, Oikonomopoulos S, et al. Extensive characterization of NF-κB binding uncovers non-canonical motifs and advances the interpretation of genetic functional traits. *Genome Biol* 2011; **12**: R70.
38. Wang Q, Oliveira T, Jankovic M, et al. Epigenetic targeting of activation-induced cytidine deaminase. *Proc Natl Acad Sci USA* 2014; **111**: 18667–18672.
39. Ramiro AR, Barreto VM. Activation-induced cytidine deaminase and active cytidine demethylation. *Trends Biochem Sci* 2015; **40**: 172–181.
40. Casellas R, Basu U, Yewdell WT, Chaudhuri J, Robbiani DF, Di Noia JM. Mutations, kataegis and translocations in B cells: understanding AID promiscuous activity. *Nat Rev Immunol* 2016; **16**: 164–176.
41. Sapoznik S, Bahar-Shany K, Brand H, et al. Activation-induced cytidine deaminase links ovulation-induced inflammation and serous carcinogenesis. *Neoplasia* 2016; **18**: 90–99.
42. Schoetz U, Deliolanis NC, Ng D, et al. Usefulness of a Darwinian system in a biotechnological application: evolution of optical window fluorescent protein variants under selective pressure. *PLoS ONE* 2014; **9**: e107069.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

© 2018 The Authors
Immunology & Cell Biology published by John Wiley & Sons Australia, Ltd on behalf of Australasian Society for Immunology Inc.
 This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.