Identification of Core DNA Elements That Target Somatic Hypermutation

Kristin M. Kohler, Jessica J. McDonald, Jamie L. Duke, Hiroshi Arakawa, Sally Tan, Steven H. Kleinstein, Jean-Marie Buerstedde and David G. Schatz

*J Immunol* published online 19 October 2012
http://www.jimmunol.org/content/early/2012/10/19/jimmunol.1202082

Supplementary Material  
http://www.jimmunol.org/content/suppl/2012/10/19/jimmunol.1202082.DC1

Subscription  
Information about subscribing to *The Journal of Immunology* is online at: http://jimmunol.org/subscription

Permissions  
Submit copyright permission requests at: http://www.aai.org/About/Publications/JI/copyright.html

Email Alerts  
Receive free email-alerts when new articles cite this article. Sign up at: http://jimmunol.org/alerts
Identification of Core DNA Elements That Target Somatic Hypermutation

Kristin M. Kohler,*1 Jessica J. McDonald,* Jamie L. Duke,† Hiroshi Arakawa,‡ Sally Tan,* Steven H. Kleinstein,†,§ Jean-Marie Buerstedde,* and David G. Schatz*1

Somatic hypermutation (SHM) diversifies the V region of Ig genes and underlies the process of affinity maturation, in which B lymphocytes producing high-affinity Abs are generated and selected. SHM is triggered in activated B cells by deamination of deoxycytosine residues mediated by activation-induced deaminase (AID). Whereas mistargeting of SHM and AID results in mutations and DNA damage in many non-Ig genes, they act preferentially at Ig loci. The mechanisms responsible for preferential targeting of SHM and AID activity to Ig loci are poorly understood. Using an assay involving an SHM reporter cassette inserted into the Ig L chain locus (IgL) of chicken DT40 B cells, we have identified a 1.9-kb DIVAC (diversification activator) element derived from chicken IgL, that supports high levels of AID-dependent mutation activity. Systematic deletion analysis reveals that targeting activity is spread throughout much of the sequence and identifies two core regions that are particularly critical for function: a 200-bp region within the IgL enhancer, and a 350-bp 3′ element. Chromatin immunoprecipitation experiments demonstrate that whereas DIVAC does not alter levels of several epigenetic marks in the mutation cassette, it does increase levels of serine-5 phosphorylated RNA polymerase II in the mutation target region, consistent with an effect on transcriptional elongation/pausing. We propose that multiple, dispersed DNA elements collaborate to recruit and activate the mutational machinery at Ig gene variable regions during SHM. The Journal of Immunology, 2012, 189: 000–000.

The ability of AID to trigger the formation of DNA mutations, breaks, and deletions in the genome suggests a propensity to contribute to genomic instability (15, 16), the risks of which would be greatly increased if AID acted outside of the Ig loci. Hence, mechanisms to restrict AID action to Ig genes would seem highly desirable.

Results from studies in the past 15 years indicate that although such mechanisms exist, they are imperfect. AID has been linked to mutations and large-scale genome rearrangements (particularly chromosomal translocations) involving both Ig and non-Ig genes in B cell lymphomas (17), and numerous non-Ig genes have been found to be mutated in an AID-dependent manner in normal germinal center B cells (18). A large-scale sequencing analysis of germinal center B cells found that approximately one-quarter of the genes analyzed sustained mutations as a result of SHM and...
that at least one-half of them were deaminated at detectable levels by AID (19). Ig genes, however, were found to undergo SHM and be deaminated by AID at frequencies 10- to 1000-fold higher than those of non-Ig genes, consistent with the existence of mechanisms to target AID and/or AID activity preferentially to Ig genes (19). These mechanisms remain poorly defined, leaving a significant gap in our understanding of how the genome is protected from AID-mediated instability.

Numerous studies have pursued the hypothesis that DNA sequences associated with Ig loci function as SHM/AID targeting elements (20–23). Such analyses have ruled out an essential role for the IgV exon itself (24) or the IgV promoter (25), although the IgV promoter can enhance the efficiency of AID-mediated diversification (26). Analyses of the targeting function of Ig gene enhancers have yielded conflicting results, leaving their role as mutation targeting elements uncertain (20–22). The E box sequence CANNTG, as well as the E2a-encoded E box binding proteins E12/E47, have been linked to SHM targeting (19, 27–31), although the magnitude and mechanism of their contribution to AID/SHM targeting remains unknown.

Several groups have taken advantage of the relatively small size of the chicken IgL locus, and the ease of manipulating the genome of the DT40 chicken B cell line (3), to search for IgL-associated DNA elements capable of targeting AID-mediated sequence diversification activity. Collectively, these studies indicate that sequences downstream of the chicken IgL C region exon function as a diversification activator (DIVAC) element, capable of conferring AID-mediated GCV/SHM on a nearby transcription unit driven by the IgV promoter or a heterologous promoter (33–38). DIVAC is active in multiple locations in the DT40 genome (34, 36) and contains evolutionarily conserved sequence elements (38) and binding sites for a number of transcription factors, some of which have been suggested to be relevant to DIVAC function (35, 37, 38). These studies, however, did not converge on particular sequence motifs or protein factors that account for DIVAC function. Furthermore, the mechanism of action of DIVAC is not understood at even a simplistic level, although it is clear that DIVAC contains functions beyond those of a transcriptional enhancer (38).

In this study, we used an established SHM assay involving a heterologous expression cassette targeted to the IgL locus of DT40 cells (34) to identify DNA sequences downstream of the chicken IgL C region exon critical for DIVAC function. This allowed us to define a 1928-bp composite element with strong DIVAC activity, which we refer to as DIVAC 1928. Extensive deletion and mutation analyses revealed that functionally important elements are broadly dispersed throughout DIVAC 1928 and also identified two “core” regions that are particularly important for DIVAC function. Chromatin immunoprecipitation experiments demonstrate that DIVAC increases levels of serine-5 phosphorylated Pol II within the mutating region but does not alter many other parameters of the cassette including levels of activating and repressive epigenetic modifications or the Pol II-associated factor Spt5. The striking dispersion of DIVAC function throughout DIVAC 1928 and the DIVAC core regions strongly suggests that multiple DNA sequences, and possibly multiple protein factors, collaborate to recruit and activate the mutation machinery during SHM/GCV and helps explain why the precise definition of SHM targeting elements has proved so elusive.

**Materials and Methods**

**Targeting constructs**

Targeting constructs were derived by modifying the pLG(-)GFP2 plasmid (34). Test DNA fragments were PCR amplified using Phusion polymerase (New England Biolabs) and cloned into unique Nhel/Spel sites in pLG(-)GFP2 (primers and DNA templates, Supplemental Table IA). Fragment deletions depicted in Figs. 2–7 were assembled by PCR. Primer sets were made to include the sequence 5′ of the deletion (reaction 1: F1/R1 primer set) and 3′ of the deletion (reaction 2: F2/R2 primer set) where roughly 15 bp of the R1 and F2 primers overlap (Supplemental Table IA). PCR was performed using Phusion polymerase, and products were separated on a 1% agarose gel. Desired products were excised and extracted using the QIAquick gel extraction kit (Qiagen), and 0.2-μl of the gel extraction products for reaction 1 and reaction 2 were used as template in a third assembly reaction using the F1/R2 primers. Correct orientation and assembly of all test DNA fragments were confirmed by sequencing. 2b-2 fragment multimerization was accomplished by ligating 2b-2 digested with Nhel/Spel in the pLG(-)GFP2 2b-2 vector digested with Nhel. The resulting plgLG(-)GFP2 2b-2 vector was then digested with Nhel/Spel and re-ligated into a 2b-2 dimer, which was then ligated into a plgLG(-)GFP2 2b-2 vector digested with Nhel.

**Cell culture**

Cells were cultured in RPMI 1640 with 10% FBS, 1% chicken serum, 2 mM L-glutamine, 0.1 mM 2-mercaptoethanol, and penicillin/streptomycin at 41°C with 5% CO2. Transfections were performed by electroporating 107 cells with 40-μg linearized plasmid DNA at 25 μF and 700 V (Bio-Rad Gene Pulser). Stable transfecteds were selected with 10–15 μg/ml blasticidin (Sigma) for 6–9 d and subsequently screened by duplicate plating in 1 μg/ml puromycin (Sigma) to confirm deletion of the puromycin resistance gene within the rearranged IgL locus. Genotypes of clones with targeted integration were confirmed by PCR of the targeting construct and VIIB intervening sequence of the un-rearranged locus as previously described (34). The IgL(-)GFP2 (ΔV IgL(-)GFP2), IgL(-)GFP2 W (ΔV IgL(-)GFP2), and IgL(+)/GFP2 AID- (ΔV IgL(-)GFP2), and ΔV IgL puro+ DT40 cells have been described previously (34).

**Flow cytometry**

GFP levels for each targeting construct were assessed by FACS (FACS-Calibur; BD Biosciences) for at least two independent targeted transfectants. These primary targeted transfectants were subcloned by limiting dilution, and at least 12 subclones for each primary transfectant were grown in cell culture on day 14 post-transfection (with the exclusion of subclones Figs. 2B and 3, which were analyzed on day 21 post-subcloning). At least 100,000 events were collected, and the gate for GFP low/negative cells was drawn one log below the mean fluorescence intensity for the main population of green fluorescent cells (FlowJo software). Subclones where more than 50% of live cell events fell into the GFP low or negative gate were excluded from analysis due to the possibility of expansion of a precursor cell expressing mutated GFP at the time of subcloning (34).

**Chromatin immunoprecipitation**

Abs used for chromatin immunoprecipitation (ChIP) analysis were H3K4me3 (04-745; Millipore), H4ac (06-866; Millipore), H3K9ac (06-942; Millipore), H3K9me3 (ab8898; Abcam), Ub-H2B (05-1312; Millipore), H3K79me (ab3594; Abcam), Pol II (sc-899x, Santa Cruz Biotechnology), RNA Pol II CTD Phospho-S2 (ab5095; Abcam), RNA Pol II CTD Phospho-S5 (ab5131; Abcam), CDK9 (sc-8338; Santa Cruz Biotechnology), and normal rabbit IgG (12-370; Millipore). A polyclonal rabbit anti-Spt5 Ab (39) was a generous gift from Dr. Yuko Yamaguchi (Tokyo Institute of Technology, Tokyo, Japan).

ChIP was performed as described previously (40), with modifications. Briefly, cells were cross-linked with 1% formaldehyde, quenched with 0.125 M glycine, then washed and resuspended in RIPPA buffer [10 mM Tris-HCl (pH 7.4), 1 mM EDTA, 0.1% SDS containing 0.5 M NaCl]. Sonication was performed using a water bath sonicator (Diagenode), precleared with Protein G Dynabeads (Invitrogen), and chromatin from roughly 15 × 106 cells was incubated with specific Ab or normal rabbit IgG overnight at 4°C (an aliquot of 10% of the chromatin volume of an immunoprecipitate (IP) was set aside as the input sample). Immune complexes were isolated after a 3-h incubation at 4°C with Protein G Dynabeads and extensive washing [twice for 10 min with each of the following buffers: RIPPA, 0.125 M NaCl RIPPA, 0.5 M NaCl RIPPA, LiCl buffer (0.25M LiCl, 0.5% Nonidet P-40, 0.5% sodium deoxycholate), and TE (10 mM Tris-Cl, pH 7.5, 1 mM EDTA)]. After cross-link reversal and RNA purification, duplicate TaqMan quantitative PCR (qPCR) reactions were performed with PerfeCta qPCR supermix with Ung (Quanta Biosciences) using company-specified cycling parameters with a 95°C 5-min initial denaturation step) and 30 cycles of 95°C for 30 s, 60°C for 30 s, and 72°C for 1 min. qPCR reactions were performed with PerfeCta qPCR supermix with Ung (Quanta Biosciences) using company-specified cycling parameters with a 95°C 5-min initial denaturation step) and 30 cycles of 95°C for 30 s, 60°C for 30 s, and 72°C for 1 min. qPCR reactions were performed with PerfeCta qPCR supermix with Ung (Quanta Biosciences) using company-specified cycling parameters with a 95°C 5-min initial denaturation step) and 30 cycles of 95°C for 30 s, 60°C for 30 s, and 72°C for 1 min.
by dividing the IP/Inputcorr for a test PCR region by the IP/Inputcorr calculated for γ-actin with the same Ab. Two-tailed unpaired t tests were used in Fig. 8 to compare data for IgL(−)GFP2 versus IgL(−)GFP2 W and IgL(+)GFP2 AID−/− at PCR regions a–f and to compare data for region e versus region f using the combined data for IgL(−)GFP2 W and IgL(+)GFP2 AID−/−.

RT-PCR analysis

Total RNA was isolated from 5 × 10⁶ DT40 cells using the RNeasy Mini Kit (Qiagen). RNA concentration was measured using a NanoDrop spectrophotometer (Thermo Scientific), and 1 µg RNA was treated with RNase-free DNase I (Invitrogen) and reverse transcribed using Superscript II (Invitrogen) primed with random hexamers. TaqMan qPCR reactions were performed in duplicate using PerfeCTa qPCR supermix with Ung (Quanta Biosciences) and company-specified cycling parameters with 50 ng of first-strand product (primers and probes, Supplemental Table IB). The starting quantity of the initial cDNA sample was calculated from primer-specific standard curves with the 3000 XP thermocycler (Stratagene) data analysis software. cDNA values were normalized by dividing the calculated amount of input cDNA for GFP by the calculated input cDNA of 18S rRNA in each sample.

Results

DIVAC assay

To assess DNA sequences for DIVAC function, we took advantage of a previously developed assay in which a heterologous expression cassette, flanked by the DNA sequence of interest, is inserted into the IgL locus of DT40 cells (34). The expression cassette, termed GFP2, consists of the strong, enhancer-independent Rous sarcoma virus promoter driving expression of GFP, followed by an internal ribosome entry site, the blasticidin resistance gene, and an SV40 virus polyadenylation signal (Fig. 1A). To perform the assay, a targeting vector containing GFP2, the test fragment, and appropriate homology arms, is transfected into the ψV⁺ IgL⁻ puro⁺ DT40 cell line, in which the entirety of the chicken IgL locus has been deleted and replaced by a puromycin resistance gene (Fig. 1A), and which constitutively expresses AID (34). Homologous recombinants are identified as blasticidin-resistant, puromycin-sensitive clones, with proper insertion into the recombined IgL allele confirmed by PCR. At least two independent primary transfecants for each targeting vector were subcloned, and GFP expression was assessed by flow cytometry in multiple subclones 14–21 d later. The frequency of cells with decreased GFP fluorescence (hereafter, “GFP loss”) provides a sensitive measure of the mutation frequency of GFP (34).

We confirmed that cells containing GFP2 without any flanking IgL sequences (cell line IgL(−)GFP2) exhibit very little GFP loss, with median frequencies <0.1% 21 d after subcloning (Figs. 1B, 2B) and <0.06% 14 d after subcloning (Fig. 2D). The data presented in this study were derived from cells cultured for 14 d after subcloning except data of Figs. 1B, 2B, and 3B, which

![FIGURE 1. The DT40 DIVAC experimental system. (A) Top, Schematic diagram of the ψV⁺ IgL⁺ locus. Exons for IgL [leader (L), VJ, and constant (C)] and the surrounding genes VpreB and CAXV (carbonic anhydrase XV; predicted) are depicted as black boxes. The IgL enhancer and chicken repeat region 1 (CR1) are shown as an open circle and open square, respectively. The location of the W fragment (34) (GenBank accession number FJ482234) is depicted with a line below the main diagram. The position and size of each of these elements are drawn to scale. Bottom, Map of the "empty" IgL⁺ locus in the ψV⁺ IgL⁻ puro⁺ cell line, targeting construct including the GFP2 reporter, and IgL⁺ locus after targeted insertion of the GFP2 reporter are shown. (B) FACS analysis of representative primary transfecants (21 d of culture) for the indicated targeting constructs. (C) Transcript levels of the GFP coding region assayed by TaqMan quantitative RT-PCR. Bars show average signals for GFP transcripts from two independently derived RNA samples from each cell line, with SEM indicated. Transcript levels were normalized to the 18S rRNA signal for each respective sample.](http://www.jimmunol.org/Downloadedfrom)
derive from 21-d cultured cells. We observed a reproducible ∼2-
fold increase in median GFP loss frequencies between days 14 and
21 of culture (data for 2b-2 in Figs. 2B and 4B, and data not
shown). If GFP2 is flanked by the 9.8-kb “W” fragment (spanning
IgL sequences downstream of the VL promoter; Fig. 1A) or by
a subfragment of W with strong DIV AC function (e.g., the first
panel of Fig. 1B), median GFP loss is increased at least 100-fold
(Fig. 2D). In the absence of AID, GFP loss is reduced to extremely
low levels (0.005% at 14 d of culture) even when GFP2 is
flanked by the entirety of the W fragment [IgL(+)/GFP2 AID
2/2 in
Fig. 1B and Fig. 2B, 2D]. Our data, combined with that reported
previously (34), demonstrate that in this assay system, mutation of
GFP and GFP loss are strongly dependent on DIVAC sequences
and AID.

Importantly, the effects of DIVAC and AID on GFP loss and
GFP mutation are not due to changes in levels of GFP transcrip-
tion. GFP fluorescence levels of the main GFP+ population
do not vary substantially depending on the presence or absence
of a test fragment flanking GFP2, the length or nature of the test
fragment, or the presence or absence of AID (Ref. 34, Fig. 1B, and
data not shown). Furthermore, direct measurement of GFP tran-
script levels showed no differences between IgL(−)/GFP2 cells (no
test fragment), IgL(−)/GFP2 W cells (W as a test fragment), and
IgL(+)/GFP2 AID−/− cells (W sequences flanking GFP2 but no
AID) (Fig. 1C, first three bars). GFP transcript levels showed at
most small variations in the other transfectants examined (Fig.
1C), and these variations did not correlate with DIVAC function
or with the particular test fragment they contained. Finally,
GFP transcript levels were previously shown to be equivalent between
sorted GFP-high and GFP-low populations, demonstrating that
GFP loss is not due to downregulation of GFP transcription (34).

Strong DIVAC function arises from synergy between discrete
DNA elements

We focused our analysis on a 4.4-kb region starting downstream
of the IgL C region exon and terminating within a CR1 retro-
transposon element (Fig. 2A) based on a previous low-resolution
deletion analysis of the W fragment (34). The 4.4-kb region was
divided into three 1.5- to 1.6-kb fragments with 100-bp overlaps
at their ends (fragments 1, 2, and 3; Fig. 2A), which were indi-
vidually tested for DIVAC function. Fragment 1 was devoid of
activity, whereas fragments 2 and 3 were active (Fig. 2B) but at
levels well below that seen for W (Fig. 2D). Strikingly, either the
entire 4.4-kb region (DIVAC 1-3) or a combination of fragments 2

FIGURE 2. Localization of DIVAC activity within
the W fragment. (A and C) Schematic diagram of the
region 3′ of the chicken IgL locus. The C region exon,
IgL enhancer, and CR1 element are shown. The position
and length of each DNA fragment tested are
specified below the locus diagram, with dotted lines
indicating missing sequence. Position and size of the
subfragments are drawn to scale. Note that fragment 1-
3 was discovered to contain a small (∼100 bp) duplica-
tion of sequences from the 5′ end of fragment 2,
inserted at the position indicated in (C). (B and D)
Fluctuation analysis of GFP loss in subclones. Each
circle represents percent GFP loss for one subclone and
the median GFP loss value of all subclones from the
same primary transfectant is indicated with a horizontal
bar and stated numerically at the top of the graph. The
name of the test fragment adjacent to the GFP2 cas-
ette, from which each cell line derives its name, is
indicated at the bottom of the graphs. Data in (B) and
(D) derive from subclones cultured for 21 and 14 d,
respectively. Note that the y-axis for (D) and all sub-
sequent plots of percent GFP loss are depicted on a log
scale. In (D), the endpoint for the deletion in 2-3ΔN
corresponds to the 5′ end of fragment “N” (34), and the
deletion removes a region that increased GFP loss ac-
tivity in that study.

4 CORE SHM TARGETING ELEMENTS
and 3 (DIVAC 2-3) yielded levels of GFP loss that were nearly equivalent to that supported by the entire W fragment (Fig. 2D) and were substantially higher than the sum of the activities of the individual fragments (note that the results for the individual fragments in Fig. 2B are from day 21 whereas those for the combined fragments in Fig. 2D are from day 14). These results indicate that sequences in fragments 2 and 3 synergize with one another to yield strong DIVAC function.

To confirm that the high levels of GFP loss driven by DIVAC 2-3 were accompanied by frequent point mutations in GFP, we cultured IgL(2)GFP2 2-3 and IgL(2)GFP2 cell lines for 21 d and sequenced a 1217-bp region spanning GFP. As expected, IgL(2)GFP2 cells (which contain no DIVAC element) showed little evidence for mutation (5.7 × 10^{-2} mutations/bp), whereas IgL(2)GFP2 2-3 cells exhibited a mutation frequency of 1.25 × 10^{-2} mut/bp, similar to the frequency of 1.24 × 10^{-3} mut/bp reported previously for IgL(+)GFP2 cells (which contain all W sequences) cultured for 6 wk (34) (Supplemental Fig. 1). The mutation spectrum in IgL(2)GFP2 2-3 cells was similar to that of IgL(+)GFP2 cells, with almost all mutations occurring at G and C residues and C mutations being biased toward C→G transversions. DIVAC 2-3 supported an increased frequency of G→A mutations compared with W (Supplemental Fig. 1), which might be due to the smaller sample size in our analysis and differences in the size of the region sequenced. We conclude that DIVAC 2-3 has activity roughly comparable to that of W as assessed by GFP loss and point mutation of GFP.

**Deletion analysis to identify minimal DIVAC sequences**

We then performed deletion experiments to localize DIVAC function in fragments 2 and 3. Fragment 2 was divided into three overlapping fragments (2a, 2b, and 2c), and essentially all GFP loss activity was found associated with 2b (Fig. 2B). This suggests that fragments 2a and 2c lack functionally important elements, a conclusion further supported by the finding that deletion 2a, 2c, or 2a and 2c had no effect on the activity of DIVAC 2-3 (samples 2-3Δ2a, 2-3Δ2c, and 2-3Δ2a/2c, Fig. 2D). Subdivision of 2b into fragments 2b-1, 2b-2, and 2b-3 revealed that most activity resided in 2b-2, which, like fragment 2b, was as active as the entirety of fragment 2 (Fig. 2B). Fragment 2b-2 is therefore a small (316 bp) region with detectable DIVAC function (33), although it contains only ~10% as much activity as DIVAC 2-3.

When a series of six ~50-bp deletions were made through the entirety of the 2b-2 element (Fig. 4A), GFP loss activity was found to be most reduced in the Δ101–150 and Δ202–254 deletions (Fig. 4B). A 154-bp fragment spanning these two deletion intervals and the intervening DNA (2b-2 154) was as active as 2b-2 (Fig. 4B) and therefore contains elements with DIVAC function. This 154-bp sequence contains an E-box motif (CAGCTG), an NF-κB binding site, and a binding site for the transcription factors PU.1 and IRF4 termed the Ets–IRF composite element (EICE) (Fig. 4A). This NF-κB site (35) and EICE element (37) have previously been suggested to contribute to DIVAC function, and considerable data link E boxes to SHM (see Introduction). Further analysis of the role of these individual motifs was difficult because of the low level of GFP loss associated with the 2b-2 fragment. We are currently developing a more sensitive DIVAC assay with which to identify functionally important motifs in DNA fragments with low activity. Notably, a 4× multimer of the 2b-2 fragment exhibited an average of ~1.5% GFP loss, more than 3-fold higher than the 0.4–0.45% GFP loss seen with a single copy of 2b-2 (Fig. 4B). This suggests that DIVAC function can be modulated in an
additive fashion, with the number of active DNA motifs, not just
the presence of the motifs, being important.

Because much of 2b-2 lies within the chicken IgL locus, we
tested mouse Igk locus enhancers (iE<sub>k</sub>,<sup>3</sup>E<sub>k</sub>) and mouse Igλ
locus enhancers (E<sub>λ</sub>2-4, E<sub>λ</sub>3-1) for DIV AC function. These
enhancers supported low levels of GFP loss, ranging from slightly
above background to similar to that of 2b-2 in the case of 3E<sub>k</sub>
(Supplemental Fig. 2). This argues that the murine enhancers are
not strong DIV AC elements in DT40 cells by themselves, but
leaves open the possibility that they can collaborate with other
sequences to target SHM to Ig L chain genes.

Fragment 3 was also subjected to deletion analysis to identify
regions important for DIVAC function. Analysis of three over-
lapping fragments spanning all of fragment 3 except CR1
sequences (3a, 3b, and 3c; Fig. 3A) revealed that only fragment 3c
exhibited substantial activity, but at levels well below the entirety
of fragment 3 (Fig. 3B). Deletion of 3a (3D<sub>a</sub>) or 3b (3D<sub>b</sub>) from
fragment 3 yielded regions with slightly more activity than 3c,
whereas deletion of 3c and CR1 sequences (3a+b) resulted in
loss of most activity. (Fig. 3B). Adding CR1 sequences to 3c (3c3) did
not enhance activity, and a fragment spanning all of CR1 (PG) did
not have substantial activity (Fig. 3B). Together, these results in-
dicate that functionally relevant DNA sequences are dispersed
in fragment 3, with fragment 3c containing particularly important
DNA elements that require other sequences, likely within both
fragments 3a and 3b, to attain the full DIVAC function found in
fragment 3. The idea that functionally important elements are
dispersed in fragment 3 is supported by deletion analyses carried
out in the context of DIVAC 2-3, where deletion of most of
fragment 3 (2-3ΔN), or the 3′ portion of fragment 3 (2-3Δ3′N), or
5′ sequences of fragment 3 (2-3Δ5′N), resulted in a substantial
drop in GFP loss, whereas deletion of only the CR1 sequences
(2-3ΔCR1) did not (Fig. 2D).

**Broad dispersion of the elements contributing to DIVAC function**

The data described earlier indicate that DIVAC 2-3 requires
sequences in both fragments 2 and 3 for strong activity, that the
activity of fragment 2 resides within the central part of fragment 2b,
and that sequences from regions 3a, 3b, and 3c are important for the
DIVAC function of fragment 3. In an attempt to create a smaller
fragment that retained strong DIVAC function, the unique sequences
of fragment 2b (those not contained in the overlap with 2a or 2c)
were combined with most of fragment 3 (lacking the overlap with
fragment 2 as well as CR1 sequences) to create DIVAC 1703 (Fig.
5A, Supplemental Fig. 3A). In Figs. 5A and 6A, certain transcrip-
tion factor recognition motifs are depicted to serve as landmarks.
This includes E boxes, which are broken down into those conforming
to the motif CASSTG (S = G or C), which has been demonstrated
to be the preferential binding site for E<sub>2a</sub>-encoded proteins (41),
and those that do not conform to this motif.

**FIGURE 4.** Deletion and mutation analysis of
fragment 2b-2. (A) Schematic diagram of the location
of fragment 2b-2 in the region 3′ of the chicken IgL
locus. The 2b-2 fragment is enlarged in the lower part,
and the structures of the various deletion mutants
analyzed are indicated, with dotted lines indicating
missing sequence. The positions of DNA motifs of
interest are indicated with symbols, as indicated in the
key to the right. (B) Fluctuation analysis of GFP loss
in subclones with data depicted as in Fig. 2D.
suggests that this portion of 3c contains important DNA sequences. However, decreased activity was also seen with D\textsubscript{1001–1200} (removing part of 3b) and D\textsubscript{1401–1703} (removing the 3\textsuperscript{9} part of 3c), again supporting the idea that multiple elements in fragment 3 contribute to DIVAC function. The D\textsubscript{401–600} deletion also reduced activity somewhat. The strong loss of activity seen with D\textsubscript{201–400} and D\textsubscript{1201–1400} further supports the idea that sequences from fragments 2 and 3 work together synergistically to constitute DIVAC function. However, the data argue that DNA motifs outside of D\textsubscript{201–400} and D\textsubscript{1201–1400} also contribute to DIVAC function.

In an attempt to create a fragment smaller than DIVAC 2-3 that retained its full activity, we joined all of fragment 2b (including the overlap with 2a and 2c) to fragment 3 (lacking CR1 sequences) to create DIVAC 1928 (Fig. 6A, Supplemental Fig. 3B). DIVAC 1928 (4.3–6.7% GFP loss; Fig. 6B) was found to be as active as DIVAC 2-3 (4.0–6.9% GFP loss; Fig. 2D). In an attempt to identify with greater resolution the DNA motifs responsible for DIVAC function, a series of eighteen 50-bp deletions were made across DIVAC 1928 (Fig. 6A), focusing on the regions suggested by the data of Fig. 5B to be most important for function. Remarkably, almost every 50-bp deletion caused at least a moderate drop in the frequency of GFP loss (Fig. 6B), suggesting that functionally relevant sequences are distributed broadly in DIVAC 1928. Two regions were identified in which deletions caused the greatest loss.

**FIGURE 5.** Deletion analysis of DIVAC 1703. (A) Schematic diagram of the location of the DIVAC 1703 fragment in the region 3\textsuperscript{9} of the IgL locus. The DIVAC 1703 fragment is enlarged in the lower part to illustrate the endpoints of deletions (black vertical lines) and relevant DNA motifs (see key; S = G or C). Dotted lines indicate missing sequence. (B) Fluctuation analysis of GFP loss in subclones with data depicted as in Fig. 2D. The cell lines and test fragments are named for the roughly 200-bp segment that was deleted from DIVAC 1703.

**FIGURE 6.** Deletion analysis of DIVAC 1928. (A) Schematic diagram of the DIVAC 1928 element with positions of DNA motifs indicated. Each bar below the DIVAC 1928 line (Δ1–Δ18) shows a 50-bp DNA fragment and associated DNA motifs deleted from the DIVAC 1928 element. (B) Fluctuation analysis of GFP loss in subclones with data depicted as in Fig. 2D. The cell lines and test fragments are named for the 50-bp segment that was deleted from DIVAC 1928.
of DIVAC function: Δ1–Δ3 and Δ10–Δ16. Δ1–Δ3 corresponds closely to 2b-2 154 (Fig. 4A) and lies within the Δ201–400 interval (Fig. 5A), while Δ10–Δ16 lies within fragment 3c (Fig. 3A) and the Δ1201–1400 and Δ1401–1703 intervals (Fig. 5A). On the basis of these data, we defined functionally important “core” regions of fragments 2 and 3 as the F2 core (Δ201–400) and the F3 core (350 bp spanned by Δ10–Δ16).

The analysis of DIVAC 1703 indicated that sequences from 1–200 and 601–1000 contributed little to DIVAC function in this context (Fig. 5B), and the analysis of DIVAC 1928 indicated that Δ17 and Δ18 did not reduce activity substantially (Fig. 6B). However, deletion of all of these apparently unimportant sequences from DIVAC 1928 reduced activity substantially, to ∼2.1–2.7% GFP loss (data not shown). We have thus far not been able to identify a DNA fragment substantially smaller than DIVAC 1928 that retains high levels of activity.

**A critical role for the F2 and F3 core regions**

To determine the extent to which the F2 and F3 core regions contribute to the activity of large, highly active DIVAC elements, they were deleted either individually or in combination from DIVAC 1-3 and DIVAC 2-3 (Fig. 7A). In both contexts, deletion of either core alone reduced activity, and deletion of both cores resulted in a nearly complete loss of activity (Fig. 7B). Similarly, deletion of the F2 core or a critical portion of the F3 core (the 1201–1400 region) from DIVAC 1928 reduced the frequency of GFP loss (the double core deletion was not tested in this context) (Fig. 7A, 7B). In all three contexts, deletion of the F3 core caused a greater loss of activity than deletion of the F2 core. These results demonstrate that when the core regions are removed, the remaining sequences (nearly 3.9 kb in the case of DIVAC 1-3) have no ability to support AID-mediated sequence diversification, and hence that the core regions play essential roles in DIVAC function. To determine whether the core regions are sufficient for strong DIVAC function in a smaller fragment, they were combined together with some flanking DNA sequences to yield DIVAC 751 (Fig. 7A). DIVAC 751 (2.6–3.0% GFP loss) was approximately half as active as DIVAC 2-3 (Fig. 7B). We conclude that sequences outside of the core regions contribute to DIVAC function.

**DIVAC and epigenetic modifications**

To begin to explore the mechanism by which DIVAC functions, we tested the hypothesis that DIVAC alters levels of epigenetic modifications within the mutation cassette, taking advantage of the availability of cell lines harboring identical GFP2 transcription units either flanked or not flanked by W sequences. ChIP experiments were performed using three cell lines: IgL(−)GFP2, IgL(−)GFP2 W (W sequences flanking GFP2), and IgL(+)GFP2 AID−/− (W sequences flanking GFP2, lacking AID). Six different regions of the locus were analyzed (Fig. 8A), four in the GFP2 cassette and two within the F2 or F3 cores. [Note that in IgL(−)GFP2 cells, W sequences are assessed

---

**Figure 7.** Analysis of F2 core and F3 core DIVAC elements. (A) Schematic diagram of the position of the F2 core and F3 core regions in the region 3′ of the IgL locus (drawn to scale). The core elements were either singly (ΔF2 Core or ΔF3 Core) or doubly (ΔF2/F3 Core) deleted from DIVAC 1-3 or DIVAC 2-3 and individually deleted from DIVAC 1928 (the equivalent of the DIVAC 1703 Δ1201–1400 deletion was performed for DIVAC 1928 in place of the ΔF3 core deletion). Deleted DIVAC core regions are depicted with gray shading. The Δ1201–1400 equivalent deletion is shown below DIVAC 1928. The endpoints of DIVAC 751, which is made up primarily of the two core elements, is also shown. (B) Fluctuation analysis of GFP loss in subclones with data depicted as in Fig. 2D. The DIVAC 1-3, 2-3 and 1928 data sets are the same as those in Figs. 2D and 6B, respectively, and are shown for ease of comparison. (C) Schematic diagram of the 3′ portion of the IgL locus, drawn approximately to scale, indicating the location of fragments 1, 2, and 3, the F2 and F3 cores, region A from Refs. 35, 37, and the 5′ mutational enhancer element (5′ MEE) and 3′ mutational enhancer element (3′ MEE) from Ref. 38. The 222-bp element identified as functionally important in Ref. 38 lies within the 3′MEE. Dashed lines are used to indicate that the boundaries of the MEEs have not been defined.
only on the un-rearranged IgL allele, whereas in IgL(−) GFP2 W and IgL(+)GFP2 AID−/− cells, W sequences are present on both the GFP2 knock-in allele and the non-rearranged IgL allele. We analyzed six histone modifications—H3K4me3, acetylated H4 (H4ac), acetylated H3 lysine 9 (H3K9ac), H3 lysine 9 trimethylation (H3K9me3), monoubiquitinated H2B (Ub-H2B), and H3 lysine 79 dimethylation (H3K79me2)—based on previous data linking them to AID-mediated diversification processes and transcription (29, 42–49). Values were normalized to the signal obtained from the γ-actin gene, a well expressed gene whose
modifications should be similar in the three cell lines. Statistical analyses (see Materials and Methods) were used to compare signals from IgL(+)GFP2 cells to both the IgL(−)GFP2 W and IgL(+)GFP2 AID−/− cells and to compare the signals obtained at the two core regions for cell lines containing W sequences adjacent to GFP2.

H3K4me3, H4ac, and H3K9ac are marks of open, transcriptionally active chromatin that have been suggested to enhance AID-mediated diversification processes (29, 42–45). We did not observe significant differences in the levels of these marks at any region of GFP2 between IgL(−)GFP2 cells (white bars) and IgL(+)GFP2 W (black bars) or IgL(+)GFP2 AID−/− cells (hashed bars) (Fig. 8B, 8C, 8D). Levels of H3K4me3 were, however, higher at the F3 core than the F2 core (Fig. 8B), as were levels of H4ac (Fig. 8C). H3K9me3, a repressive mark generally found in heterochromatin, is found in S regions undergoing CSR (46) where it is thought to facilitate the recruitment of AID (47). No DIVAC-dependent differences in levels of H3K9me3 were detected in GFP2, and this mark was virtually undetectable at the F2 and F3 core regions (Fig. 8E). Ub-H2B is a consequence of transcriptional activation (50) and affects the movement of Pol II through a coding region by directing histone H2A/H2B exchange and chromatin condensation (51). Ub-H2B and Ub-H2A were previously found associated with AID-targeted loci in a human cell line and with S regions in stimulated primary mouse B cells (48). Low levels of Ub-H2B were detected at the promoter and GFP coding region whereas higher levels were seen at the IRES and Bsr segments of GFP2; however, none of the signals exhibited a dependence on the presence of DIVAC (Fig. 8F). H3K79me2 marks intragenic regions of actively transcribed genes and is associated with transcription elongation (49). Relatively low and comparable levels of H3K79me2 were seen at the promoter and GFP coding region in the three cell lines, with higher levels detected at the downstream elements of the GFP2 cassette, particularly IRES (Fig. 8G). In addition, IRES displayed significantly higher levels of H3K79me2 in the absence of W compared with when W was present (Fig. 8G), raising the possibility that transcription elongation within IRES is altered by W sequences. It is unclear what feature of the IRES element dictates its particularly high levels of Ub-H2B and H3K79me2.

In summary, the analysis of six different histone marks failed to reveal any DIVAC-dependent differences at the promoter or at the GFP gene, where DIVAC-dependent mutagenesis occurs.

**DIVAC and the transcription apparatus**

Because of the tight link between transcription and the action of AID, we used ChIP to investigate the hypothesis that DIVAC mediates changes in Pol II recruitment, promoter clearance, or transcriptional elongation. The C-terminal domain (CTD) of Pol II is hypophosphorylated upon initial recruitment to a promoter and undergoes phosphorylation at serine 5 (S5P) during promoter clearance by general transcription factors (52). S5P Pol II complexes accumulate ~40 nt downstream of the transcription start site, partially due to association with the negative-acting elongation factor complex (NELF) and the DRB-sensitivity inducing complex (DSIF; composed of Spt4 and Spt5) (53). The release of Pol II pausing occurs after a second phosphorylation event on serine 2 (S2P) of the CTD that is mediated by P-TEFb, a complex containing the kinase Cdk9 (54). P-TEFb also phosphorylates DSIF, and the combination of S2P Pol II and phosphorylated DSIF causes the dissociation of NELF and activation of elongation.

We first used an Ab to the N-terminal portion of Pol II to measure total Pol II levels at the GFP2 cassette and the F2 and F3 core regions. Total Pol II levels were highest in GFP and did not vary significantly between the cell lines containing (black or shaded bars) or lacking DIVAC (white bars) at any of the regions tested (Fig. 8H), consistent with the similar levels of GFP transcripts detected in these lines (Fig. 1C). We also assessed levels of S5P and S2P Pol II and found that the ChIP signals for these two modifications differed significantly between the cell lines only within the GFP coding region, where S5P Pol II was elevated ~2-fold in the presence of DIVAC compared with its absence, and S2P Pol II was also increased in a DIVAC-dependent manner, although only reaching statistical significance in one of the two DIVAC-containing cell lines (Fig. 8I, 8J). The DSIF component Spt5 was recently shown to facilitate interactions between AID and Pol II at sites of Pol II stalling (8), making it of particular interest in relation to the mechanism by which DIVAC recruits SHM. There were, however, no significant DIVAC-dependent differences in Spt5 levels in any of the regions analyzed (Fig. 8K).

The P-TEFb component Cdk9 was also analyzed, revealing a consistent trend of higher levels of Cdk9 in the presence of DIVAC compared with its absence, although none of the individual differences were statistically significant (Fig. 8L).

Overall, these experiments indicate that in the presence of DIVAC, the GFP mutation target region displays increased levels of the stalled (S5P) form of Pol II but not of the Pol II-associated factor Spt5. They also raise the possibility that the presence of DIVAC is associated with an increase in the pause release kinase Cdk9 and its product S2P Pol II.

**Discussion**

**Properties of DIVAC**

We have performed an extensive deletion analysis of DIVAC, a region of the chicken IgL locus previously implicated in supporting efficient AID-mediated mutagenesis of a flanking GFP transcription unit (34). The findings reported in this study reveal a number of novel features of DIVAC. First, the nearly 10-kb W fragment can be reduced in size to less than 2 kb (DIVAC 1928) with only modest loss of activity. Second, sequence elements that contribute to DIVAC function are not found in one small region but instead are widely dispersed. This is most clearly illustrated by our observation that many different 50-bp deletions in DIVAC 1928 cause a measurable drop in activity (Fig. 6) but is also apparent from deletion analyses in the context of relatively small fragments (Figs. 2–4) and another large fragment (Fig. 5). Third, despite the dispersal of functional elements, two relatively small regions of particular functional significance exist: the 200-bp F3 core spanning the 5’ half of the canonical IgL enhancer, and the 350-bp F3 core that lies downstream of the enhancer, ~1.2 kb away. In several different contexts, deletions affecting the cores are more detrimental to activity than comparably sized deletions in other regions. Strikingly, whereas removal of one of the two cores usually has only a modest effect on activity, deletion of both cores, even in the context of quite large fragments, virtually eliminates GFP loss activity. These observations indicate that whereas other sequences in the fragments analyzed in this study can partially compensate for loss of one core, they cannot compensate for the loss of both. We cannot rule out the possibility that sequences outside of the 4.4-kb DIVAC 1-3 region could compensate for deletion of both cores. It is also apparent that the consequences of deletion of the F2 core are context dependent, causing a relatively mild (2- to 3-fold) drop in activity in the context of DIVAC 1-3, DIVAC 2-3, and DIVAC 1928 (Fig. 7B) but a severe (30-fold) decrease in the context of DIVAC 1703 (Fig. 5B). This indicates that the 225 bp present in DIVAC 1928 but missing from DIVAC 1703 play an important role in compensating for the loss of the F2 core. The existence of sequences with re-
Several previous studies have searched the chicken IgL locus for DNA elements that is disrupted by deletions between them. A recent study by Kothapalli et al. (38) identified a 222-bp region downstream of the IgL enhancer that substantially increased the frequency of GCV in a stop codon reversion assay (although it had only a marginal effect on mutation frequency as assessed by DNA sequencing). This 222-bp region, located within the 3′MEE (38) (Fig. 7C), lies between the F2 and F3 core elements and spans the junction between the ∆601–800 and ∆801–1000 deletion intervals, neither of which had a substantial effect on DIVAC function (Fig. 5B). Hence, we do not detect a major, nonredundant role for the 222-bp region in our assay. Nor does our data lead us to assign substantial DIVAC function to DNA sequences in the area designated as the 5′MEE (38) (Fig. 7C). While the MEEs do not correspond to either the F2 core or the F3 core, the precise degree of overlap remains to be determined since the boundaries of the MEEs have not been mapped.

Methodological differences might help explain the apparent discrepancy between the sequences that we assign as the DIVAC cores and those assigned as the MEEs. One important difference is that Kothapalli et al. (38), as well as several other relevant studies in DT40 cells (35–37), used a transcription unit driven by the IgL promoter. This promoter is strongly dependent on distal sequences for its transcriptional activity (33), whereas the GFP2 cassette used in this study and previously (34) does not require other flanking elements versus effects due to loss of the deleted sequences themselves. Given the numerous different contexts in which deletion of some or all of the F2 core or the F3 core causes a substantial drop in activity, it is unlikely that the cores function primarily as spacers; rather, it is likely that the core sequences are themselves important for DIVAC function. This is supported by our finding that DIVAC 751, which contains little besides the cores, has substantial activity. This result with DIVAC 751 also suggests that the natural spacing between the two cores is not essential for their ability to cooperate to some extent with one another. Other data, however, are consistent with the notion that a minimal intercore spacing is important for maximal activity. Almost any manipulation that reduces the intercore distance below that seen in DIVAC 1928 results in a drop in activity, from ~5% GFP loss in DIVAC 1928 to 2.5–3.5% GFP loss in ∆5–∆9 (Fig. 6B), DIVAC 1703 (Fig. 5B), and DIVAC 751 (Fig. 7B). This raises the possibility that DIVAC 1928 contains a favorable, minimal spacing between the core regions that is disrupted by deletions between them.

Comparisons with other studies of the DIVAC region

Several previous studies have searched the chicken IgL locus for sequences capable of directing AID-mediated sequence diversification (26, 33–38). The results of these previous studies are in general agreement with those reported in this study in localizing DIVAC function to the region 3′ of the IgL C region exon. In one set of studies, the most important region was found to span the enhancer (“Region A” in Refs. 35, 37). Another study identified an important 3′ regulatory region consisting of the 4 kb 3′ of the enhancer (33), and a follow-up analysis identified mutation enhancer elements (MEEs) both 5′ and 3′ of the enhancer (5′MEE and 3′MEE) (38). The relationship between these regions and the sequences identified as important in our experiments is summarized in Fig. 7C. Our finding that DIVAC is dispersed over many sequences helps explain the failure of previous studies to converge on a single functional element.

Previous studies implicated NF-κB subunits, IRF4, and the NF-κB and EICE binding sites located in the IgL enhancer as important for GCV in DT40 cells (35, 37). The results of our deletion analysis of 2b-2 (Fig. 4) and the location of the NF-κB and EICE sites within the F2 core are consistent with the idea that these binding sites are relevant to DIVAC function. Numerous studies have implicated the E box sequence CANNTG and the E box binding proteins encoded by the E2a locus in SHM/GCV, both in mouse germinal center B cells and chicken DT40 cells (19, 27–31). Our data are consistent with this but also suggest that the relationship between E boxes and DIVAC function is not a simple one. The ∆401–600 and ∆1201–1400 deletions each remove five E boxes from DIVAC 1703, but the latter reduces GFP loss activity almost 10-fold, whereas the former has less than a 2-fold effect (Fig. 5); and in the context of DIVAC 1928, 50-bp deletions that remove one or more E boxes typically reduce activity, but two of the deletions with the strongest deleterious effects (∆2 and ∆13) do not contain E boxes (Fig. 6).

One limitation of our study is the difficulty in distinguishing between effects of a deletion caused by altering the spacing of flanking elements versus effects due to loss of the deleted sequences themselves. Given the numerous different contexts in which deletion of some or all of the F2 core or the F3 core causes a substantial drop in activity, it is unlikely that the cores function primarily as spacers; rather, it is likely that the core sequences are themselves important for DIVAC function. This is supported by our finding that DIVAC 751, which contains little besides the cores, has substantial activity. This result with DIVAC 751 also suggests that the natural spacing between the two cores is not essential for their ability to cooperate to some extent with one another. Other data, however, are consistent with the notion that a minimal intercore spacing is important for maximal activity. Almost any manipulation that reduces the intercore distance below that seen in DIVAC 1928 results in a drop in activity, from ~5% GFP loss in DIVAC 1928 to 2.5–3.5% GFP loss in ∆5–∆9 (Fig. 6B), DIVAC 1703 (Fig. 5B), and DIVAC 751 (Fig. 7B). This raises the possibility that DIVAC 1928 contains a favorable, minimal spacing between the core regions that is disrupted by deletions between them.

Comparisons with other studies of the DIVAC region

Several previous studies have searched the chicken IgL locus for sequences capable of directing AID-mediated sequence diversification (26, 33–38). The results of these previous studies are in general agreement with those reported in this study in localizing DIVAC function to the region 3′ of the IgL C region exon. In one set of studies, the most important region was found to span the enhancer (“Region A” in Refs. 35, 37). Another study identified an important 3′ regulatory region consisting of the 4 kb 3′ of the enhancer (33), and a follow-up analysis identified mutation enhancer elements (MEEs) both 5′ and 3′ of the enhancer (5′MEE and 3′MEE) (38). The relationship between these regions and the sequences identified as important in our experiments is summarized in Fig. 7C. Our finding that DIVAC is dispersed over many sequences helps explain the failure of previous studies to converge on a single functional element.

Previous studies implicated NF-κB subunits, IRF4, and the NF-κB and EICE binding sites located in the IgL enhancer as important for GCV in DT40 cells (35, 37). The results of our deletion analysis of 2b-2 (Fig. 4) and the location of the NF-κB and EICE sites within the F2 core are consistent with the idea that these binding sites are relevant to DIVAC function. Numerous studies have implicated the E box sequence CANNTG and the E box binding proteins encoded by the E2a locus in SHM/GCV, both in mouse germinal center B cells and chicken DT40 cells (19, 27–31). Our data are consistent with this but also suggest that the relationship between E boxes and DIVAC function is not a simple one. The ∆401–600 and ∆1201–1400 deletions each remove five E boxes from DIVAC 1703, but the latter reduces GFP loss activity almost 10-fold, whereas the former has less than a 2-fold effect (Fig. 5); and in the context of DIVAC 1928, 50-bp deletions that remove one or more E boxes typically reduce activity, but two of the deletions with the strongest deleterious effects (∆2 and ∆13) do not contain E boxes (Fig. 6).

A recent study by Kothapalli et al. (38) identified a 222-bp region downstream of the IgL enhancer that substantially increased the frequency of GCV in a stop codon reversion assay (although it had only a marginal effect on mutation frequency as assessed by DNA sequencing). This 222-bp region, located within the 3′MEE (38) (Fig. 7C), lies between the F2 and F3 core elements and spans the junction between the ∆601–800 and ∆801–1000 deletion intervals, neither of which had a substantial effect on DIVAC function (Fig. 5B). Hence, we do not detect a major, nonredundant role for the 222-bp region in our assay. Nor does our data lead us to assign substantial DIVAC function to DNA sequences in the area designated as the 5′MEE (38) (Fig. 7C). While the MEEs do not correspond to either the F2 core or the F3 core, the precise degree of overlap remains to be determined since the boundaries of the MEEs have not been mapped.

Methodological differences might help explain the apparent discrepancy between the sequences that we assign as the DIVAC cores and those assigned as the MEEs. One important difference is that Kothapalli et al. (38), as well as several other relevant studies in DT40 cells (35–37), used a transcription unit driven by the IgL promoter. This promoter is strongly dependent on distal sequences for its transcriptional activity (33), whereas the GFP2 cassette used in this study and previously (34) does not require other flanking elements versus effects due to loss of the deleted sequences themselves. Given the numerous different contexts in which deletion of some or all of the F2 core or the F3 core causes a substantial drop in activity, it is unlikely that the cores function primarily as spacers; rather, it is likely that the core sequences are themselves important for DIVAC function. This is supported by our finding that DIVAC 751, which contains little besides the cores, has substantial activity. This result with DIVAC 751 also suggests that the natural spacing between the two cores is not essential for their ability to cooperate to some extent with one another. Other data, however, are consistent with the notion that a minimal intercore spacing is important for maximal activity. Almost any manipulation that reduces the intercore distance below that seen in DIVAC 1928 results in a drop in activity, from ~5% GFP loss in DIVAC 1928 to 2.5–3.5% GFP loss in ∆5–∆9 (Fig. 6B), DIVAC 1703 (Fig. 5B), and DIVAC 751 (Fig. 7B). This raises the possibility that DIVAC 1928 contains a favorable, minimal spacing between the core regions that is disrupted by deletions between them.

Comparisons with other studies of the DIVAC region

Several previous studies have searched the chicken IgL locus for sequences capable of directing AID-mediated sequence diversification (26, 33–38). The results of these previous studies are in general agreement with those reported in this study in localizing DIVAC function to the region 3′ of the IgL C region exon. In one set of studies, the most important region was found to span the enhancer (“Region A” in Refs. 35, 37). Another study identified an important 3′ regulatory region consisting of the 4 kb 3′ of the enhancer (33), and a follow-up analysis identified mutation enhancer elements (MEEs) both 5′ and 3′ of the enhancer (5′MEE and 3′MEE) (38). The relationship between these regions and the sequences identified as important in our experiments is summarized in Fig. 7C. Our finding that DIVAC is dispersed over many sequences helps explain the failure of previous studies to converge on a single functional element.
activity associated with these sequences might be due to the greater sensitivity of our assay. We have not been able to identify regions of extended sequence similarity between chicken DIVAC sequences and mouse or human Ig loci. The possibility that mouse L chain gene enhancers contribute to DIVAC function warrants further investigation.

Although our analysis represents by far the most extensive deletion and mutation analysis of DIVAC sequences to date, it does not exclude the possibility that sequences outside of DIVAC 1-3 also have DIVAC (mutation enhancer) activity. Indeed, our finding that sequence elements that contribute to DIVAC function are broadly dispersed renders this possibility likely.

Mechanism of action of DIVAC

DIVAC might act to increase AID-mediated diversification of a nearby transcription unit by multiple different mechanisms, including AID recruitment, AID activation, rendering the transcribed DNA a better substrate for AID, and/or tilting the balance of deoxyuracil repair in favor of pathways that have a mutagenic outcome over those that result in high-fidelity repair. Despite extensive efforts, we have thus far been unable to detect AID binding reproducibly within the GFP2 cassette by ChIP (in the same cell lines used for the ChIP analyses of Fig. 8), and hence have not been able to evaluate the hypothesis that DIVAC acts to recruit AID. A previous study detected a small (~2-fold) increase in AID recruitment to the chicken IgL locus attributable to fragment A (Fig. 7C) using an assay involving an AID-DAM methylase fusion protein (36). We note that IgL(-)GFP2 cells (which express AID) exhibit GFP loss frequencies that, although low, are consistently an order of magnitude higher than in IgL+(+GFP2 AID(-) cells (Figs. 2, 3–7). This suggests that AID is recruited to GFP2, at least at low levels, in the absence of DIVAC, and is consistent with the finding that AID associates at low levels with thousands of active promoters in ex vivo-activated splenic B cells (11).

How does DIVAC facilitate mutation of a nearby transcription unit? Our data indicate that in our system it does not function by increasing transcription, as measured by steady-state RNA levels, GFP fluorescence levels, or total Pol II association. Nor does it decrease the levels of six different covalent histone modifications associated with GFP fluorescence levels, or total Pol II association. Despite this, we find that AID associates at low levels with thousands of active promoters in ex vivo-activated splenic B cells (11).

How does DIVAC facilitate mutation of a nearby transcription unit? Our data indicate that in our system it does not function by increasing transcription, as measured by steady-state RNA levels, GFP fluorescence levels, or total Pol II association. Nor does it decrease the levels of six different covalent histone modifications associated with GFP fluorescence levels, or total Pol II association. Despite this, we find that AID associates at low levels with thousands of active promoters in ex vivo-activated splenic B cells (11).

One appealing model is that DIVAC acts through the recruitment of trans-acting factors. If this is the case, then our data suggest that there are likely multiple relevant binding sites dispersed over the DIVAC region. There are 20 E boxes in DIVAC 1928, and as noted earlier, some deletions that reduce activity contain E boxes and others do not (Fig. 6). An important challenge for the future will be to determine to what extent E boxes, and non-E box sequences, contribute to DIVAC function. Although our data are broadly consistent with an additive contribution of multiple DNA motifs to DIVAC function (e.g., the increase in activity seen upon multimerization of 2b-2; Fig. 4B), they are also suggestive of regions with distinct and synergistic activities (e.g., fragments 2 and 3; Fig. 2). In this regard, it is noteworthy that the F2 and F3 cores were found to immunoprecipitate with distinct factors and epigenetic marks—the F2 core with RNA Pol II, and the F3 core with Cdk9, acetylated H4, H3K4me3—perhaps reflecting distinct contributions to DIVAC function. An interesting possibility raised by our findings is that DIVAC is constructed from well-known transcription factor binding sites, with Ig locus specificity achieved through the arrangement, number, and density of these binding motifs. A corollary of this is that the preferential mistargeting of SHM to particular non-Ig loci (19) might be the result of the presence of “degenerate” DIVAC elements in such loci.

Acknowledgments

We thank Dr. Y. Yamaguchi for providing anti-Spt5 Abs, Dr. M. Scharff and Dr. F.L. Kuang for reagents and advice, and members of the Schatz laboratory for advice and assistance during the course of this study.

Disclosures

The authors have no financial conflicts of interest.

References

**Fig. S1. Sequence analysis of mutations supported by DIVAC 2-3.** Genomic DNA was isolated from (A) IgL(-)GFP2 2-3 and (B) IgL(-)GFP2 cell lines on day 21 post-subcloning. A 1217 bp region of the GFP2 expression cassette was PCR amplified, cloned, and sequenced. The pie chart shows the proportion of sequences containing the indicated number of mutations; the middle circle denotes the number of sequences analyzed. Mutation spectra are depicted in the box diagrams (left, sequences from this analysis; right, published data from (34)). Further information concerning the mutations and mutation frequencies are detailed below the mutation spectra.
**Fig. S2. Analysis of DIVAC function associated with murine Ig light chain enhancers.** (A) Schematic diagram of the murine Ig light chain enhancers. The sizes of the fragments containing the intronic Igκ enhancer (iEκ), 3' Igκ enhancer (3'Eκ), Igλ 2-4 enhancer (Eλ2-4), and Igλ 3-1 enhancer (Eλ3-1), are indicated, as are the locations of E boxes conforming (closed symbols), or not conforming (open symbols) to the motif CASSTG. (B) Fluctuation analysis of GFP loss in subclones with data depicted as in Fig. 2D.
Fig. S3

A

AGCAGGCCGCACTGACAGCTCCATCAGCA
CAGCTG
GGGCCACACAAAGAGCTGGGTTACTGTGGGCAGCAGGCTGAAACCCGAAAACAAGAGCTGGGGCCTCAGAATAGCCCCGGGAGCAGGCAGGGCCTGGGGGAGAGGGCAAGCACAGGCCCAGGGCCACACAGCCCTTCCAGGAAGGCACAGCGCTGTCAGGGTGCAGCACGCTCAGCCCCACCATG

CB

GCTG
TGCGGCCGGGGCATCCCCAAGCTAAATTTACTTCTCAGTCTCCAATCAGAAACTGAAGCTGAGGGGCCCACGCCGGCCAAAAAAAGGAAACGAAACAGTCTCCAGAAAGCACTGACGTGTGAAGCAGAGCGAGCGCCGCGCAAACCGGCCGCCATGTCACACACCTCAGGTGGGGCTTTGCCAGACTGAGCTTTGCTGCTGCTCGGGGTGGGTGCCCACGGCCTGGG
CACATG
GGATGGGGTACACACGTACACACACTTG
CACACCCACACCCCAACACTTCAGGTG
ATGCTGGTG

C

GCTGACCGCAGCCACGGGGGCAACAGTTTCTCTTGCTAAAAATTGTAGCCGGGAAGAAAA
CACGTG
GCAACTTCGGCCAAACAG
CAGCTG
GAGGACAGGAATAGCCGTGGCCACGGCACGCTCTGCTTCCTCGGCACAAACATTCCAGTACGTGGC
ACCACGAGCGCCGCTGCCCGGCACAGCAGCAAGCAGAGCCAGGAGCAGGAAATGCTGATTTGGGCCCTTGGCCATGGCTGAGAGAAGAGGCTTCCAGGGAGCTGGTCAGCTTGGTCCCCAAGCTGTGGCTTGGGGAAATGATG GGAGGGGATTGCCACTGCCCACCCTGCAGAGCAGGCTCTGGTCCCATCTCACTGCAGGGCACCAGGGCGTTTGCACTGCAGCAATTCACAGAAACATTGAAATGGCTCCT
Fig. S3. Sequences of key DIVAC fragments. (A) DIVAC 1703. (B) DIVAC 1928. In both, the F2 core is the 5’ underlined section and the F3 core is 3’ underlined section. E boxes are indicated in bold.
<table>
<thead>
<tr>
<th>Construct Name</th>
<th>Primer Sequence</th>
<th>Template DNA</th>
<th>Product Size (bp)</th>
</tr>
</thead>
</table>
| 1              | F: cccgctagcctgggaacaggggagttctg  
R: cccactagtctctgtgtccgctcagtt | DT40 Genomic | 1493 |
| 2              | F: cccgctagctggagaagccgacaccgtctgtaat  
R: cccactagttgcttttgggtctacagtt | DT40 Genomic | 1476 |
| 3              | F: cccgctagcggcggagccccagcttgag  
R: cccactagtgcgggtgcctccccaggatacc  
| DT40 Genomic | 1633 |
| 2a             | F: 2 F  
R: cccactagttcgtctgatgccatc  
| 2 | 590 |
| 2b             | F: cccgctagcaacatccccccacaccagacaagcag  
R: cccactagtcaacagataagctggagact  
| 2 | 650 |
| 2c             | F: cccgctagcccccccagcttgacccc  
R: 2 R  
| 2 | 350 |
| 2b-1           | F: 2b F  
R: cccactagttgctttgctctgccccctgtg  
| 2 | 227 |
| 2b-2           | F: cccgctagcagccccggagcggcagggctct  
R: cccactagtcaacagctcctgctgcaacagcccccctgtg  
| 2 | 316 |
| 2b-3           | F: cccgctagcagacgagcccctgctgaaaccctgcccctgc  
R: 2b R  
| 2 | 244 |
| 1-3            | F: 1 F  
R: 3 R  
| DT40 Genomic | 4384 |
| 2-3            | F: 2 F  
R: 3 R  
| DT40 Genomic | 2989 |
| 2-3Δ2a         | F: 2b F  
R: 3 R  
| 2-3 | 2452 |
| 2-3Δ2c         | F: 2 F  
R: 2b R  
| 2-3 (added to 3) | 2640 |
| 2-3Δ2a/2c      | Digested 2b vector and inserted into 3  
| -- | 2103 |
| 2-3ΔN          | F: 2 F  
R: cccactagtcaacagctcctgctgctctctgtg  
| 2-3 | 1610 |
| 2-3Δ3′N        | F: 2 F  
R: cccactagttaacatggcagaaacaacttaac  
| 2-3 | 2118 |
| 2-3Δ5′N        | F1: 2 F  
R1: tccaaacatgccaaagggctgtttgctgctgac  
F2: tgaccgaacagccccctttgctgttggctgctg  
R2: 3 R  
| 2-3 | 2347 |
| 2-3ΔCR1        | F: 2 F  
R: cccactagttagagccattctcaatgttctgtga  
| 2-3 | 2626 |
| 2b-2 Δ1-50      | F: cccgctagcagccccccggaggacagcagccag  
R: 2b-2 R  
| 2b-2 | 265 |
| 2b-2 Δ51-100    | F1: 2b-2 F  
R1: ctgcagttttggtcagggccttctggctgct  
F2: tgcagccccccacaccctgtgctgctgctgctg  
R2: 2b-2 R  
| 2b-2 | 266 |
| 2b-2 Δ101-150   | F1: 2b-2 F  
R1: ctggagagactgagaagggcgttgctgacccctgca  
F2: cccgctagctctcttcctctctcctctcaatca  
R2: 2b-2 R  
| 2b-2 | 266 |
| 2b-2           | F1: 2b-2 F  
| 2b-2 | 265 |
| Δ151-201 | R1: actgtttgtttcttttttaatattagtttggttga  
|          | F2: cccgctagcacaagggaaacagaaactagt  
|          | R2: 2b-2 R |
| Δ202-254 | F1: 2b-2 F  
|          | R1: gcggccgggttgtcgcgcttggccggc  
|          | F2: cccgctagccgcggcacaaccggcgccgc  
|          | R2: 2b-2 R |
| Δ255-316 | F: 2b-2 F  
|          | R: cccactagtctctgctcttcactac  
| 2b-2 154 | F1: cccgctagctcagcccccaacctcagac  
|          | R1: cccactagtctctgctcttcactac |
| 2b-2 263 | F: 2b-2 F  
|          | R: gcggccgggttgtcgcgcttggccggc  
|          | F2: cccgctagccgcggcacaaccggcgccgc  
|          | R2: 2b-2 R |
| 2b-2 3a  | F: 3 F  
|          | R: cccactagtctctgctcttcactac |
| 2b-2 3b  | F: cccgctagccgcggcacaaccggcgccgc  
|          | R: cccactagtctctgctcttcactac |
| 2b-2 3c  | F: cccgctagccgcggcacaaccggcgccgc  
|          | R: cccactagtctctgctcttcactac |
| 3aΔa     | F: cccgctagctcagcccccaacctcagac  
|          | R: 3 R  |
| 3a+b     | F: 3 F  
|          | R: 3b R  |
| 3c3      | F: 3c F  
|          | R: 3 R  |
| PG       | F1: cccgctagctcagcccccaacctcagac  
|          | R1: cccactagtctctgctcttcactac  
|          | W: 700 |
| DIVAC 1703 | F1: cccgctagctcagcccccaacctcagac  
|          | R1: cccactagtctctgctcttcactac  
|          | DIVAC 2-3 | 1703 |
| DIVAC 1703 Δ1-200 | F: cccgctagctcagcccccaacctcagac  
|          | R: 2-3ΔCR1 R |
| DIVAC 1703 Δ201-400 | F1: DIVAC 1703 F1  
|          | R1: tctgcggccagccccccacctgacacgcgcttg  
|          | F2: cacgccgggcttgtgctctggccgcag  
|          | R2: 2-3ΔCR1 R |
| DIVAC 1703 Δ401-600 | F1: DIVAC 1703 F1  
|          | R1: cacgccgggcttgtgctctggccgcag  
|          | F2: cacgccgggcttgtgctctggccgcag  
|          | R2: 2-3ΔCR1 R |
| DIVAC 1703 Δ601-800 | F1: DIVAC 1703 F1  
|          | R1: tctgcggccagccccccacctgacacgcgcttg  
|          | F2: cacgccgggcttgtgctctggccgcag  
<p>|          | R2: 2-3ΔCR1 R |
| DIVAC 1703 | F1: DIVAC 1703 F1 |
|          | DIVAC 1703 | 1503 |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th>R1:</th>
<th>F1:</th>
<th>F2:</th>
<th>R2:</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1703</td>
<td>Δ801-1000</td>
<td>cagtgagacaccccaagcaggtgtagctc</td>
<td>DIVAC 1703 F1</td>
<td>gagctacacacctgcttgggtgtctcactg</td>
<td>2-3ΔCR1 R</td>
<td>DIVAC 1703</td>
</tr>
<tr>
<td></td>
<td>1703</td>
<td>Δ1001-1200</td>
<td>gcccctccaaacatgcggaccagcccatgtc</td>
<td>DIVAC 1703 F1</td>
<td>agcagacgtgtggcacttttggaggggcagctgct</td>
<td>2-3ΔCR1 R</td>
</tr>
<tr>
<td></td>
<td>1703</td>
<td>Δ1201-1400</td>
<td>gcccctccaaacatgcggaccagcccatgtc</td>
<td>cagtgagacaccccaagcaggtgtagctc</td>
<td>gagctacacacctgcttgggtgtctcactg</td>
<td>2-3ΔCR1 R</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1503</td>
<td>F1:</td>
<td>DIVAC 1703 F1</td>
<td>DIVAC 1703 1503</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R1:</td>
<td>cagtgagacaccccaagcaggtgtagctc</td>
<td>DIVAC 1703 1503</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>F2:</td>
<td>gagctacacacctgcttgggtgtctcactg</td>
<td>DIVAC 1703 1503</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R2:</td>
<td>2-3ΔCR1 R</td>
<td>DIVAC 1703 1503</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1400</td>
<td>F:</td>
<td>DIVAC 1703 F1</td>
<td>DIVAC 1703 1400</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R:</td>
<td>cccactagtgccacggctattct</td>
<td>DIVAC 1703 1400</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1928</td>
<td>F:</td>
<td>2b F</td>
<td>DIVAC 1928</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R:</td>
<td>2-3ΔCR1 R</td>
<td>DIVAC 1928</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1878</td>
<td>F1:</td>
<td>DIVAC 1928 F1</td>
<td>DIVAC 1928 1878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R1:</td>
<td>cctttttttggccggca ccctgacagcgctgtg</td>
<td>DIVAC 1928 1878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>F2:</td>
<td>agcgctcgctcttgctgggctggccaaaaaaggaaac</td>
<td>DIVAC 1928 1878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R2:</td>
<td>2-3ΔCR1 R</td>
<td>DIVAC 1928 1878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1878</td>
<td>F1:</td>
<td>DIVAC 1928 F1</td>
<td>DIVAC 1928 1878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R1:</td>
<td>cctttttttggccgggc ttggggatgccccggc</td>
<td>DIVAC 1928 1878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>F2:</td>
<td>ggggcatccccaagccgccccggccaaaaaaggaaac</td>
<td>DIVAC 1928 1878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R2:</td>
<td>2-3ΔCR1 R</td>
<td>DIVAC 1928 1878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1878</td>
<td>F1:</td>
<td>DIVAC 1928 F1</td>
<td>DIVAC 1928 1878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R1:</td>
<td>ggcgctcgctctgctcgtgggcccctcagcttc</td>
<td>DIVAC 1928 1878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>F2:</td>
<td>gctgaggggcccacgagcagagcgagcgccccgc</td>
<td>DIVAC 1928 1878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R2:</td>
<td>2-3ΔCR1 R</td>
<td>DIVAC 1928 1878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1878</td>
<td>F1:</td>
<td>DIVAC 1928 F1</td>
<td>DIVAC 1928 1878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R1:</td>
<td>gtctggcaaagcccctcacacgtcagtgctttc</td>
<td>DIVAC 1928 1878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>F2:</td>
<td>agcactgacgtgtgaggggctttgccagactga</td>
<td>DIVAC 1928 1878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R2:</td>
<td>2-3ΔCR1 R</td>
<td>DIVAC 1928 1878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1878</td>
<td>F1:</td>
<td>DIVAC 1928 F1</td>
<td>DIVAC 1928 1878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R1:</td>
<td>cctgtgtgtgctgtgggaccagcccatgtcac</td>
<td>DIVAC 1928 1878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>F2:</td>
<td>acatgggctggtccacagcagcacagccaggg</td>
<td>DIVAC 1928 1878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R2:</td>
<td>2-3ΔCR1 R</td>
<td>DIVAC 1928 1878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1878</td>
<td>F1:</td>
<td>DIVAC 1928 F1</td>
<td>DIVAC 1928 1878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R1:</td>
<td>ggatggggttttcttattgccccagggctcca</td>
<td>DIVAC 1928 1878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>F2:</td>
<td>agccctggggcaataagaaaaccccatccgta</td>
<td>DIVAC 1928 1878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R2:</td>
<td>2-3ΔCR1 R</td>
<td>DIVAC 1928 1878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1878</td>
<td>F1:</td>
<td>DIVAC 1928 F1</td>
<td>DIVAC 1928 1878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R1:</td>
<td>ctcagcccatgcagaggagtgcagggcagctgctg</td>
<td>DIVAC 1928 1878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>F2:</td>
<td>cagctgccctgcactcctctgcatgggctgac</td>
<td>DIVAC 1928 1878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R2:</td>
<td>2-3ΔCR1 R</td>
<td>DIVAC 1928 1878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1878</td>
<td>F1:</td>
<td>DIVAC 1928 F1</td>
<td>DIVAC 1928 1878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R1:</td>
<td>cccctccaaacatgcctgcagagagatgggaag</td>
<td>DIVAC 1928 1878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>F2:</td>
<td>cccatctctctgcaggcatgttttgaggccag</td>
<td>DIVAC 1928 1878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R2:</td>
<td>2-3ΔCR1 R</td>
<td>DIVAC 1928 1878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1878</td>
<td>F1:</td>
<td>DIVAC 1928 F1</td>
<td>DIVAC 1928 1878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R1:</td>
<td>tggggctggcggccaaaaaaggcaaaaacatggcagaa</td>
<td>DIVAC 1928 1878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>F2:</td>
<td>ctgccatctctctgcaccctgct</td>
<td>DIVAC 1928 1878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R2:</td>
<td>2-3ΔCR1 R</td>
<td>DIVAC 1928 1878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1878</td>
<td>F1:</td>
<td>DIVAC 1928 F1</td>
<td>DIVAC 1928 1878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R1:</td>
<td>ctgttgcccccgtggcactgtggcgtacccgta</td>
<td>DIVAC 1928 1878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>F1:</td>
<td>F2:</td>
<td>R1:</td>
<td>R2:</td>
<td>Δ</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>----</td>
<td></td>
</tr>
<tr>
<td>Δ10</td>
<td>gggtaccgccacgtgccacggggcagagtttc</td>
<td>2-3ΔCR1 R</td>
<td>gggtaccgccacgtgccacggggcagagtttc</td>
<td>2-3ΔCR1 R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIVAC 1928 Δ11</td>
<td>F1: gggtaccgccacgtgccacggggcagagtttc</td>
<td>2-3ΔCR1 R</td>
<td>F1: 2b F</td>
<td>1878</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIVAC 1928 Δ12</td>
<td>F1: 2b F</td>
<td>2-3ΔCR1 R</td>
<td>F1: 2b F</td>
<td>1878</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIVAC 1928 Δ13</td>
<td>F1: 2b F</td>
<td>2-3ΔCR1 R</td>
<td>F1: 2b F</td>
<td>1878</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIVAC 1928 Δ14</td>
<td>F1: 2b F</td>
<td>2-3ΔCR1 R</td>
<td>F1: 2b F</td>
<td>1878</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIVAC 1928 Δ15</td>
<td>F1: 2b F</td>
<td>2-3ΔCR1 R</td>
<td>F1: 2b F</td>
<td>1878</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIVAC 1928 Δ16</td>
<td>F1: 2b F</td>
<td>2-3ΔCR1 R</td>
<td>F1: 2b F</td>
<td>1878</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Δ17</td>
<td>F1: 1 F</td>
<td>3 R</td>
<td>F1: 1 F</td>
<td>4184</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ΔF2 core</td>
<td>F1: 1 F</td>
<td>3 R</td>
<td>F1: 1 F</td>
<td>4034</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ΔF3 core</td>
<td>F1: 1 F</td>
<td>3 R</td>
<td>F1: 1 F</td>
<td>3834</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ΔF2/F3 core</td>
<td>F1: 2 F</td>
<td>3 R</td>
<td>F1: 2 F</td>
<td>2789</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ΔF2 core</td>
<td>F1: 2 F</td>
<td>3 R</td>
<td>F1: 2 F</td>
<td>2639</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ΔF3 core</td>
<td>F1: 2 F</td>
<td>3 R</td>
<td>F1: 2 F</td>
<td>2439</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ΔF2/F3</td>
<td>F1: 2 F</td>
<td>3 R</td>
<td>F1: 2 F</td>
<td>2159</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**DIVAC 1928**

1-3 ΔF2 core

1-3 ΔF3 core

1-3 ΔF2/F3 core

2-3 ΔF2 core

2-3 ΔF3 core

2-3 ΔF2/F3
<table>
<thead>
<tr>
<th>Gene</th>
<th>PCR Fragment</th>
<th>Primer</th>
<th>Ta</th>
<th>Product Size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>γ-actin</td>
<td>--</td>
<td>F: acagcattcgggtggaggtg&lt;br&gt;R: ttgagcatcctctctac&lt;br&gt;&lt;b&gt;Probe:&lt;/b&gt; cctcagccccctcctcagccc</td>
<td>60</td>
<td>127</td>
</tr>
<tr>
<td>RSV Promoter (in IgLGFP2)</td>
<td>a</td>
<td>F: ttcttcatgcaattgtcggt&lt;br&gt;R: ctgctccctgctgtg&lt;br&gt;&lt;b&gt;Probe:&lt;/b&gt; tgtctgccgcaactactcagcc</td>
<td>60</td>
<td>95</td>
</tr>
<tr>
<td>GFP2 (in IgLGFP2)</td>
<td>b</td>
<td>F: tgaccctgaaagtcatctgc&lt;br&gt;R: gaagttcgtcgtcatg&lt;br&gt;&lt;b&gt;Probe:&lt;/b&gt; cccaccctctgagccaccct</td>
<td>60</td>
<td>125</td>
</tr>
<tr>
<td>IRES (in IgLGFP2)</td>
<td>c</td>
<td>F: ccctaggaatgtctgcaag&lt;br&gt;R: cctaaacgtaactgcccgaag&lt;br&gt;&lt;b&gt;Probe:&lt;/b&gt; tttccgggccctcactacattgc</td>
<td>60</td>
<td>139</td>
</tr>
<tr>
<td>BSR (in IgLGFP2)</td>
<td>d</td>
<td>F: aatggctttctgccaaacacag&lt;br&gt;R: gcgcagacagaaagattcata&lt;br&gt;&lt;b&gt;Probe:&lt;/b&gt; cgaattgcgcgtccccacatg</td>
<td>60</td>
<td>141</td>
</tr>
<tr>
<td>2b</td>
<td>e</td>
<td>F: tccaatcagaaactgaaagtg&lt;br&gt;R: cgctctgcttcacagcttc&lt;br&gt;&lt;b&gt;Probe:&lt;/b&gt; aggaaacggaggaagtcag</td>
<td>60</td>
<td>80</td>
</tr>
<tr>
<td>3</td>
<td>f</td>
<td>F: agcggagacaggagatgc&lt;br&gt;R: ctgctgctctgctgcc&lt;br&gt;&lt;b&gt;Probe:&lt;/b&gt; cgctgcatctctgccagca</td>
<td>60</td>
<td>109</td>
</tr>
<tr>
<td>W fragment, upstream of cores (DIVAC 1)</td>
<td>g</td>
<td>F: gatctgtcacagccacttc&lt;br&gt;R: gctgatgagacataagg&lt;br&gt;&lt;b&gt;Probe:&lt;/b&gt; cagcaacactccagccatc</td>
<td>60</td>
<td>113</td>
</tr>
<tr>
<td>W fragment, between cores (Δ601-800)</td>
<td>h</td>
<td>F: aacagctccaaactgagacc</td>
<td>R: cacggcacaaggttgttatt</td>
<td>Probe: ccgagcttgtccacaggg</td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>---</td>
<td>-------------------------</td>
<td>-------------------------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>18S</td>
<td>--</td>
<td>F: taaaggaattgacgggaagg</td>
<td>R: tgtcaatctgtccgtgctc</td>
<td>Probe: cgaggctccactctgtggt</td>
</tr>
</tbody>
</table>

**Table S1. Primer and probe sequences for experiments.** (A) Primers for the creation of all tested DIVAC fragments. (B) qPCR primer pairs and probes for gene expression and ChIP. All sequences are written 5’ to 3’.