Cell Reports, Volume *29*

Supplemental Information

Topologically Associated Domains Delineate Susceptibility to Somatic Hypermutation

Filip Senigl, Yaakov Maman, Ravi K. Dinesh, Jukka Alinikula, Rashu B. Seth, Lubomira Pecnova, Arina D. Omer, Suhas S.P. Rao, David Weisz, Jean-Marie Buerstedde, Erez Lieberman Aiden, Rafael Casellas, Jiri Hejnar, and David G. Schatz

Figure S1

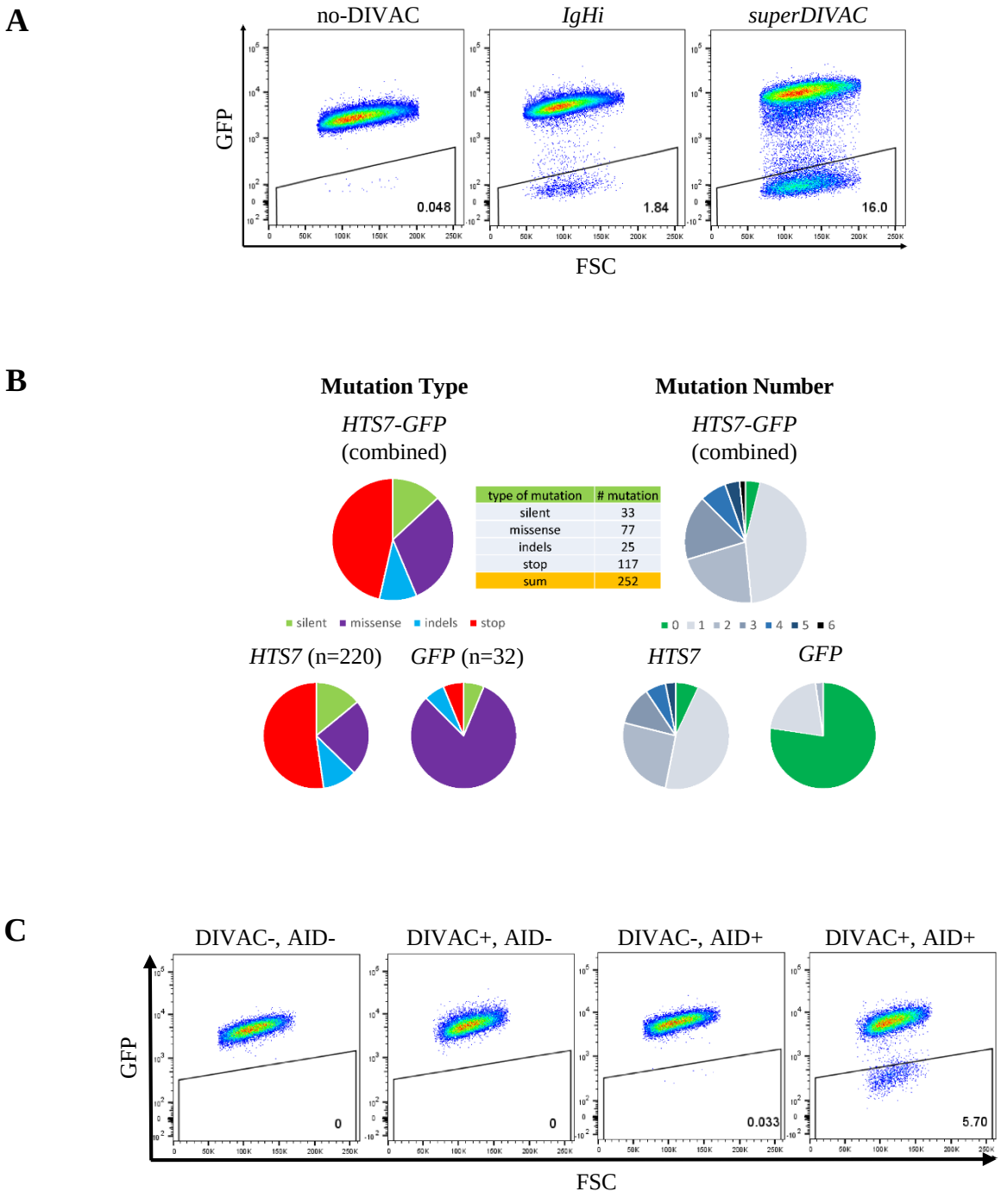

Figure S1. Verification of the DIVAC- and AID-dependence of *GFP7* vector GFP fluorescence loss *Related to Figure 1*

(A) Representative examples of FACS analysis for GFP fluorescence loss with WT Ramos clones with *GFP7* either lacking DIVAC or containing *superDIVAC* or *IgHi*.

(B) Type and frequency of mutations in *HTS7-GFP* region of the no-DIVAC-*GFP7* vector in sorted GFP-negative Ramos cells. 128 sequences were analyzed. At top are shown the types of mutations (left) and number of mutations per sequence (right) for the combined *HTS7-GFP* region, while below are shown the same data for the *HTS7* and *GFP* regions separately.

(C) Representative examples of FACS analysis of either AID-deficient (AID-) or wild-type (AID+) clones of DT40 cells infected with *GFP7* lacking (DIVAC-) or containing (DIVAC+) a chicken *IgL* DIVAC element (Blagodatsky et al., 2009).

Figure S2

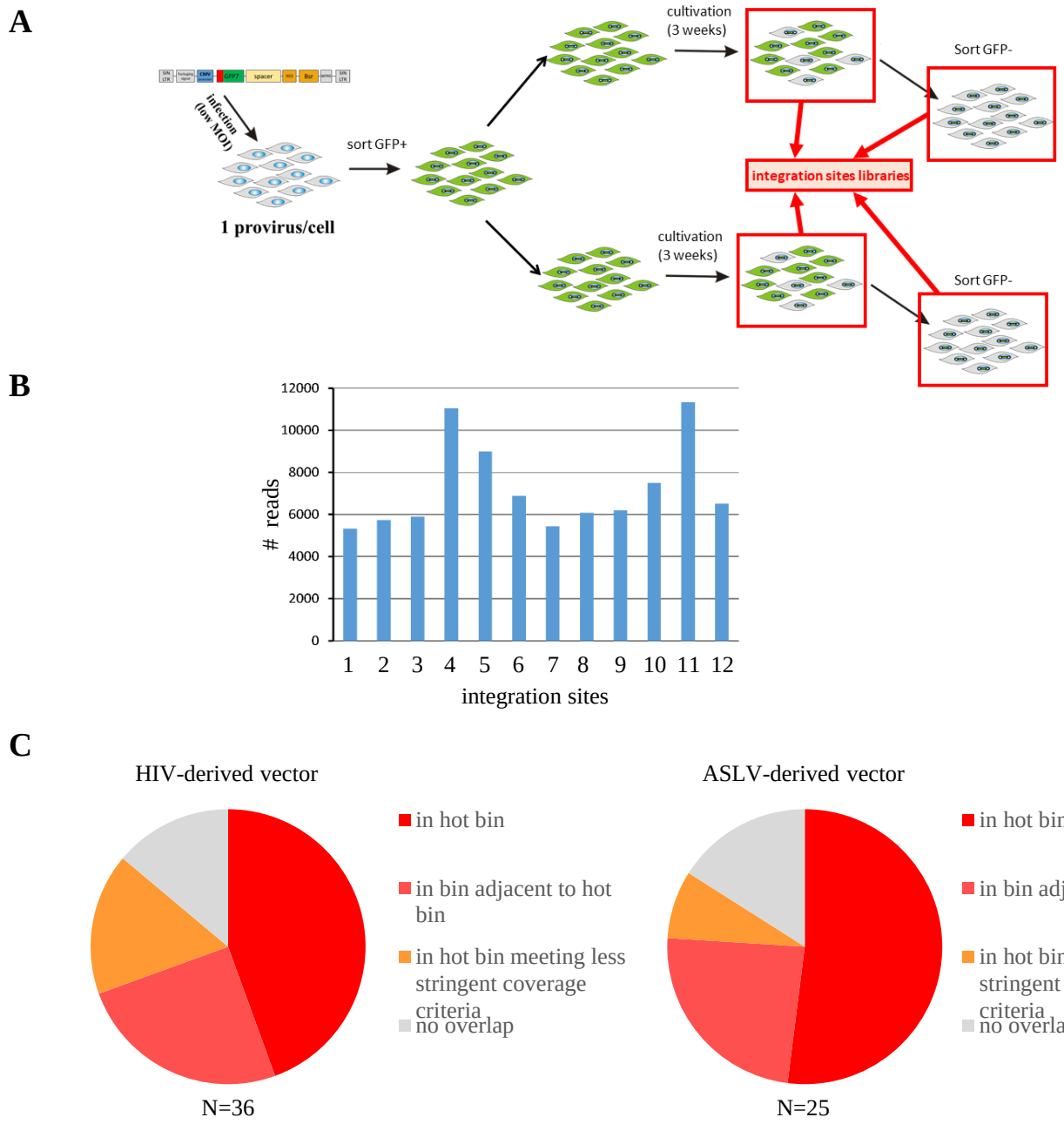

Figure S2. DIVAC-trap HTISA assay *Related to Figure 1*

(A) Schematic diagram of the DIVAC-trap HTISA assay. Cells were infected at low multiplicity to achieve one copy of *GFP7* per cell, sorted for GFP-positive cells, and split into two parallel populations at the time of sorting. After 3 weeks of growth, cells were selected with blasticidin (not shown) to remove cells with transcriptionally silent *GFP7* after which GFP-negative cells were sorted. Vector integration site libraries were prepared from the unsorted (Total) and GFP-negative (GFP-) populations using HTISA method to assess the frequency of individual integration sites in the population.

(B) Numbers of *GFP7* vector integration site sequence reads from each integration site obtained by HTISA sequencing of a mix of equimolar amounts of genomic DNAs from 12 Ramos clones with different vector integration sites.

(C) Analysis of no-DIVAC *GFP7* HIV vector (left) and ASLV vector (right) insertion sites that support substantial SHM, showing the proportion of integrations that occur within hot bins, within bins adjacent to hot bins, and within bins that meet criteria for SHM-susceptibility but not the stringent criteria required for coverage, and hence are not among the 175 high-confidence hot bins. A small fraction of insertion sites do not overlap with hot bins but none of these are in cold bins (Tables S1 and S2).

Figure S3

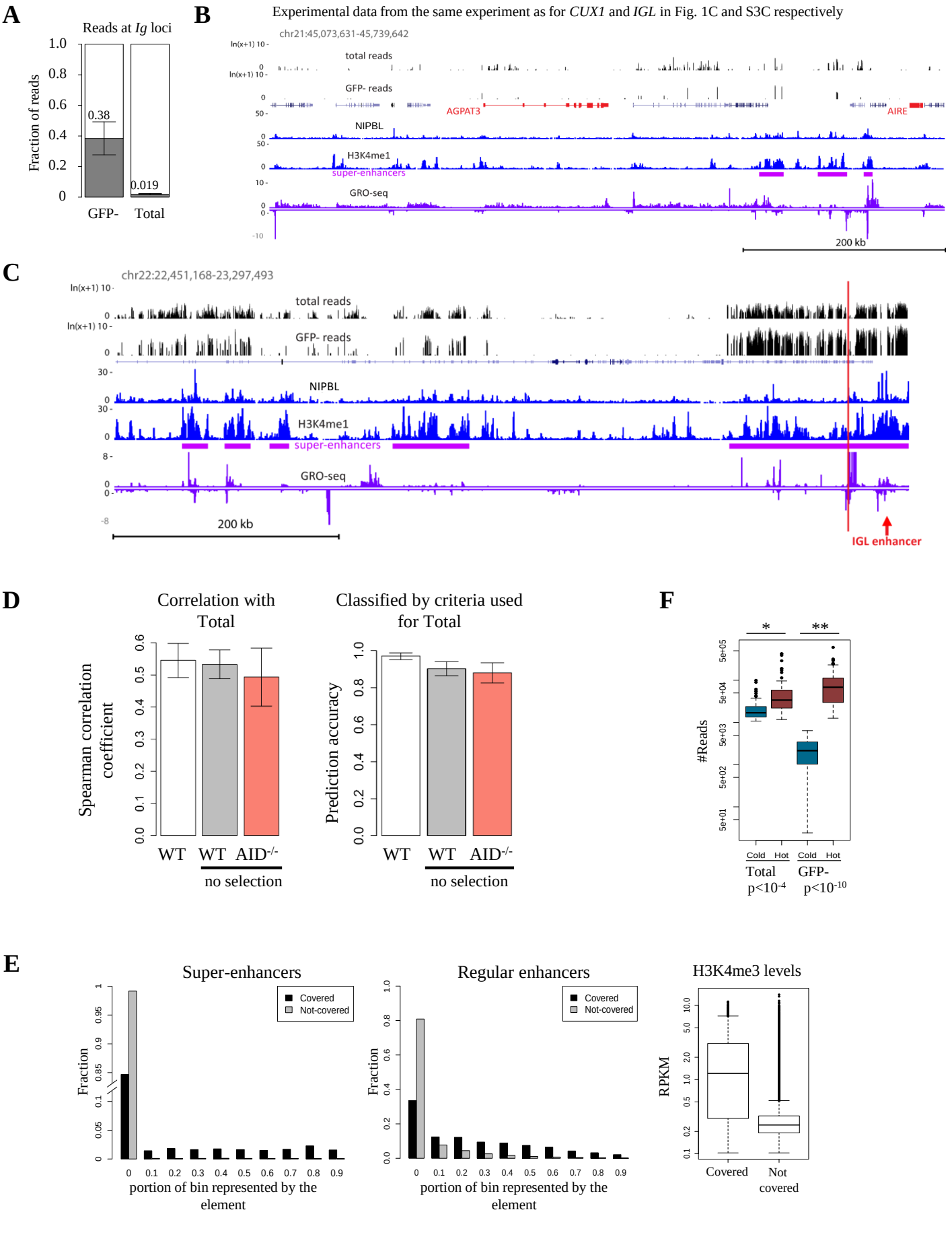
A
Reads at Ig loci
Fraction of reads
1.0
0.8
0.6
0.4
0.2
0
0.38
0.019
GFP-
Total
B
Experimental data from the same experiment as for CUX1 and IGL in Fig. 1C and S3C respectively
chr21:45,073,631-45,739,642
total reads
GFP- reads
AGPAT3
AIRE
NIPBL
H3K4me1
super-enhancers
GRO-seq
200 kb
C
chr22:22,451,168-23,297,493
total reads
GFP- reads
NIPBL
H3K4me1
super-enhancers
GRO-seq
200 kb
IGL enhancer
D
Correlation with Total
Spearman correlation coefficient
WT
WT
AID-/-
no selection
Classified by criteria used for Total
Prediction accuracy
WT
WT
AID-/-
no selection
F
*
**
#Reads
Cold
Hot
Cold
Hot
Total
p<10-4
GFP-
p<10-10
E
Super-enhancers
Covered
Not-covered
Fraction
portion of bin represented by the element
Regular enhancers
Covered
Not-covered
Fraction
portion of bin represented by the element
H3K4me3 levels
RPKM
Covered
Not covered

Figure S3. Mapping of SHM-susceptible regions in the human B cell genome *Related to Figures 1 and 2*

(A) Proportion of *GFP7* integration site sequence reads in *Ig* loci in GFP-negative and Total populations.

(B) DIVAC-trap HTISA assay data for the *AGPAT3* locus derived from the same experiment as *CUX1* and *IGL* loci in Fig. 1C and S3C respectively.

(C) Example of DIVAC-trap HTISA assay data for the *IGL* locus. No-DIVAC *GFP7* integration site sequence read tracks for Total and GFP-negative populations (log scale) are shown above tracks for NIPBL, H3K4me1, super-enhancers, and GRO-seq (sense and antisense above and below the line, respectively). *IGL* locus region is characterized by abundant integration sites in both Total and GFP-negative. 140 kb region deleted in the rearranged *IGL* locus (red vertical line) and *IGL* enhancer (red arrow) are indicated.

(D) At left is shown a plot of correlation coefficients obtained by comparing the distribution of the genome integration sites of a pool of reads from all Total libraries and a single Total library (WT), and by comparing the integration sites for the same pool with samples from cells that did not undergo blasticidin selection and were collected two days post infection, thus representing *GFP7* vector integration preference (no selection) in either wild-type (WT) or AID-deficient (AID-/-) cells (left chart). The data demonstrate that the correlation between selected and unselected samples is nearly as high as between two selected libraries. At right is shown a plot of how well a classification based on the integration sites from the selected pool of all Total libraries is able to predict the integration sites arising from the same three libraries analyzed in the graph to left. The unselected library integration sites are predicted nearly as well those from the selected library.

(E) Occupancy of covered and uncovered bins by active super-enhancers and regular enhancers (left and middle bar charts). H3K4me3 levels in covered and uncovered bins (right box-plot). The data demonstrate that super-enhancers are virtually completely absent from bins that are not covered, and enhancers and H3K4me3 are much more abundant in covered than non-covered bins (two-tailed T-test, $p<1e-20$).

(F) Numbers of *GFP7* integration site sequence reads in Total and GFP-negative samples in hot and cold TADs. Two-tailed T-test, with P-values shown below the plot.

Figure S4

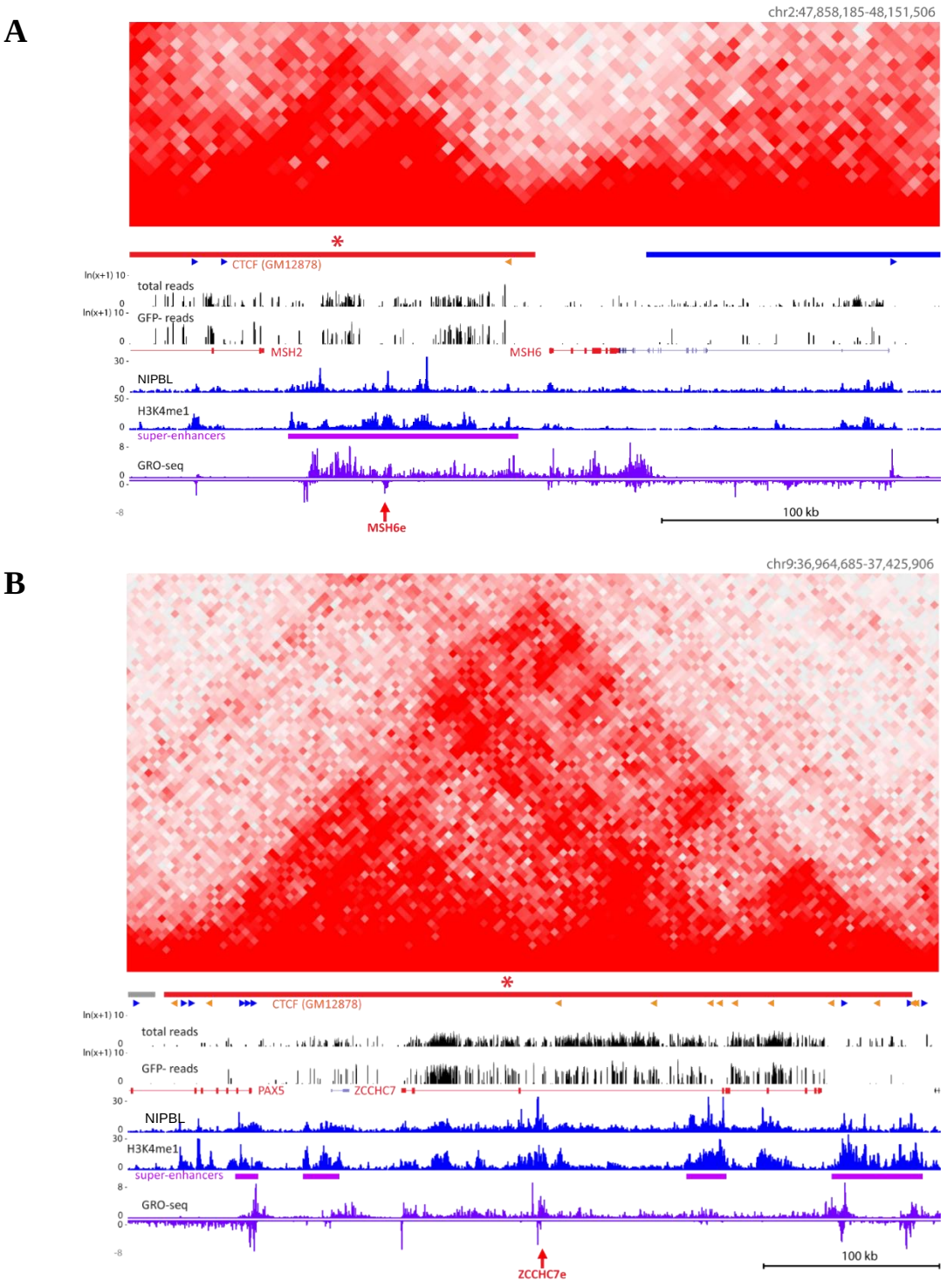

Figure S4. Examples of SHM-susceptible (hot) and SHM-resistant (cold) TADs *Related to Figures 3 and 5*

(A, B) Hi-C matrices are shown above DIVAC-trap HTISA data for cold (blue) and hot (red) TADs, with high confidence hot and cold TADs marked with an asterisk. CTCF motif orientations (sense, blue; antisense, orange) overlapping with CTCF ChIP-seq peaks for the GM12878 human lymphoblastic cell line are indicated, with data tracks for NIPBL, H3K4me1, super-enhancers, and GRO-seq (sense and antisense above and below the line, respectively). (A) Example of juxtaposed hot and cold TADs. (B) Example of hot region. (A, B) Hot TADs where candidate enhancer elements were isolated (indicated by red arrows).

Figure S5

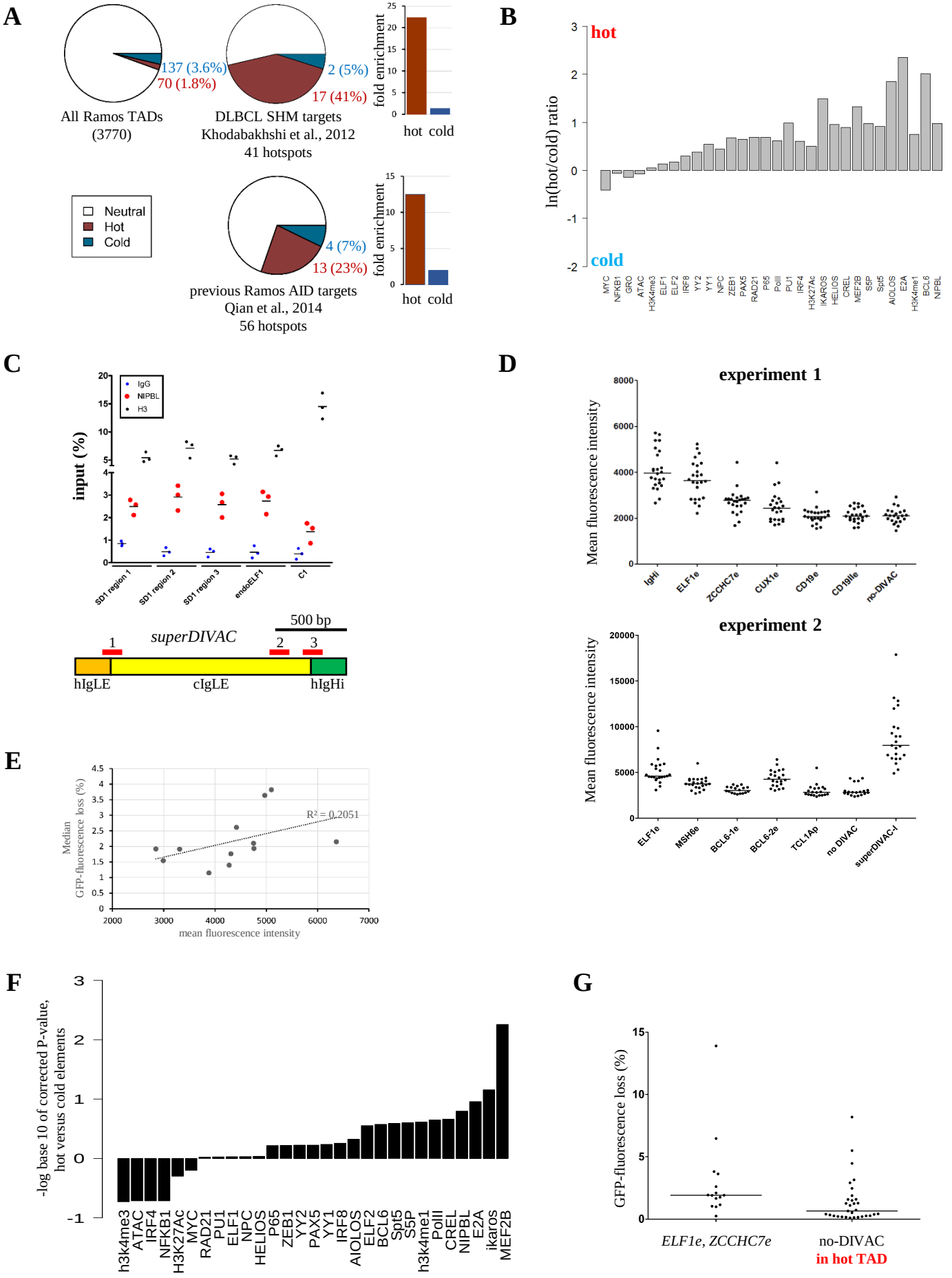
A
137 (3.6%)
70 (1.8%)
All Ramos TADs
(3770)
2 (5%)
17 (41%)
DLBCL SHM targets
Khodabakhshi et al., 2012
41 hotspots
fold enrichment
hot cold
Neutral
Hot
Cold
4 (7%)
13 (23%)
previous Ramos AID targets
Qian et al., 2014
56 hotspots
B
hot
cold
ln(hot/cold) ratio
MYC NFKB1 GRO ATAC H3K4me3 ELF1 ELF2 IRF8 YY2 YY1 NPC ZEB1 PAX5 RAD21 P65 PolII PU1 IRF4 H3K27Ac IKAROS HELIOS CREL MEF2B S5P Spt5 AIOLOS E2A H3K4me1 BCL6 NIPBL
C
IgG
NIPBL
H3
input (%)
SD1 region 1
SD1 region 2
SD1 region 3
endoELF1
C1
500 bp
superDIVAC
1
2
3
hIgLE
cIgLE
hIgHi
D
experiment 1
Mean fluorescence intensity
IgH ELF1e ZCCHC7e CUX1e CD19e CD19Ile no-DIVAC
experiment 2
Mean fluorescence intensity
ELF1e MSH6e BCL6-1e BCL6-2e TCL1Ap no DIVAC superDIVAC-I
E
Median GFP-fluorescence loss (%)
R² = 0.2051
mean fluorescence intensity
F
-log base 10 of corrected P-value, hot versus cold elements
h3k4me3 ATAC IRF4 NFKB1 H3K27Ac MYC RAD21 PU1 ELF1 NPC HELIOS P65 ZEB1 YY2 PAX5 YY1 IRF8 AIOLOS ELF2 BCL6 Spt5 S5P h3k4me1 PolII CREL NIPBL E2A ikaros MEF2B
G
GFP-fluorescence loss (%)
ELF1e, ZCCHC7e
no-DIVAC
in hot TAD

Figure S5. Overlap with previously identified AID targets and characterization of non-Ig DIVAC elements ***Related to Figures 3 and 5***

(A) Pie charts in the center indicate the proportion and number of previously identified SHM targets in DLBCL (Khodabakhshi et al., 2012) or repair-deficient Ramos cells (Qian et al., 2014) that lie within hot or cold TADs. For comparision, at left is a pie chart showing the fraction of all TADs represented by hot and cold TADs. At right, bar graphs show fold enrichment of the previously identified SHM targets in hot and cold TADs compared to that expected for a random distribution in all TADs.

(B) Enrichment analysis in hot versus cold TADs. The signal for each parameter or factor (log(RPKM)) was compared between hot and cold TADs. Ratios of RPKM signal are plotted. All data are derived from Ramos and include ChIP-seq for the indicated transcription or chromatin factors or modified histones, GRO-seq (GRO), and ATAC-seq (ATAC). NPC, nuclear pore components; PolII, total RNA Pol 2; S5P, serine-5-phosphorylated Pol2.

(C) ChIP-qPCR analysis of NIPBL binding in three different regions (red bars) of the *superDIVAC* element inserted into a cold TAD on chr22 (see Fig. 7A). Each data point represents an independent measurement and the mean of the three biological replicas is indicated with a bar, with values plotted as percent of signal of input. C1, NIPBL-non-binding region (based on NIPBL ChIP-seq data). H3, histone H3; IgG, control ChIP with non-specific antibody.

(D) Mean fluorescence intensity of Ramos clones with *GFP7* containing the candidate enhancer elements indicated below the graphs. Since absolute fluorescence intensity values can differ between datasets gathered at different times, data from the two different experiments performed are shown on separate graphs. Independent groups of cell clones containing *GFP7-ELF1e* were analyzed in the two experiments. Bars represent the median of the data; each data point is a separate clone.

(E) Plot showing comparison of GFP fluorescence loss and mean fluorescence intensity of independent *ELF1e-GFP7* Ramos clones. Each clone was sub-cloned and the median of GFP-fluorescence loss of 12 sub-clones was plotted.

(F) Enrichment analysis in the four active non-Ig enhancer elements versus the five with minimal DIVAC function. The signal for each parameter or factor (log(RPKM)) was compared between these two sets of enhancer elements (two-tailed T-test, with P-values corrected for multiple hypothesis testing). Corrected P-values are plotted. All data are derived from Ramos and include ChIP-seq for the indicated transcription or chromatin factors or modified histones, GRO-seq (GRO), and ATAC-seq (ATAC). NPC, nuclear pore components; PolII, total RNA Pol 2; S5P, serine-5-phosphorylated Pol2.

(G) GFP-fluorescence loss of Ramos clones infected with *GFP7* containing *ELF1e* or *ZCCHC7e* integrated in various integration sites compared to that of Ramos clones containing no-DIVAC-*GFP7* integrated in hot TADs. Bars represent the median of the data; each dot is a separate clone.

Figure S6

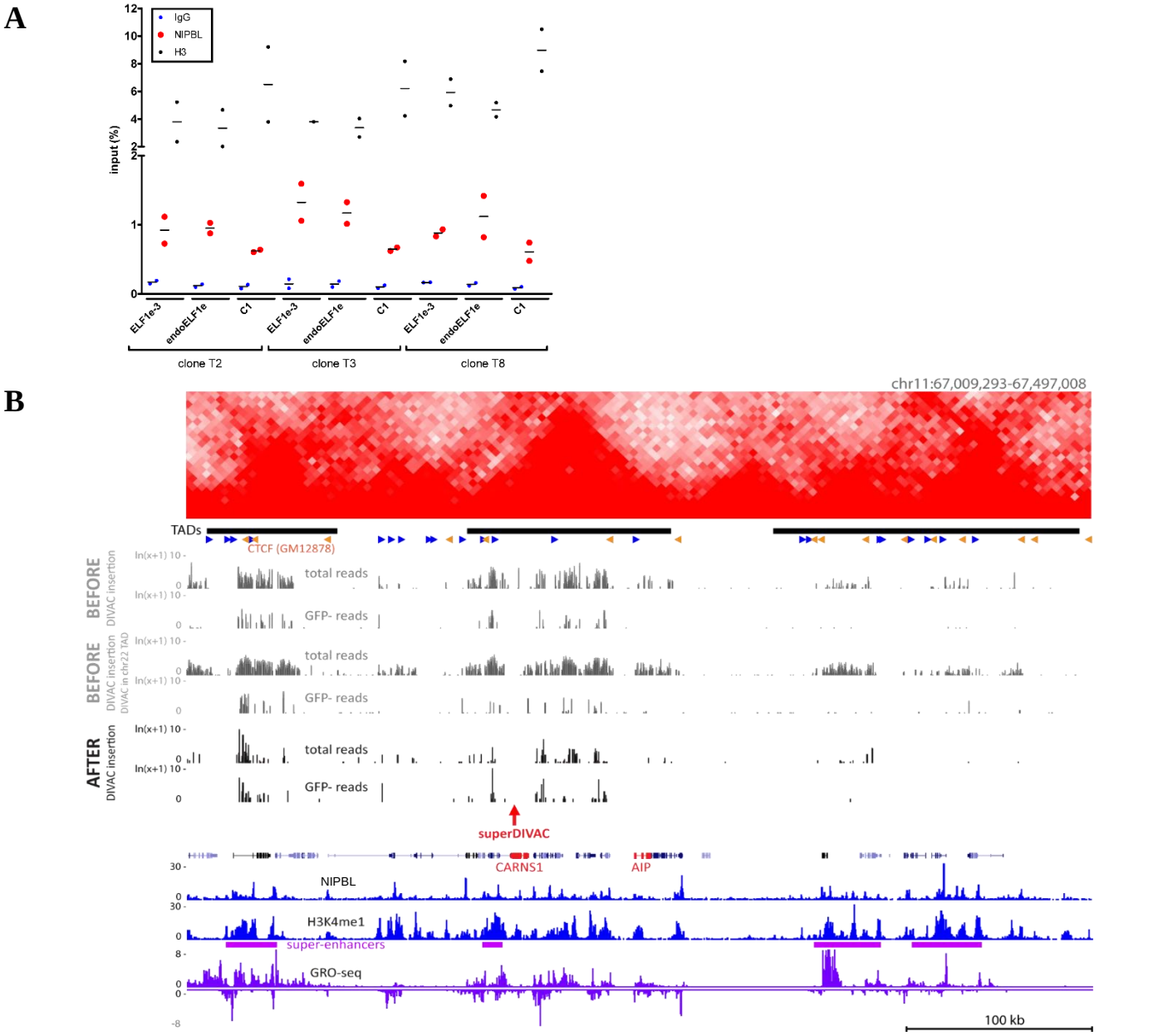

Figure S6. NIPBL binding to truncated *ELF1e*, DIVAC insertion transforms a cold TAD into a SHM-susceptible genomic region ***Related to Figures 6 and 7***

(A) ChIP-qPCR analysis of NIPBL binding in three independent Ramos clones harboring *GFP7-ELF1e-3* integrated in different cold TADs. Binding at endogenous *ELF1e* (*endoELF1e*) and truncated 250 bp vector *ELF1e* (*ELF1e-3*) was assessed using primers located within the major NIPBL peak; vector *ELF1e* contained two 5 bp substitutions to allow the design of primers specific for the ectopic *ELF1e* element. Each data point represents an independent measurement (average of duplicate technical replicas) and the mean of the two biological replicas is indicated with a bar, with values plotted as percent of signal of input. C1, a NIPBL-non-binding region (based on NIPBL ChIP-seq data). H3, histone H3; IgG, control ChIP with non-specific antibody.

(B) DIVAC-trap HTISA data, both before and after *superDIVAC* insertion, for the chromosome 11 region into which *superDIVAC* was inserted. Coverage, as assessed by read numbers in the total cell population, was much higher across the genome in the chr. 22 *superDIVAC* insertion experiment and in the experiment with unmodified Ramos than in the experiment with the chromosome 11 DIVAC insertion. The relatively weak coverage obtained in the chromosome 11 DIVAC insertion experiment makes it difficult to appreciate the strong increase in SHM susceptibility that occurs in targeted TAD, but this is captured quantitatively in Fig. 7B. Hi-C and other data presented as in Fig. S4. Red arrow, location of *superDIVAC* insertion.

Figure S7

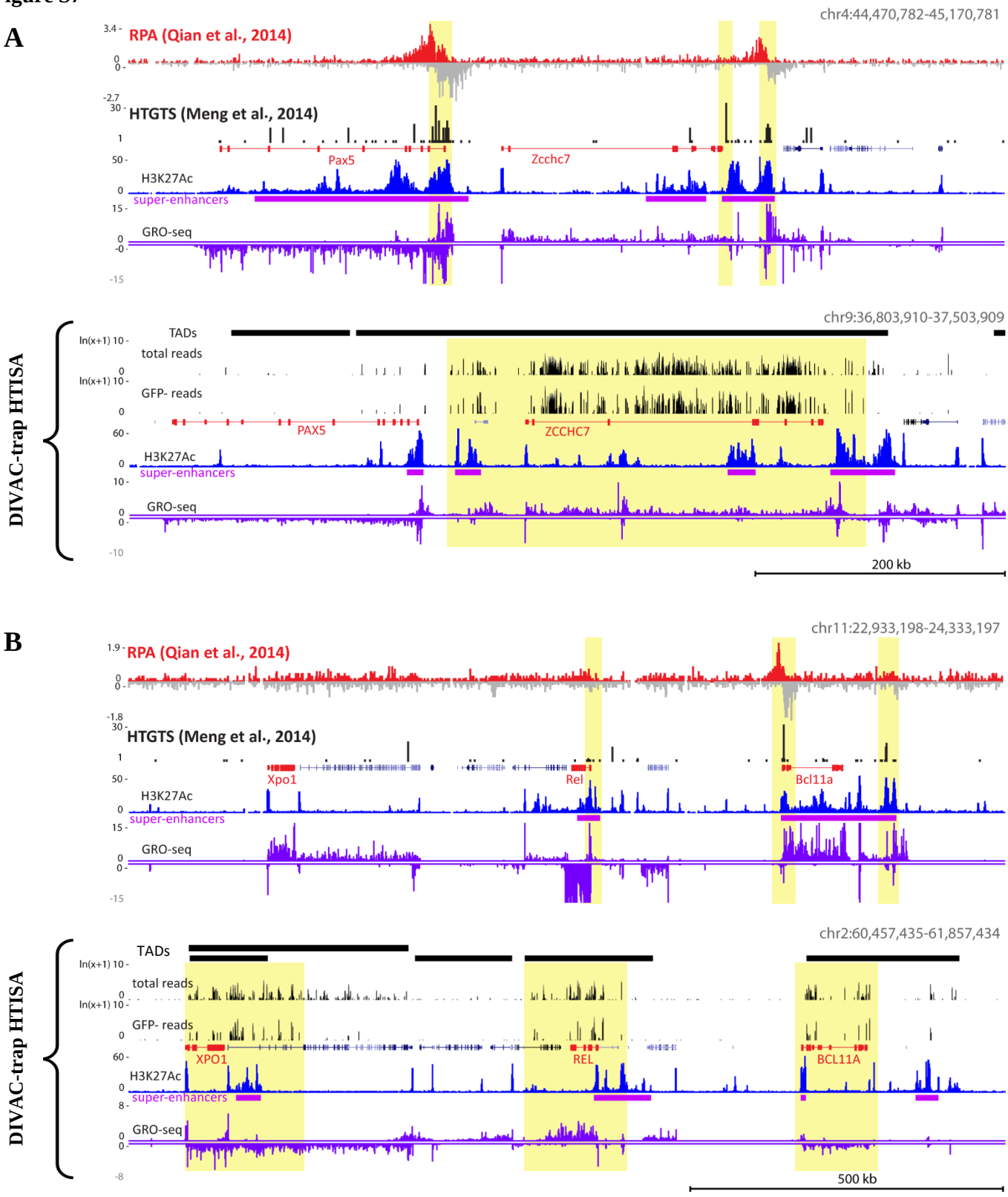

Figure S7. Comparison of DIVAC-trap HTISA assay data with previous studies *Related to Figures 2, 3 and 7*

(A, B) DIVAC-trap HTISA data for the human *PAX5* (A) and *REL* (B) loci from this study are shown aligned with AID targeting data obtained for the corresponding mouse loci in previous studies (Qian et al., 2014, Meng et al., 2014). Regions identified as AID-targeted (previous studies) or SHM-susceptible (this study) are highlighted with yellow. The regions of the human and mouse loci depicted are of equal size.