Supporting Information

Kato et al. 10.1073/pnas.1120791109

SI Materials and Methods

Cell Culture. The BL2 Burkitt's lymphoma cell line was cultured in RPMI medium 1640 containing 10% FCS (vol/vol), 100 μM NEAA (Gibco), 1 mM sodium pyruvate, 100 units per mL penicillin, and 100 μg/mL streptomycin. BL2-ΔC-AIDER cells, which are a BL2 clone harboring Jp8BdelER (Jp8Bdel fused with the hormone-binding domain of the estrogen receptor [ER]), were cultured with 0.5 μg/mL puromycin. To activate AIDER proteins, 4-hydroxytamoxifen (4-OHT) was added to a concentration of 1 μM.

Microarray. DNA was labeled according to the Affymetrix sample preparation protocol and hybridized to Human Promoter 1.0R Arrays. Slides were scanned with an Affymetrix GeneChip scanner. Six independent experiments were performed for each sample. Microarray data were analyzed using CisGenome.

Analysis of Microarray Data. Normalized signal intensities were generated from Affymetrix CEL files using CisGenome (1). The parameters used to detect peaks are as follows: moving average (MA) is used to combine neighboring probes, false discovery rates is estimated from permutation test, the window boundary is set to 250 bp, the MA cutoff is set to 3, the max allowable gap within a region is set to 200 bp, the max run of insignificant probes with a region is set to 5, the minimum region length is set to 150 bp, and the minimum number of significant probes within a region is set to 5. We retained only peaks where the FDR is <0.3. All probe sequences were mapped to the hg19 genome assembly.

SOLiD Sequencing. Templated microbeads were prepared according to Applied Biosystem's standard protocol. Sequencing runs were performed on the SOLiD 3 Plus system (Applied Biosystems) under standard conditions.

Analysis of SOLiD DNA-Sequencing Data. Linkers were removed from the sequenced tags by using a custom Perl script, and the resulting tags were mapped to the human genome (assembly hg19) using the Bowtie program with standard parameters (2). Breakage points were summed over 100-bp intervals across the entire genome in 10-bp increments for each control and 4-OHT-treated replicate. Significant differences in breakage frequency were calculated using the EdgeR program (3), with background dispersion values for both conditions calculated over shifting 10,000-bp intervals. This program provided *P* values and FDR values for each 100-bp interval, measuring the likelihood of observing the differences in breakage points across the two conditions given the selected background breakage rates. Regions with low FDR values overlapping the promoter (defined as 500 bases upstream of the RefSeq-defined transcriptional start site) and gene definitions were extracted. To increase confidence in the potential AID targets, regions with significant *P* values were clustered via a single-linkage clustering procedure, which joined any regions within 1 kb of another peak on the genome. The resulting clusters of regions were then overlaid with low-FDR regions, and the most promising candidate genes containing the lowest FDR values (FDR < 0.1), high numbers of *P* value clusters, or both were selected for further validation testing (Table S2). This layer of additional clustering was chosen to facilitate the testing of genes with the highest likelihood of breakage. Indeed, both the *MALAT1* and *SNHG3* loci (Fig. S3*A*) displayed these patterns, which were likely to be consistent with high levels of AID-induced cleavage activity.

Statistical Parameters of SOLiD DNA-Sequencing Data. In terms of data integrity, our technique is consistent with existing technologies. A summary of the characteristics of the sequenced libraries is provided in Table S8. The read redundancy ranges between 1.23 and 1.28; this redundancy rate is slightly higher than most ChIP-seq experiments (1.05–1.15) but much lower than other sequencing technologies. The slightly higher redundancy rates likely result from high numbers of breakages found in repeat regions (Fig. S5*A*). Importantly, the low redundancy values argue against problems arising from the number of cycles used, indicating the library does not suffer from amplification bias. An overview annotating the genomic locations of breakage sites in each library is also provided (Fig. S5*B*). The source of the discrepancies in target identification between this dataset and others instead likely stems from lack of depth. Each sequenced library covers the entire genome at roughly 0.8× coverage. Although the coverage of the as-currently undefined AID-targeted "break-ome" is not known, at present coverage levels, it appears possible to detect regions that are substantially affected by AID-induced breakage while not capturing everything that has been identified.

We have plotted the distribution of the square root of the absolute common dispersion values across the two conditions in all local genome windows which were used to approximate local background breakage rates for comparisons across samples (Fig. S5*C*). Very few genomic regions exceed an expected error rate of 20% in the measurements. We observe a bimodal distribution in the error rates, with the first peak corresponding to extremely low expected error (<1%) and the second peak found between ≈1–20%. The bimodal distribution implies that the AID breakage in a subset of genome regions is more highly reproducible relative to other regions; interestingly, this is not derived from relative expression values in the two peaks. Regardless, in either peak region the error rates are consistent with existing deep-sequencing techniques and enable robust identification of significant differences between breakage counts across conditions.

Mutation Analysis. To analyze SHM mutations, BL2-ΔC-AIDER cells were treated with 4-OHT (1 μM) for 24 h, and the genomic DNA was purified by phenol:chloroform extraction. PCR was performed by using *Pyrobest* or PrimeSTAR GXL DNA polymerase (TaKaRa) with the following amplification conditions: 95 °C for 5 min, 30 cycles at 95 °C for 30 s, 58 °C for 30 s, and 68 °C for 1 min. After purification, the PCR fragments were A-tailed and cloned with the pGEM-T Easy Vector System (Promega). Nucleotide sequences were determined with the ABI PRISM 3130xl Genetic Analyzer (Applied Biosystems). Only unique mutations were counted, and the mutation frequency was calculated from the number of mutations per total bases analyzed.

REPFIND Analysis. Analysis with the REPFIND Web server (http://zlab.bu.edu/repfind/form.html) used the following parameters: *P* value cutoff, 0.0001; minimum repeat length, 3; maximum repeat length, infinity; low complexity filter, on; statistical background, query sequence; order of background Markov model, 1.

ChIP. ChIP was performed as described (4).

1. Ji H, Wong WH (2005) TileMap: Create chromosomal map of tiling array hybridizations. *Bioinformatics* 21:3629–3636.
2. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Searching for SNPs with cloud computing. *Genome Biol* 10:R25.
3. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140.
4. Stanlie A, Aida M, Muramatsu M, Honjo T, Begum NA (2010) Histone3 lysine4 trimethylation regulated by the facilitates chromatin transcription complex is critical for DNA cleavage in class switch recombination. *Proc Natl Acad Sci USA* 107: 22190–22195.

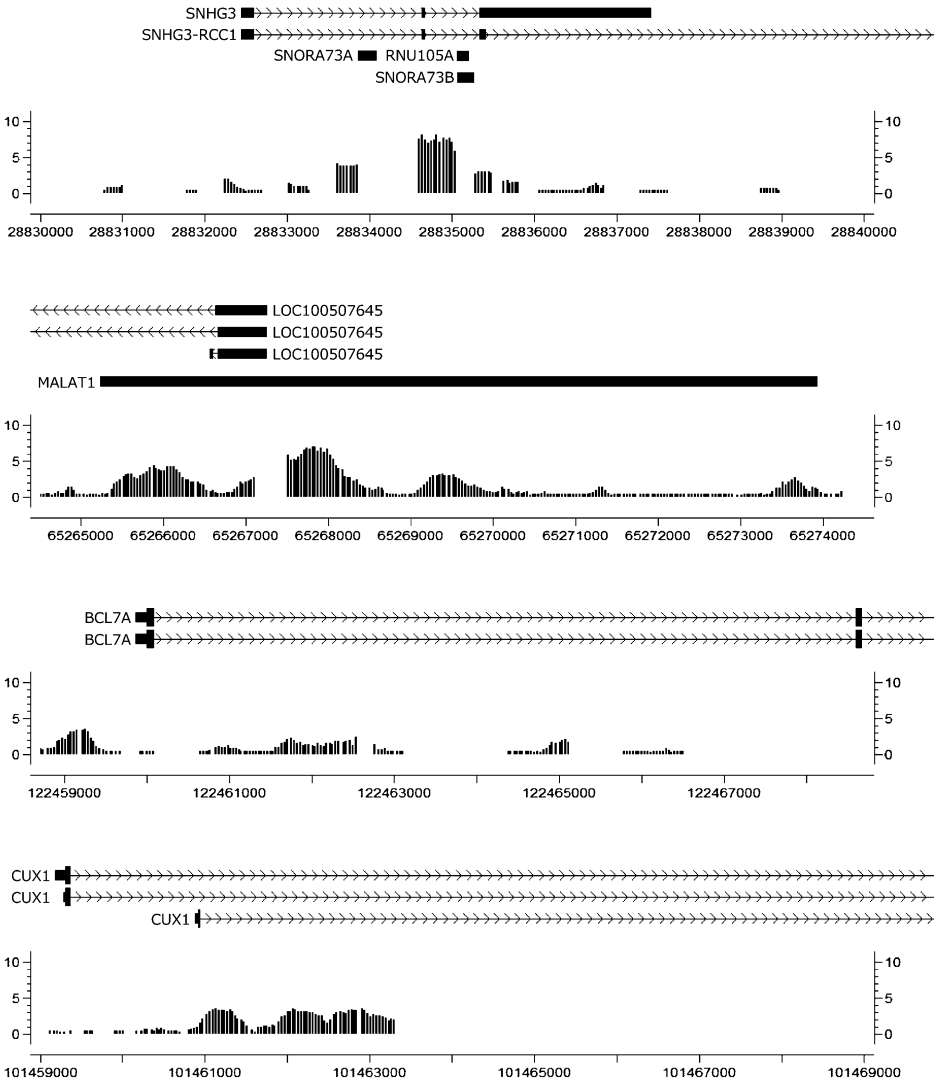

Fig. S1. Breakage signal distribution detected by promoter array. A 10-kb segment in the vicinity of detected breakage region is represented for each locus. Regions without array bars do not have array probes. *x* axis numbers indicate base positions according to hg19 assembly.

MYC

```
GGTGAAAGGGTGCTCCCTTTATTCCCCCACCAAGACCACCCAGCCGCTTTAGGGGATAGCTCTGCAAGGGGAGAGGTTCGGGACT

GTGGCGCGCACTGCGCGCTGCGCCAGGTTTCCGCACCAAGACCCCTTTAACTCAAGACTGCCTCCCGCTTTGTGTGCCCCGCTCC

                        T           A                   G       T
AGCAGCCTCCCGCGACGATGCCCCTCAACGTTAGCTTCACCAACAGGAACTATGACCTGCACTACGACTCGGTGCAGCCGTATTT

T                 A  T       A  A      A      AT T           T       A       A
CTACTGCGACGAGGAGGAGAACTTCTACCAGCAGCAGCAGCAGAGCGAGCTGCAGCCCCCGGCGCCCAGCGAGGATATCTGGAAG
          G
          T                                                G       C
AAATTCGAGCTGCTGCCCACCCCGCCCCTGTCCCCTAGCCGCCGCTCCGGGCTCTGCTCGCCCTCCTACGTTGCGGTCACACCCT

    T                     T           T                T       A
TCTCCCTTCGGGGAGACAACGACGGCGGTGGCGGGAGCTTCTCCACGGCCGACCAGCTGGAGATGGTGACCGAGCTGCTGGGAGG

                                                                          C     A   T
AGACATGGTGAACCAGAGTTTCATCTGCGACCCGGACGACGAGACCTTCATCAAAAACATCATCATCCAGGACTGTATGTGGAGC
                      G
                     CT                T                AA  A                T         T
GGCTTCTCGGCCGCCGCCAAGCTCGTCTCAGAGAAGCTGGCCTCCTACCAGGCTGCGCGCAAAGACAGCGGCAGCCCGAACCCCG

CCCGCGGCCACAGCGTCTGCTCCACCTCCAGCTTGTACCTGCAGGATCTGAGCGCCGCCGCCTCAGAGTGCATCGACCCCTCGGT

GGTCTTCCCCTACCCTCTCAACGACAGCAGCTCGCCCAAGTCCTGCGCCTCGCAAGACTCCAGCGCCTTCTCTCCGTCCTCGGAT

TCTCTGCTCTCCTCGACGGAGTCCTCCCCGCAGGGCAGCCCCGAGCCCCTGGTGCTCCATGAGGAGACACCGCCCACCACCAGCA

GCGACTCTGGTAAGCGAAGCCCG
```

SNHG3

```
ATGATGATAGGTAGCTGGAGGAAGGAGAATCGCTGGAGCCCAGGAGTGACCTATACTCAAACCTATACTCCAGTGCCACTGTACT
                                                                          C   T
                                                      T  T
                A                                 A    A  A  G
CCAACCCCAGGCGATAGCATGAGGCCCCTCGTTGAAAAAGTTTAGGGTTTTGCTGTACTAATAGATTAATATCTTGTTTTGCAGG

                        T           T  T            T  T        T
ATTTGTTAAGGATTCCAAGTAACTCTTATTTGGTGAGTAAATCTGCTAATTGTTTTTTGCTTATCAGCTCTTTGTCAATGATTTC

TGTAATGGAAATAGGATTGAAGAGACTTTTATTCTAGTTGGTCAGGATTTACCTCTGAGGCATTTAATCATTCTCAGAGCAATAG

CCAAATATCGACTTTGCTGCATTTTTGTAGGCATGTTGACATAACTTCAACATATGCTCTGTTCTGTAAAAATTGCTTTTTTTAG

TCAGCTCATTAAAAGTGCAAAGTAGTAAAAGCTGCCCTAGTGAACTGTAGGAAGCCTAATTGGCTTTATCTACATGTGTAGCCTG

  T         T                                                      A A
AGCTGAGAAAGATACTAGCCCTTGAAAATACTGTGGGTGATTAGCAATATTGGATTTGTCGGTTACTCCAATTCCTCACTAATGA

            _ A AT                                          C
GCATTCCAACGTGGATACCCTGGGAGGTCACTCTCCCCAGGCTCTGTCCAAGTGGCATAGGGGAGCTTAGGGCTCTGCCCCATGA

TGTACAGTCCCTTTCCACAACGTTGAAGATGAAGCTGGGCCTCGTGTCTGCGCCTGCATATTCCTACAGC
```

Fig. S2. (Continued)

MALAT1

```
ATAGAAGATAGAAAAATATAAAGCCAAAAATTGGATAAAATAGCACTGAAAAAATGAGGAAATTATTGGTAACCAATTTATTTTA

AAAGCCCATCAATTTAATTTCTGGTGGTGCAGAAGTTAGAAGGTAAAGCTTGAGAAGATGAGGGTGTTTACGTAGACCAGAACCA

ATTTAGAAGAATACTTGAAGCTAGAAGGGGAAGTTGGTTAAAAATCACATCAAAAAGCTACTAAAAGGACTGGTGTAATTTAAAA

AAAACTAAGGCAGAAGGCTTTTGGAAGAGTTAGAAGAATTTGGAAGGCCTTAAATATAGTAGCTTAGTTTGAAAAATGTGAAGGA

CTTTCGTAACGGAAGTAATTCAAGATCAAGAGTAATTACCAACTTAATGTTTTTGCATTGGACTTTGAGTTAAGATTATTTTTTA

AATCCTGAGGACTAGCATTAATTGACAGCTGACCCAGGTGCTACACAGAAGTGGATTCAGTGAATCTAGGAAGACAGCAGCAGAC

AGGATTCCAGGAACCAGTGTTTGATGAAGCTAGGACTGAGGAGCAAGCGAGCAAGCAGCAGTTCGTGGTGAAGATAGGAAAAGAG

TCCAGGAGCCAGTGCGATTTGGTGAAGGAAGCTAGGAAGAAGGAAGGAGCGCTAACGATTTGGTGGTGAAGCTAGGAAAAAGGAT

TCCAGGAAGGAGCGAGTGCAATTTGGTGATGAAGGTAGCAGGCGGCTTGGCTTGGCAACCACACGGAGGAGGCGAGCAGGCGTTG

TGCGTAGAGGATCCTAGACCAGCATGCCAGTGTGCCAAGGCCACAGGGAAAGCGAGTGGTTGGTAAAAATCCGTGAGGTCGGCAA

TATGTTGTTTTTCTGGAACTTACTTATGGTAACCTTTTATTTATTTTCTAATATAATGGGGGAGTTTCGTACTGAGGTGTAAAGG

GATTT
```

BCL7A

```
CTGCTGTCGGGCTGTCATCTCCAGGCCCGGGGCGCTGACATTTGGGCCACTCTCGGTCTCCCTCTTCATTCTGGGCGCGCATTAG

CTCTGGTCCGGCCGGTTCCGCTGCAGCTGAACAGCAAGATGCGGCACCCAGGTTACCCTGATCATCGCAGATTTCTCCCCGGGGC

TCTGTTCTGAGGCCTCAAAAGTGCTCCTTGTAGATGGGACCAGGGGTCATTTGGGCAGTAGCAGCGCCTGGTCTCAGTCTGGTAC

TGAAGTCAGGAATGGCTTAAGGTGAAATCGTGGTCCTCTGGTGAAGCTCAGCGAAGACCCCCTCGCCTTGTTTATGACAAGAGAA

CTTCTGGGGGCGGGAGGAAGAGTCCCTGTTACGATGCTGATCATCATTGAGCTTTTGCTGAGCAGAAAACTCTTTAGTACTCAAG

GTCGAGAGTCTCTGGTGGTCTGCCTGGCACCAGGCACCTTCCTACAACCCTAGTTTTCCAAAAGGACAAAGCCTGGGGCAGGCGA

CGTCCTAGCTCGCATTTGAACAGGGCCGCGGGCCAGCAGAGATGCGCGATGCCCAACTCTTTCCAAGAGCACCTCGCGTCCCGAA

CCGGTGCCTTCAACTCGGAGAAGTCAAGAGACCCGCAAGAAACTTGCACGACTGCACC
```

CUX1

```
GGTGAGCGGCGTGTGGGCCAGAAGTCCCGAGGTTGCAGGCGCGGAGGGAACCGGGGATGTCGGGGGGTGCCCGGGTCCCGCGGCT

TAGAATGCTCTAGGGCGGCCTGGTGCTCTGGGAGGGGATAGGAGGGTTCCTCAGGGCCCCTGGGGAACTGCAGCTACTCCCAACT

GCTAAGGTGTGCGTGAAGATAAACTTGGCTGAACTTTCCTAACCTGGAGGACTGGTGACAGTGACCTACCGCAATACCTTTGGGA

GCAAAGCTCGAGATGCAGGCTTGTATCAAACGAAGCGGGTTGGATTAAAACATATTTAAATGGAATCTGACAACTTTAATTGTAT

ATGCTTGTTTTCTGCTCTTGCCTACTAAGAACGATAAAGCCGAGTCATAGTCGTTATGAAATTTCCTGAAATTTCTCGCTTAACT

AGGAAGAAGACTATGCGAAATATGTATTTCCGATAAGATTAAAACTTAAAGGAGTTTAAGTATTTATTGACTAAAATAGATTACT

CCAAAGTGTCTATTGTGTAAATTAGATTCCTGGATGTAATGAACACAGCCGAATGGCATTTTTGATAAATTGGTCTACCCTGTTT

TAGTAAAAATAGCTCCCTTTTACCATAATTTAATTCCTGTGCTTACTGTAAAACCCGTTGTGGAGTTTTAAGGACAAACATTATT

ATTATTTTTCTCTCTTGTCTTCCTAAATGTGACATCCTAGCTTAA
```

Fig. S2. Somatic mutations and breakpoint distribution in AID target loci. Red, mutations found in cells treated with 4-OHT for 24 h; green, mutations in 4-OHT nontreated samples; blue, insertions; red line, deletions; arrowheads, break sites in samples treated with 4-OHT for 3 h; arrows, translocation breakpoints shown in Fig. 3.

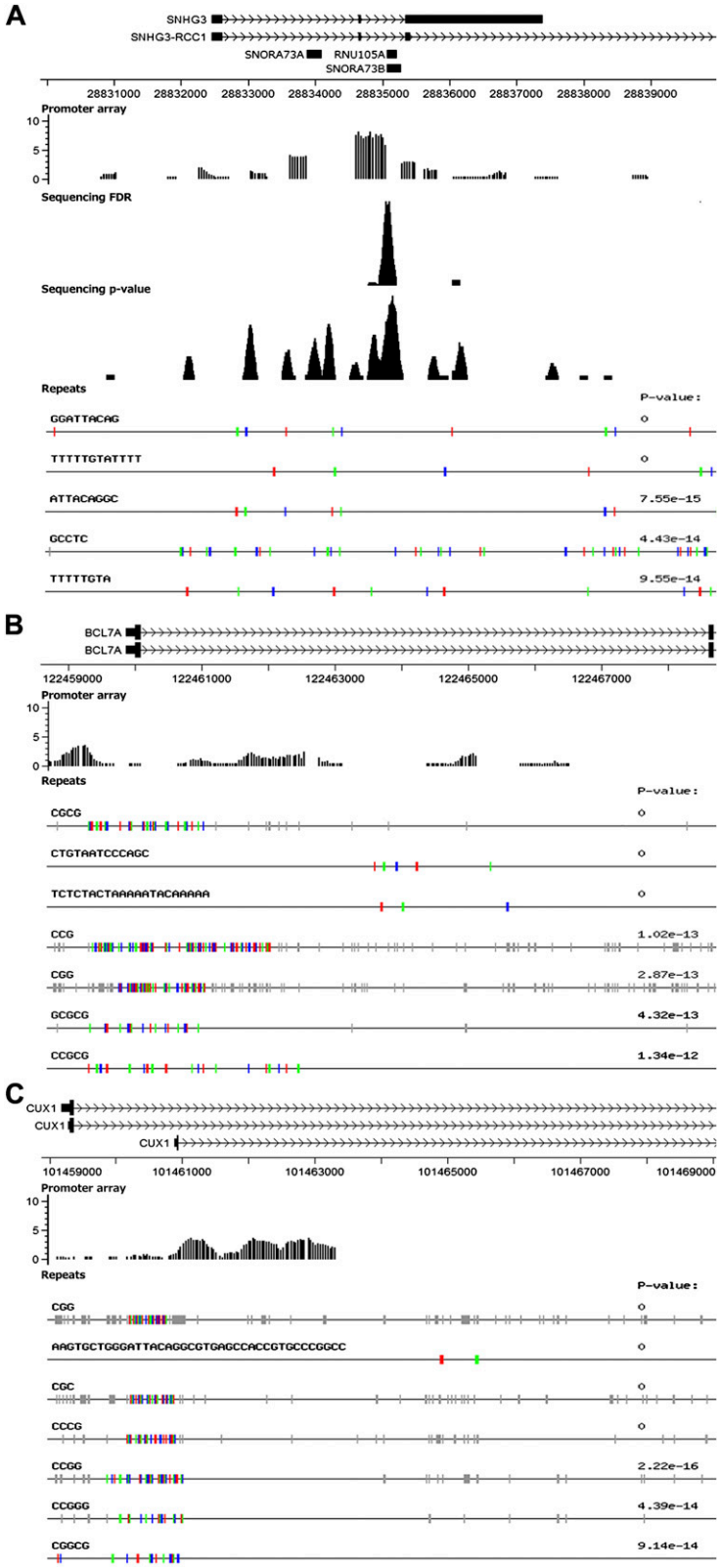

Fig. S3. Repeat sequences surrounding the breakage region in AID target genes. (*A*) *SNHG3* gene. From top to bottom: Representation of a 10-kb segment surrounding the *SNHG3* locus; Breakage signal distribution detected by promoter array (regions without bars do not have array probes); FDR regions by sequencing; *P* value peaks by sequencing; REPFIND analysis showing significant repeat clusters in the *SNHG3* locus. (*B* and *C*) *BCL7A* and *CUX1* genes, respectively. From top to bottom: Representation of a 10-kb segment surrounding the cleavage region for each gene; Breakage signal distribution detected by promoter array (regions without bars do not have array probes); REPFIND analysis showing significant repeat clusters in the same region. Motifs depicted as vertical small colored bars indicate the cluster with the most significant *P* value; individual repeats are separated by different colors. *x* axis numbers indicate base positions according to hg19 assembly.

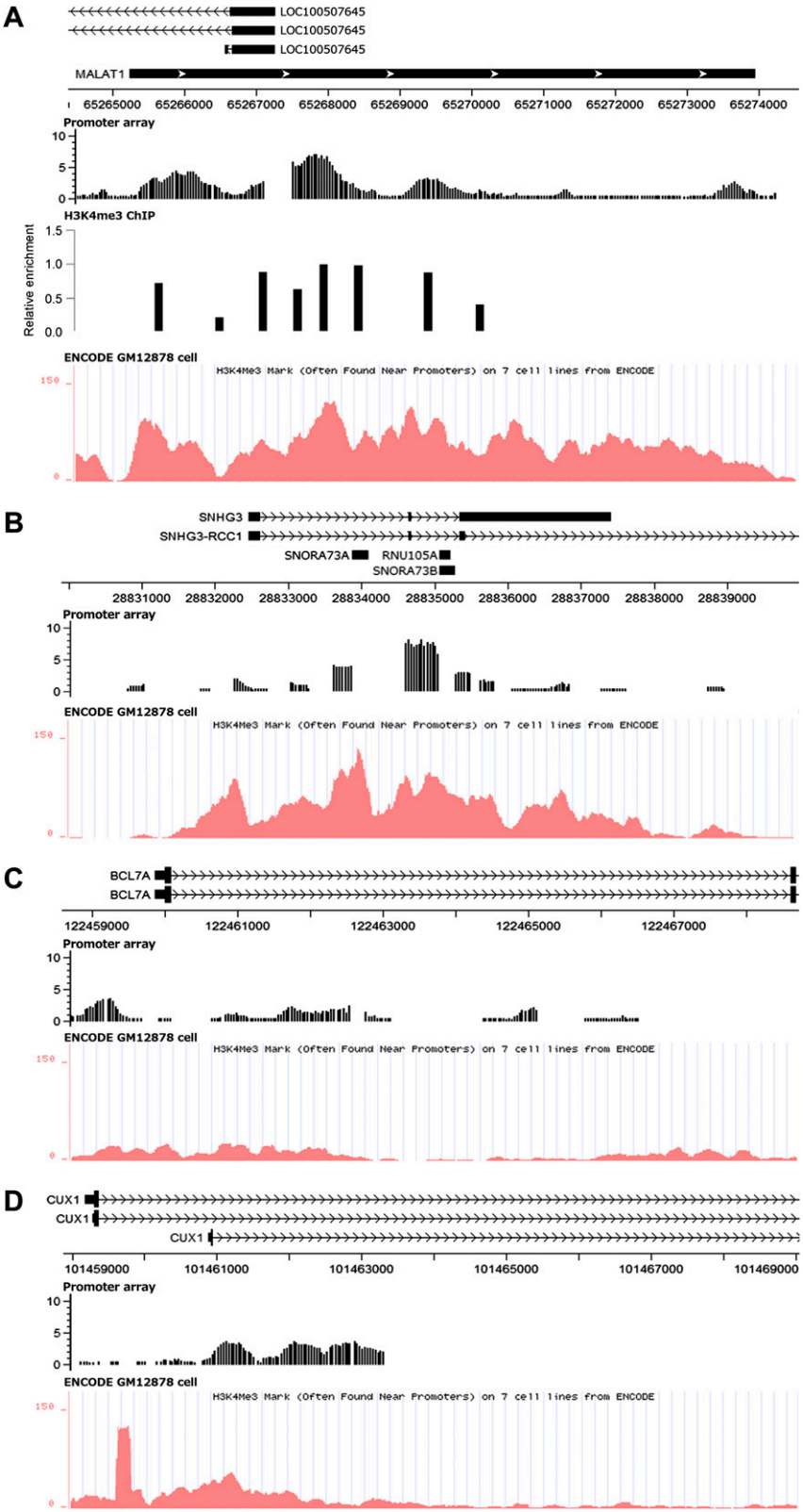

Fig. S4. H3K4me3 distribution in AID target genes. (*A*) *MALAT1* gene. From top to bottom: Representation of a 10-kb segment surrounding the *MALAT1* locus; Breakage signal distribution detected by promoter array (regions without bars do not have array probes); ChIP assay using an anti-H3K4me3 antibody; H3K4me3 status from ENCODE ChIP-seq data for GM12878 cell line. (*B–D*) *SNHG3*, *BCL7A* and *CUX1* genes, respectively. *x* axis numbers indicate base positions according to hg19 assembly.

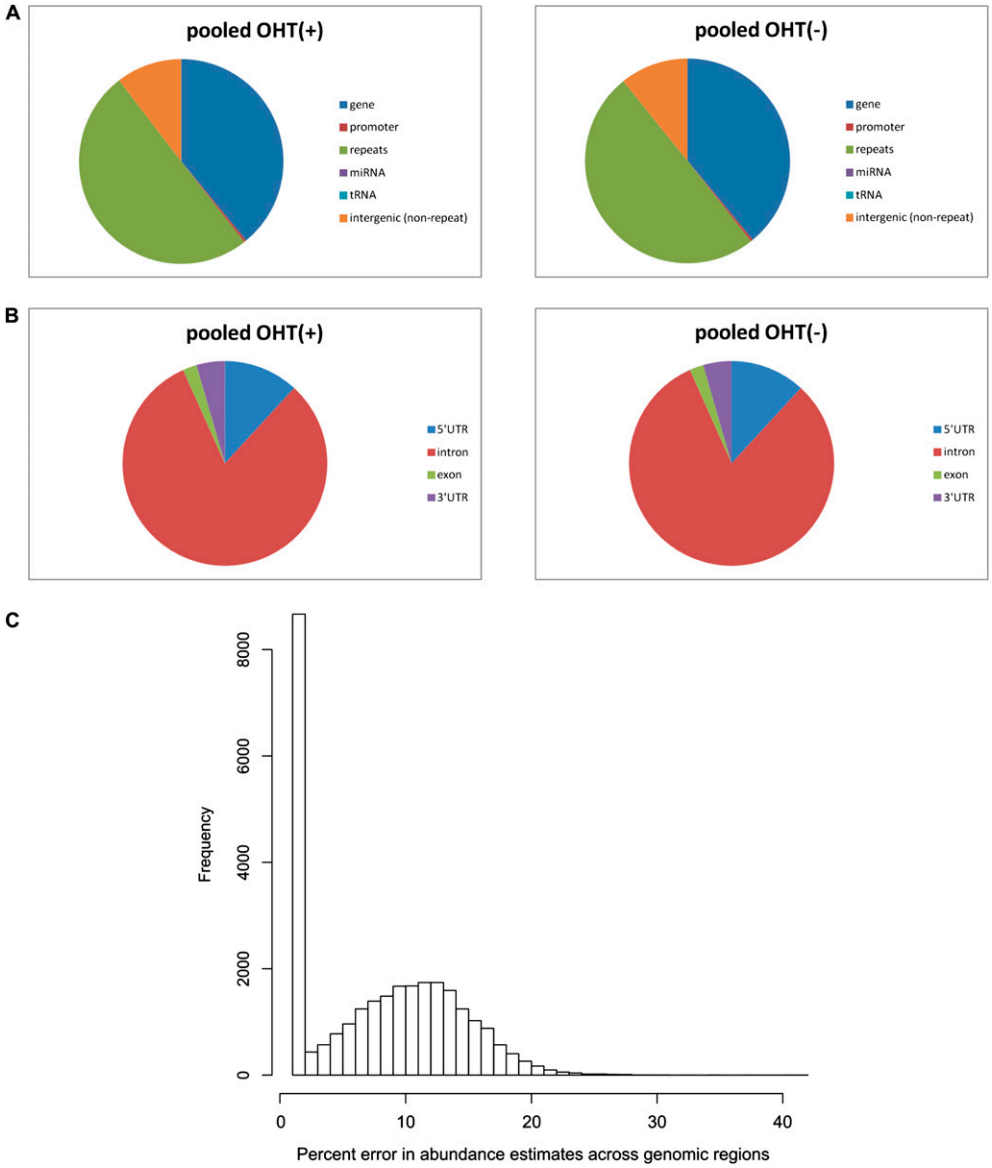

Fig. S5. Location of all mapped reads according to genome features (*A*) and distribution of gene-mapping reads to genomic locations (*B*). (*C*) Distribution of estimated percent error in abundance measurements across all analyzed genomic segments. The estimated percent error in breakage rates rarely exceeds 15–20%, enabling robust determination of significant differences.

Table S1. AID targets identified by promoter assay

Gene	Chromosome	Expression	Tilemap maxM/P	FDR value	Validation by qPCR	Translocation (partner gene)
SNHG3	chr1	+	8.18	0	+	CML (*PICALM*)
MALAT1	chr11	+	7.05	0	+	Renal cell carcinoma (*TFEB*), mesenchymal hamartoma of the liver (*ACAT2*)
NIN	chr14	+	6.25	0.08	—	CML-like myeloproliferative disorder (*PDGFRB*)
FYB	chr5	—	6.25	0.08	n.d	—
C2orf16	chr2	—	6.05	0.1	—	—
C9orf72	chr9	+	5.93	0.11	—	—
FAM119B	chr12	—	5.62	0.2	n.d	—
CFLAR	chr2	+	5.47	0.26	+	—
SNX25	chr4	+	5.45	0.26	—	—
BCL7A	chr12	+	3.53	0.83	+	Burlitt lymphoma (IgH; *MYC*-IgH)
CUX1	chr7	+	3.51	0.83	+	TLL (*FGFR1*)

FDR < 0.3 plus *BCL7A* and *CUX1*. Chr, chromosome; CML, chronic myeloid leukemia; n.d, not done; TLL, T-lymphoblastic leukemia/lymphoma.

Table S2. AID targets identified by sequencing

Gene	Chromosome	Expression	FDR value	*P* value clustering*	Validation by qPCR	Translocation (partner gene)
MALAT1	chr11	+	0.027	12	+	Renal cell carcinoma (*TFEB*), mesenchymal hamartoma of the liver (*ACAT2*)
SNHG3	chr1	+	0.015	9	+	CML (*PICALM*)
SIPA1L3	chr19	+	0.015	5	—	—
KCNC2	chr12	—	0.005	4	—	—
ZNF451	chr6	+	0.038	4	—	—
TRIO	chr5	+	0.050	4	—	—
C5orf13	chr5	+	0.072	4	—	—
CUL9	chr6	+	0.077	4	—	—
TM9SF4	chr20	+	0.026	3	—	—
ANKRD11	chr16	+	0.041	3	—	—
MYO3B	chr2	—	0.082	3	—	—
UPF2	chr10	+	0.093	3	—	—
MET	chr7	—	0.000	2	—	Gastric carcinoma (*TPR*)
AUTS2	chr7	+	0.017	2	—	ALL (*PAX5*)
RAD18	chr3	+	0.038	1	—	—
OTUD6B	chr8	+	0.041	1	n.d	—
VPS13B	chr8	+	0.044	1	—	—
CCDC41	chr12	+	0.055	1	—	—
ECT2	chr3	+	0.078	1	—	—
MRPL49	chr11	+	0.020	1	—	Close to *t*(11;17)(q13;q21) translocation in B-NHL
NECAB3	chr20	+	0.029	1	—	—
SETD8	chr12	+	0.175	1	—	—
SETBP1	chr18	—	0.134	9	—	—
PBLD	chr10	—	0.254	8	—	—
ABCG2	chr4	—	0.220	8	—	—
FAM65B	chr6	+	0.195	8	—	—
WBSCR17	chr7	—	0.162	8	—	—
ERC1	chr12	+	0.153	6	—	—
PLD2	chr17	+	0.102	5	—	—

FDR < 0.1 and/or remarkable numbers of *P* value clusters.ALL, acute lymphocytic leukemia; B-NHL, B-cell non-Hodgkin lymphoma; Chr, chromosome; CML, chronic myeloid leukemia; n.d, not done.
*Number of *P* value clusters corresponding to the FDR region.

Table S3. Mutation analysis of genes with significant increase in break signal after AID activation

	4-OHT	Mutated clone	Unique mutations, bp	Total sequenced, bp	Mutation frequency ($\times 10^{-4}$)	del(ins) (bp)	Del(ins)/ total clones	*P* value
B2M	–	1/82	1	43,674	0.23	0	0/82	0.52
	+	2/87	2	46,338	0.43	1	1/87	
Sμ	–	13/82	17	67,404	2.52	1	1/82	4.2×10^{-7}
	+	38/79	59	64,938	9.09	33(1)	7(1)/79	
V region	–	18/123	21	92,127	2.28	6	5/123	0.001
	+	38/135	51	101,115	5.04	8(10)	4(1)/135	
MYC	–	5/83	5	41,251	1.21	0	0/83	1.8×10^{-7}
	+	21/80	33	39,760	8.30	0	0/80	
SNHG3	–	3/77	3	42,633	0.70	0	0/77	0.0001
	+	16/74	21	41,170	5.10	2	2/74	
MALAT1	–	3/89	3	48,950	0.61	0	0/89	8.1×10^{-7}
	+	19/90	30	49,500	6.06	83(1)	3(1)/90	
BCL7A	–	3/84	3	44,520	0.67	0	0/84	0.006
	+	9/82	14	43,460	3.22	0	0/82	
*CUX1**	–	0/89	0	48,149	0	0	0/89	0.001
	+	8/91	11	49,231	2.23	0	0/91	
CUX1[†]	–	13/85	14	58,905	2.38	0	0/85	0.007
	+	22/84	31	58,212	5.33	4	4/84	
CFLAR	–	0/134	0	87,404	0	5	1/134	0.004
	+	5/133	8	86,878	0.92	1	1/133	

Cells were treated with or without 4-OHT for 24 h. *P* values were calculated by one-sided Fisher's exact test.
*Region detected by promoter array (Fig. S2).
[†]Region with highest peak of H3K4me3 (Fig. S4*D*).

Table S4. Linker sequences

	Linker primers (5′ to 3′)
Linker P1	biotin - TTCCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT
	ATCACCGACTGCCCATAGAGAGGAAAGCGGAGGCGTAGTGG
Linker P2	AGAGAATGAGGAACCCGGGGCAGTT
	CTGCCCCGGGTTCCTCATTCTCT
Global amplification	
(fwd)	CCACTACGCCTCCGCTTTCCTCTCTATG
(rev)	CTGCCCCGGGTTCCTCATTCT
P1-LM (LM-PCR)	CCACTACGCCTCCGCTTTCCTCTCTATG

Table S5. Gene-specific primers (5′ to 3′)

	Forward	Reverse
PCR		
Sμ	GCACAGGCTCCTAAATTCTTGGTC	CAGGCTGGCTTCCATCTTTTGTCT
B2M	CGGCTCTGCTTCCCTTAGAC	CGAAACCGCTTTGTATCACA
Mutation		
Sμ	GACATGGTAAGAGACAGGCAGCCG	GGATGGAGTTGTCATGGCCAGAAA
BL2 V region	ATCTCATGTGCAAGAAAATGAA	AGTCCCACCACGCAATCAT
MYC	CCCTCAACGTTAGCTTCACCAACA	CGCTCAGATCCTGCAGGTACAA
SNHG3	GCCCAGGAGTGACCTATACTCAAA	GGTATCCACGTTGGAATGCTCA
MALAT1	GGCAGAAGGCTTTTGGAAGA	CAACATATTGCCGACCTCACGGAT
BCL7A	ATTAGCTCTGGTCCGGCCGGTT	GGTGCAGTCGTGCAAGTTTCT
CUX1[a]	GGAGCCAGGTTGAAGGTGA	TGCCATTCGGCTGTGTTCATTA
CUX1[b]	GCTTGATCGGAAATTGATCCTC	GTCCGCGTCACCGACACAGG
CFLAR	CAGGGAAGTGTGTTAAGTGC	CATGTTGTCCTGAAGCCAGTGC
B2M	TCTCTTTCTGGCCTGGAGGCTAT	AGAGGTGCTAGGACATGCGAACTT

Table S6. Gene-specific primers LM-PCR (5′ to 3′)

	First round	Second round
MYC 5′	CCAAGCCGCTGGTTCACTAA	GAGATAGCAGGGGACTGTCCAAA
MYC 3′	GGCCCGTTAAATAAGCTGCCAA	ATCCAGCCGCCCACTTTTGACA
SNHG3 5′	GCCCAGGAGTGACCTATACTCAAA	ATGAGGCCCCTCGTTGAAAA
SNHG3 3′	ACGTTGTGGAAAGGGACTGTACAT	TAGTGAGGAATTGGAGTAACCGACA
MALAT1 5′	GCTTGAGAAGATGAGGGTGTTTA	GGCAGAAGGCTTTTGGAAGA
MALAT1 3′	CAACATATTGCCGACCTCACGGATT	ACACTGGCATGCTGGTCTAGGAT
BCL7A 5′	TGAGGCCTCAAAAGTGCTCCTTGT	ACCAGGGGTCATTTGGGCAGTA
BCL7A 3′	GGTGCAGTCGTGCAAGTTTCT	GGTCTCTTGACTTCTCCGAGTTGA
CUX1 5′	GACTCTGCCAGGTGGATGTTG	GGAGCCAGGTTGAAGGTGA
CUX1 3′	AAATGCCATTCGGCTGTGTTC	TGCCATTCGGCTGTGTTCATTA

Table S7. ChIP and qPCR primers (5′ to 3′)

	Forward	Reverse
ChIP		
V region 1	TCACCTAGGCGCCCACAGGAA	CGCCACCAGCAGGAGGAAGA
V region 2	ATCTCATGTGCAAGAAAATGAAGCACCTGT	CCCTGGGATCAGAGGCAGCCTCCCA
V region 3	CTCACTGTGGGTTTTTCTGTTCACA	GAGCCACCAGAGACAGTGCAAGTGA
V region 4	CATCAGCAGTACTAATTACTACTTGAGTTG	CGACTCTCGAGGGATGGGTTGTAGT
V region 5	GTGAAGTCTTCGGAGACCTT	ACATGGTGACTCGACTCTCG
V region 6	AGTCACCATGTCCGTAGACATGTCC	AGTCCCCCCCCTTCGAGCCACTGGT
V region 7	CTGGTTCGACTCCTGGGGCCAGGGA	ACACTCTGACCCCGAGACCCTGGCA
V region 8	TGGAGGCATTTTGGAGGTCAGGAAA	CCAGCCGAAGGAGCCCCCCAGCTGC
Cμ	CTTCCTTCCCGACTCCATCAC	CGTTCTTTTCTTTGTTGCCGT
MALAT1 1	AGAGCAGTGTAAACACTTCTGGGTG	TGGAAAGCGAGTTCAAGTGGCCT
MALAT1 2	AGGTGATCGAATTCCGGTGATGCGA	CAAGCTCCGCCTGCCCCCTCAGCA
MALAT1 3	CATTTACTAAACGCAGACGAAAATG	TTTCTTCGCCTTCCCGTACTTCTG
MALAT1 4	TTAGAAGGTAAAGCTTGAGAAGATG	AGTCCTTTTAGTAGCTTTTTGATGTG
MALAT1 5	TTCAGTGAATCTAGGAAGACAGCAG	CCTGGACTCTTTTCCTATCTTCACCA
MALAT1 6	GATTTCCGGGTGTTGTAGGTTTCTC	AAACCCACAAACTTGCCATCTACTA
MALAT1 7	TGGCAATTAGTTGGCAGTGGCCTGT	TCCATTCTAAGACTTTAAGTTCTCTG
MALAT1 8	TGTCTCTTAGAGGGTGGGCTTTTGT	GCATCTAGGCCATCATACTGCCAGGC
qPCR		
Sμ	GACTGCAGGGAACTGGGGTATCA	GGATGGAGTTGTCATGGCCAGAAA
BL2 V region	GTCAGAGTCTTGGAGGCATTTTGG	AATGCTCCAGGTGAAGCGGAGAGA
MYC	GCCGCCGCCTCAGAGTGCAT	CGGAGAGAAGGCGCTGGAGT
SNHG3	AAGCTGCCCTAGTGAACTGTAGGAAG	TAGTGAGGAATTGGAGTAACCGACA
MALAT1	AAAAGGATTCCAGGAAGGAGCGAGT	ACACTGGCATGCTGGTCTAGGAT
BCL7A	CGACGTCCTAGCTCGCATTTGAA	GGTGCAGTCGTGCAAGTTTCT
CUX1	TTCTTGTCCCTCGGCTTCCT	GCACTGAAACTTCCATACCACAA
CFLAR	AAGGGACAGGTGCAGAAAGAGTAT	CTCAACTCCAGCTGACACTGCTAAT
SIPA1L3	CTTTGCCAATGGATCTGTGTCTG	GCCGACTCAGGAACTGCTTG
KCNC2	GGGTCAGCCAATGCACCATTC	GATGCAACAGCCACTCAGTAG
ZNF451	CCTTGTTCAAGATGCTCTGAGTG	GCGCCAACATTTCAACAAGCAG
TRIO	GGAAATGAGGTCTCAGGGTTTAAG	TGTGGATGCTAAGGGAACTGAG
C5ORF13	CTCTTGCGACAGCAGTTTCC	GGTTTCTCCCTGTCAACATCAC
CUL9	GATCGCTGAGGTTAGCATACTG	TTCGTGATCTCAAAGCTCCTTC
TM9SF4	CCAGCCATGCAAAGAATGTTCC	CCTCCAGCCTTCTGTGTGTTC
ANKRD11	CACCAGATCACAGCATAAGCAC	TTTGTTTGGAGACCAGCCCTTTG
MYO3B	CTTGGTCCAACCCTTGTAGTTC	GACTACTTGAATGATGGGCACAG
UPF2	TTGCCTCTGTCAGCAATGCTC	AGAGAAGACTGCCTGGAACAAG
MET	ATGAGGCTTGAAGAGAGAGGACAAC	CACTCTGCCCTCTTCCAGTTC
AUTS2	AGCTCAAGCGATTCTCCCTC	CACTCAGCACTATACCAACCAC
RAD18	GAGCCATACCACGACTGTGC	TGCAGGGCAGTCAGTTTATTAGTG
VPS13B	GTTTCCTCCTACCCTACTCTAGC	TTCCAAAGGTGTCTGGGTGTATC
CCDC41	GCAAAGACAGTCAAGAGACAG	TGTGGACCTTCAGAATCCTATC
MRPL49	AGGCAATCATGGAGGTACAAAC	TGGTCTGCCTCTCAGGATTTC
NECAB3	TCACGTACCTGCCTACTCAC	TTGCTTTAGTTGCTGGCCCATC
SETD8	CTGAAACAGCCACCAGAGTGAC	AGCAAATGGTCTTGCAGAAGG
SETBP1	TCTGTACCTGTGTGTATCTTCCTG	CCCATAGGTGACAAGCACCATC
PBLD	GGCCCTGATCTTTGTCCATTAAAC	GCCCGGCCAATAAGAGCTTC
ABCG2	TGCCACTTTATCCAGACCTAACTC	ACTTACAGTTCTCAGCAGCTCTTC
FAM65B	AATTGCTGGCCCAGTGTAGTG	GTTTCAGTCTCTTGCCCAGG
WBSCR17	GCACTAGGTGCTGTGCATAAAC	TGCAAGCACCATGACTCAGC
B2M	CGGCTCTGCTTCCCTTAGAC	CGAAACCGCTTTGTATCACA
MAP2K4	TCTGGACATTTGAGGCAGCTCT	AAGGAAATGGTCCCTAACAGGCT
SH3KBP1	CACGTCGAGACCTGCCATTTA	GGGACTGCATGTTAGATGAGGA
TUBA1B	TCCATAACCTAGGGACTATCTGA	TTGGGTGGAGTGACTGACAT
RNMT	GCTGAGTCCTGAAACTTGT	GTTTGGGATTCTACAGCAAG
EXO1	GAGGATATTTGCCTGGCCCAGAA	GGTCCTCAAGCCACAGTTTCAGA

Table S8. Summary of general features of the sequenced libraries

Library	Total reads	Total mapped reads	Percent mapped reads	Unique mapping reads	Unique reads, %	Redundancy ratio
OHT(-) rep. 1	83,122,708	51,939,756	62.49	40,459,228	77.90	1.28
OHT(+) rep. 1	84,818,977	52,209,828	61.55	41,546,705	79.58	1.26
OHT(-) rep. 2	82,502,113	53,683,520	65.07	42,994,281	80.09	1.25
OHT(+) rep. 2	83,293,348	51,190,838	61.46	41,686,202	81.43	1.23