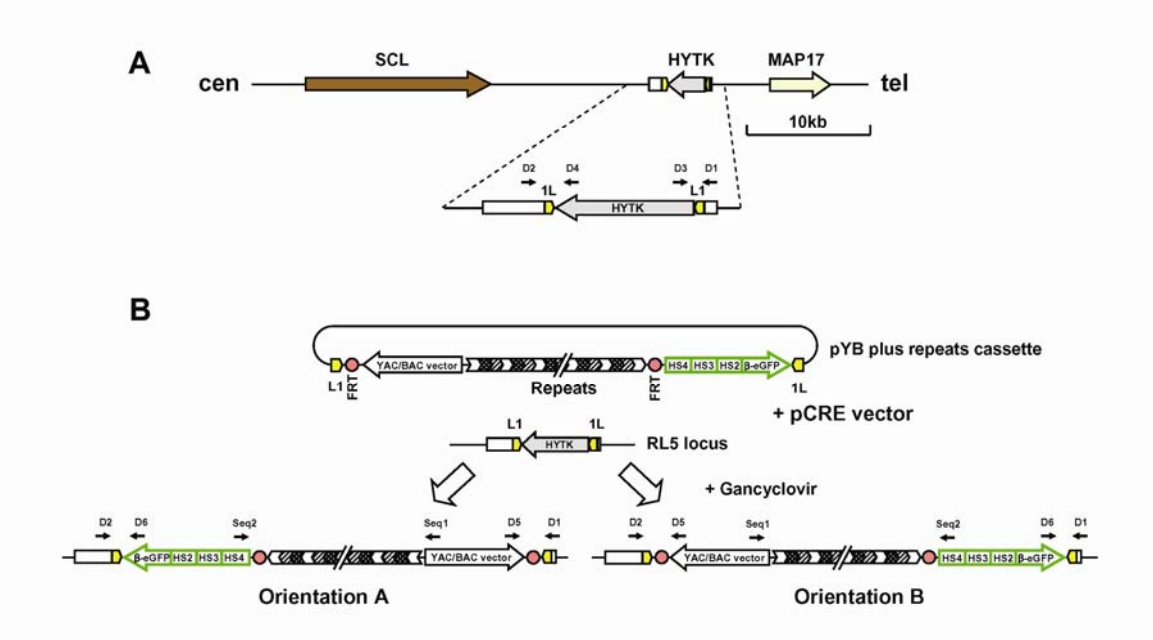


Figure S1. Physical Mapping of the RL5 Locus and a Scheme of Targeting of Amplified Repeats into the RL5 Locus

(A) The cassette containing the CMV-HYTK selectable marker (grey arrow) was previously integrated into the mouse chromosome 4 (NT_039264.4). In addition, the cassette contains ~1 kb of the L1-HYTK-1L vector DNA fragment (pBR replication origin) at the 3' of the HYTK gene region (centromere direction) and ~100 bp at the 5' end of the HYTK gene region (telomere direction) (white boxes). Mouse chromosomal flanking DNA was isolated using DNA Walking SpeedUp Premix kit (Seegene) according to manufacturer's instructions. To determine the DNA flanking sequences (upstream and downstream of the HYTK gene), nested plasmid specific primers and universal primers were used in PCR reaction that targeted unknown mouse sequences (Table S9). The correct locus was confirmed by PCR with specific chromosome 4 and L1-HYTK-1L plasmid primers (Table S9). The HYTK gene in RL5 cells maps to between 15176155 - 15176171 positions of the mouse chromosome 4 in the cytogenetic band qD1 (NT_039264.4). The L1-HYTK-1L integrated fragment is located between the SCL gene (~12.7 kb) and the Map17 gene (~4.5 kb). The insertion resulted in very little alteration in the structure of DNA at the site of integration, i.e. deletion of 15 bp endogenous DNA and insertion of 15 bp of unknown sequence to centromere direction and 6 bp to telomere direction.

(B) The cassette exchange of pYB plus repeats at MEL RL5 locus. The integrity of the insert was confirmed using the appropriate pairs of primers D1, D2, D3, D4, D5, D6, Seq1 and Seq2 (Table S9).

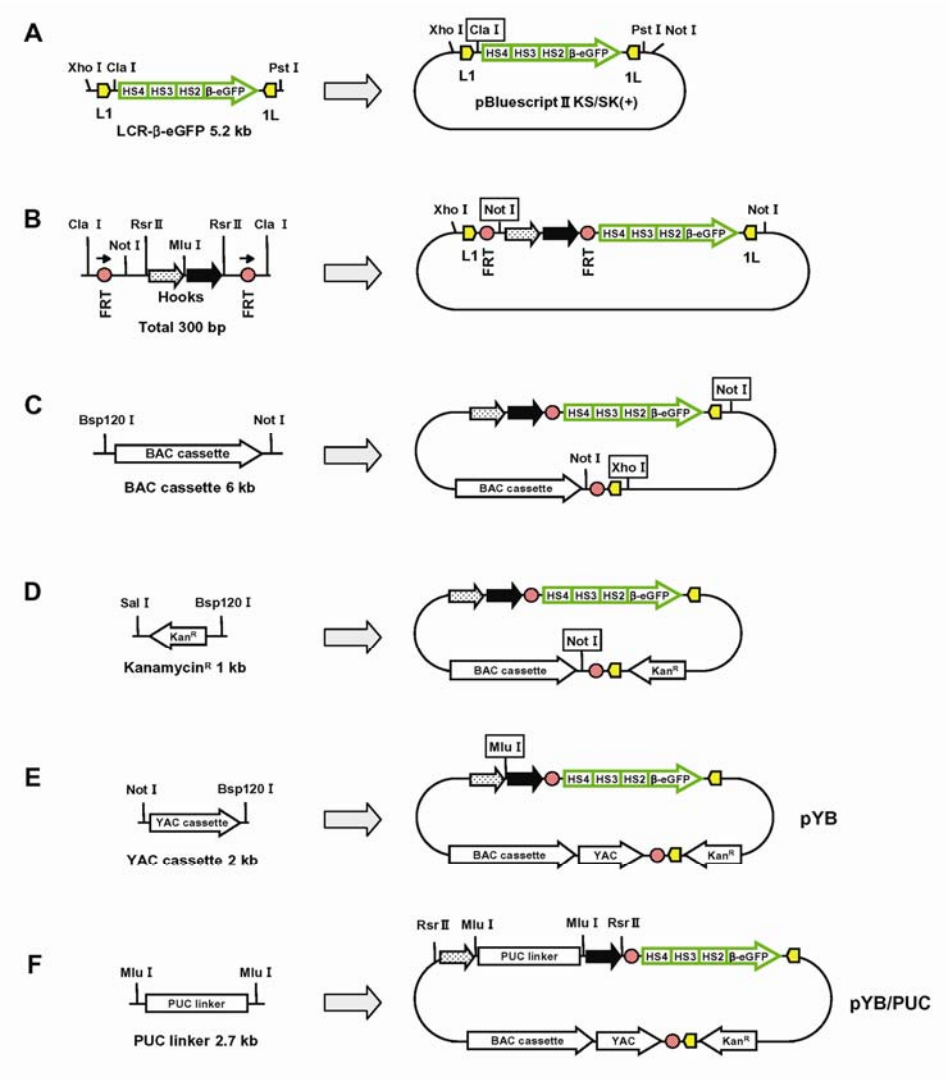


Figure S2. Construction of the pYB Cassette

(A) pYB vector was constructed as follows. 5.2 kb XhoI/PstI fragment from the p212 vector containing the human beta-globin LCR 234/promoter driving the eGFP reporter gene (LCR 234-eGFP) was cloned into the pBluescriptII KS vector (Invitrogen). This fragment has two inverted LoxP511 sites (L1 and 1L).
(B) The 'Hooks' cassette contains two FRT recognition sites and a MluI internal restriction site. A 300 bp ClaI/ClaI PCR fragment was cloned into the ClaI digested vector from step (A). The 'Hook' cassette was inserted upstream of LCR 234-eGFP.
(C) A 6 kb BAC cassette was inserted into the NotI site of the vector constructed in step (B).
(D) A 1 kb SalI/Bsp120I fragment containing the kanamycin gene (Kan^R) was inserted into the XhoI/NotI site of the vector from step (C).
(E) A 2 kb NotI/ Bsp120I fragment containing a YAC cassette (ARS, HIS3 and CEN6) was inserted into a NotI site downstream of the BAC cassette sequence.
(F) A 2.7 kb PUC linker from pBACe3.6 vector was inserted into the MluI site of the vector from step (E) to confer a high-copy number to the plasmid. The appropriate primers used for pYB construction are described in Table S7. Large synthetic repeats of mouse major satellite, human gamma 8 satellites were generated by rolling-circle amplification (RCA) and recombinational cloning in yeast (Table S5, S6 and S7) and cloned into pYB vector. pYB/PUC/cHS4 vector was constructed as follows. The chicken beta-globin locus LCR HS4 site insulator was cloned as two tandem copies of the 250 bp core into a FseI site of pYB/PUC vector between the alphoid hook and the start of the LCR region.

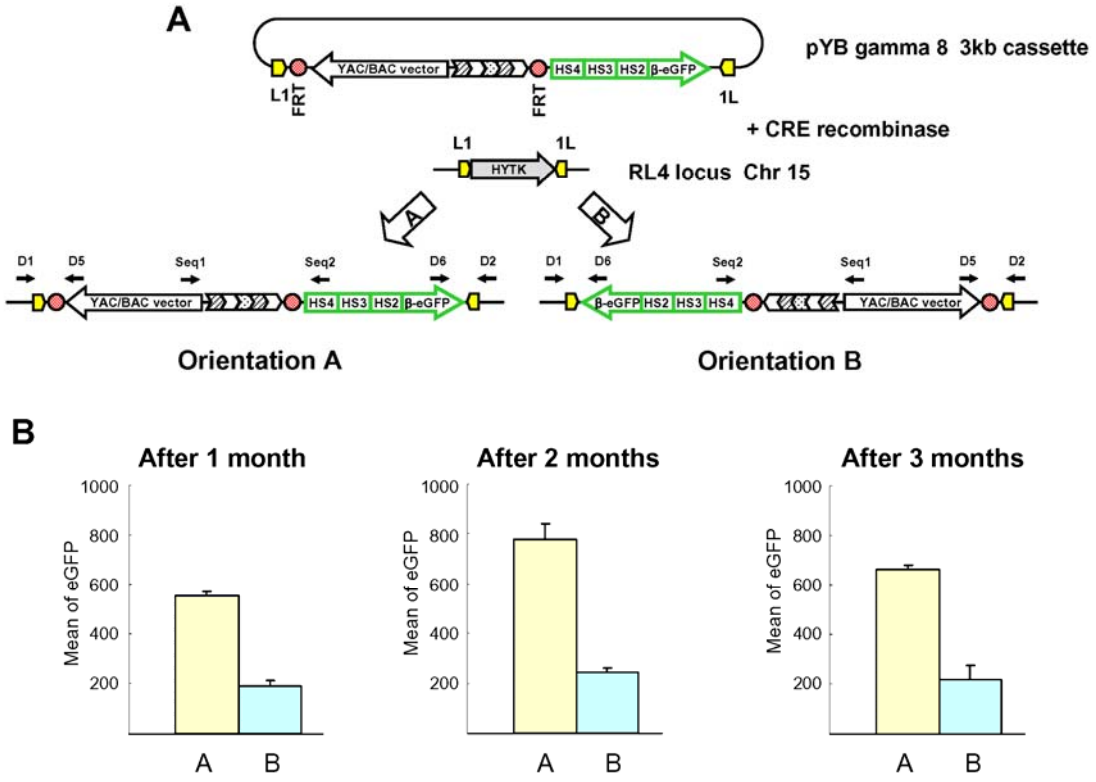


Figure S3. The eGFP Transgene Expression at the RL4 Locus

(A) The RMCE technique was used for the precise replacement of the L1-HYTK-1L cassette by 3kb gamma-satellite 8 pYB cassette at the RL4 locus (chromosome 15). After replacement, the vector produced two types of orientation, A and B.

(B) The level of eGFP in orientation A (in yellow) was higher than in orientation B (in blue) but the transgene cassette is stably expressed in both orientation during more than 3 months. A and B were distinguished by PCR (Table S9).

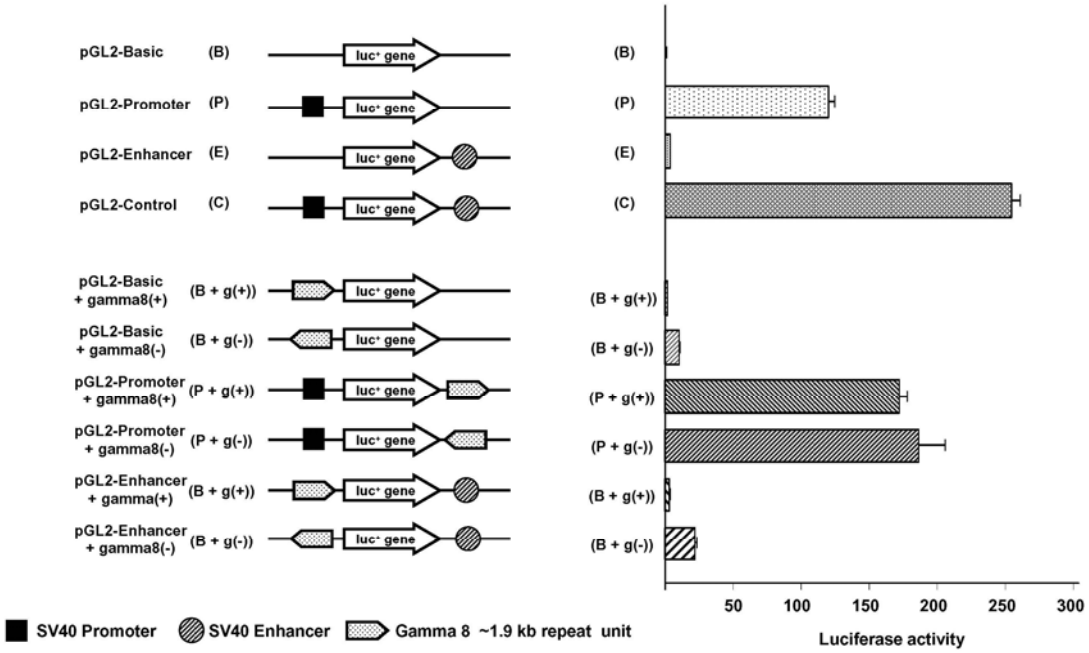


Figure S4. Analysis of gamma-satellite DNA

The ~1.9kb gamma-satellite 8 DNA fragment was linked to a firefly luciferase reporter gene in a pGL2-basic, promoter, or enhancer vector, and transfected into NIH3T3 cells with the Renilla luciferase gene in phRL-CMV as an internal standard. The schematic structures of the vectors are shown in left panel and the luciferase activity is shown in right panel. Values are means ± SD of three individual experiments in triplicates.

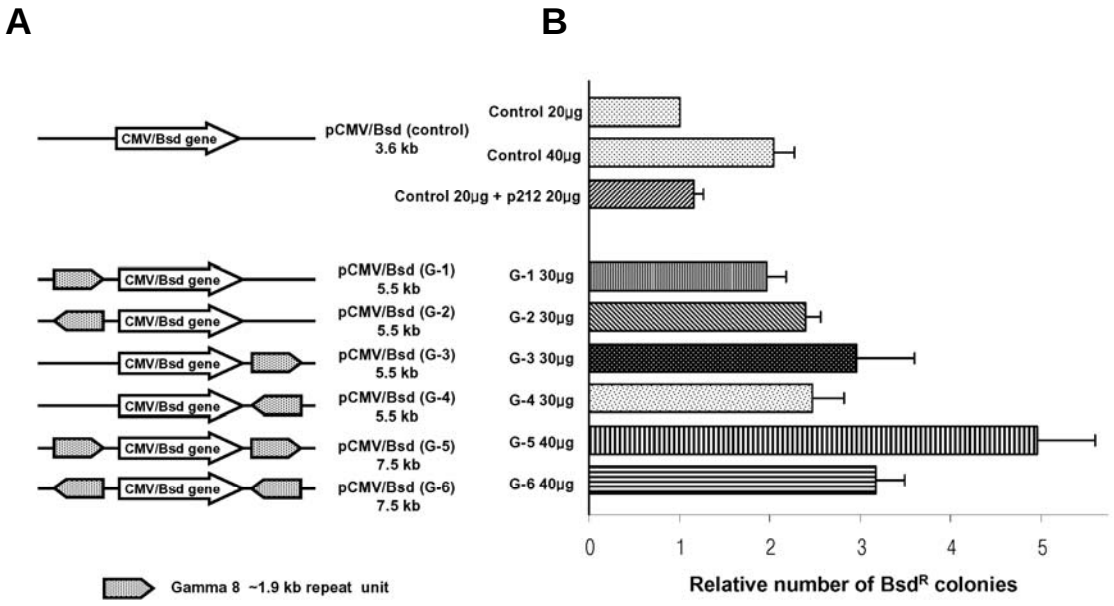


Figure S5. Gamma-Satellite DNA from Human Chromosome Increases Yield of Stable Transfectants
(A) Six vectors with human gamma-satellite DNA were constructed using the pCMV/Bsd vector (CMV promoter/Blastocydin^R gene) (Invitrogene). The ~1.9 kb gamma-satellite DNA fragment was cloned into either XhoI/BglII sites (upstream of CMV/Bsd gene) or into BamHI/XbaI sites (downstream of CMV/Bsd gene) producing G-1, G-2, G-3, and G-4 plasmids, correspondingly. Plasmids with the CMV/Bsd gene flanked by gamma-satellite fragments, G-5 and G-6, were also constructed. Before transfection, each vector was linearized by SspI digestion. The constructs were stably transfected into mouse MEL cells and Bsd-resistant colonies were counted.
(B) The relative number of Bsd-resistant colonies is shown. The number of colonies from the control construct pCMV/Bsd was arbitrarily set to 1.0. The construct containing two gamma-satellite DNA fragments flanking the Bsd gene exhibited a 3-5-fold increase in the number of Bsd-resistant colonies compared to the pCMV/Bsd vector alone. A different amount of the vector DNAs was used to equilibrate the vectors size.

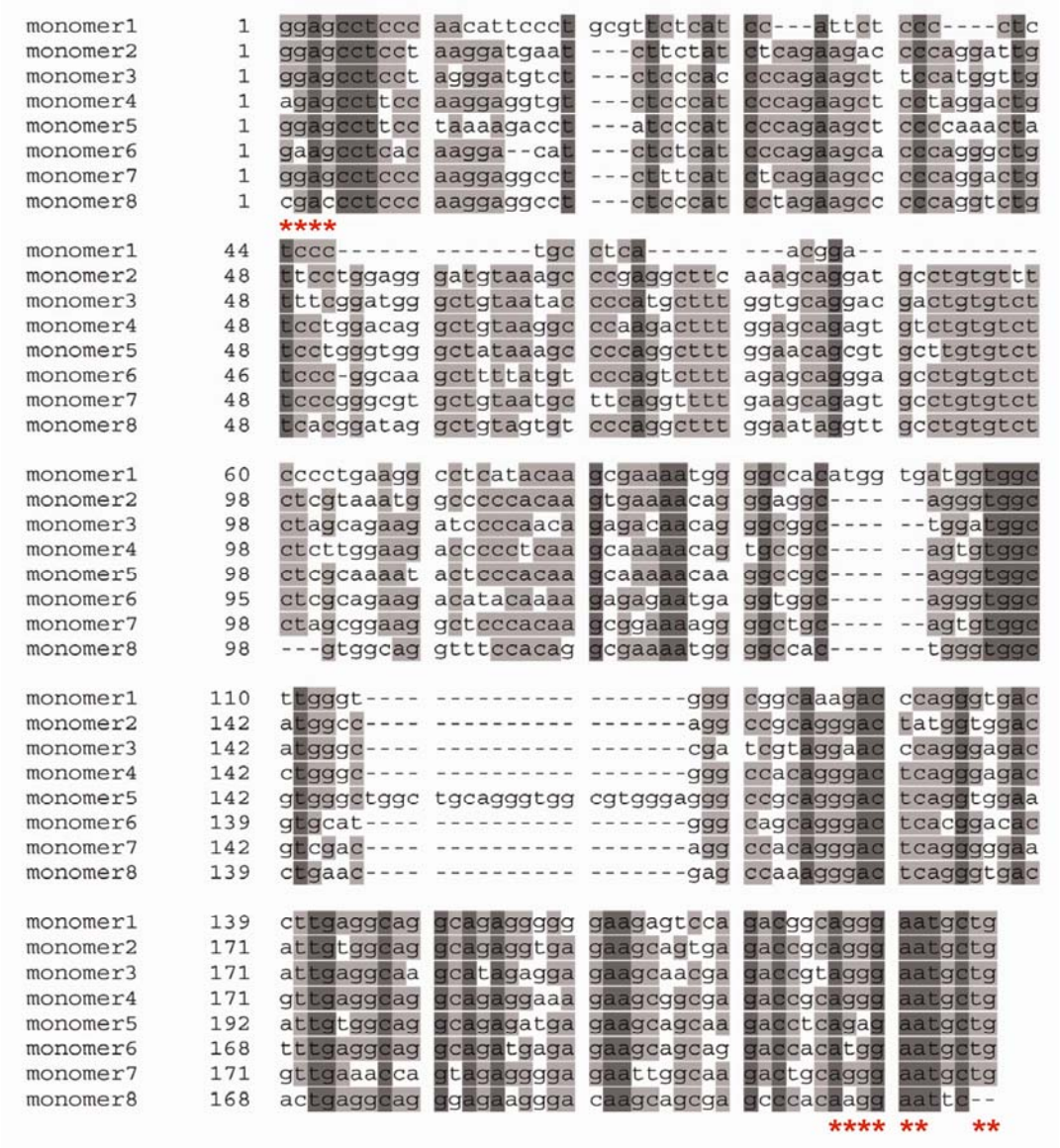


Figure S6. Alignment of eight monomers within a 1.9 kb gamma-satellite array exhibiting an anti-silencing activity.
Positions of 15 bp bipartite Ikaros binding motif in each monomer are shown in red asterisk mark.

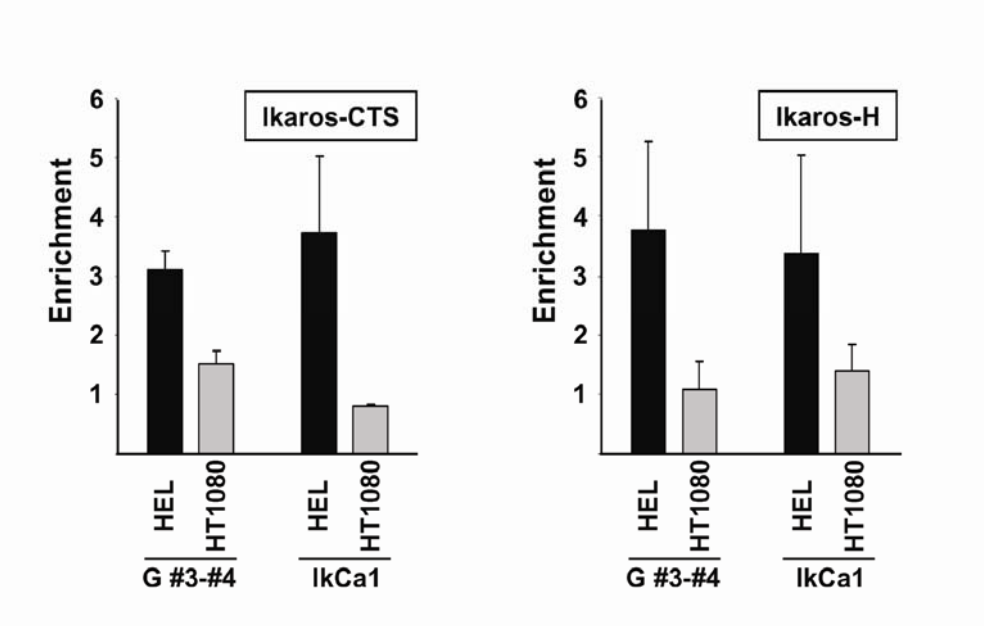
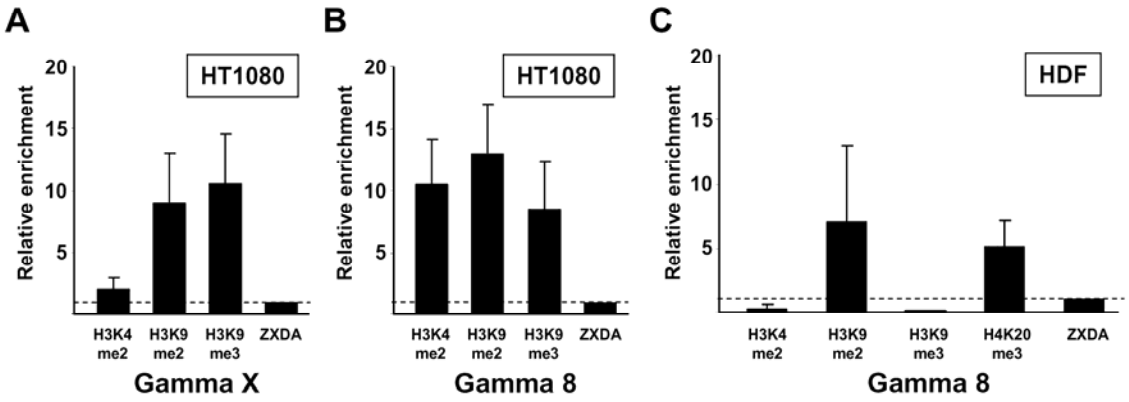


Figure S7. ChIP analysis of gamma-satellite DNA arrays on chromosome X in HEL cells.
Ikaros specific antibodies, IK-CTS and IK-H, were used for immunoprecipitation. HT1080 cells showed no enrichment with both antibodies and used as a negative control. Sequences of primers for gamma-satellite DNA and a control locus, IkCa1, are presented in Table S10.



Figures S8. ChIP analysis of gamma-satellite DNA arrays on chromosome X and 8 in human dermal fibroblats and fibrosarcoma HT1080 cells.
All data normalized to ZXDA, the most proximal human gene on Xp located in euchromatin just outside of the centromere region. Sequences of primers for gamma-satellite X and 8 DNAs are presented in Table S10.

A

Core sequence : 5′ CA/TGGGTGGCNTGGNC 3′

B

WT-8	:	CGAAAATGGGGCCACTGGGTGGCCTGAACGAGCCAAAGGGACTCAGGGTGACACT
Mu1-8	:	CGAAAATGGGGCCACT**AAA**TGG**T**CTGAACGAGCCAAAGGGACTCAGGGTGACACT
Mu2-8	:	CGAAAATGGGGCCACTG**A**GTGG**TT**TGAA**T**GAGCCAAAGGGACTCAGGGTGACACT
Mu3-8	:	CGAAAATGGGG**T**CA**T**T**A**GGTGG**T**CTGAA**T**GAGCCAAAGGGACTCAGGGTGACACT

Figure S9. Gamma-Satellite 8 Monomers Used for EMSA Analysis
(A) Alignment of eight monomers forming a repetitive structure in amplified gamma-satellite DNA (only a part of 220 bp monomer sequences is shown). A full-size sequence of these monomers is available from GenBank, Acc. number X68546. Contact "G" (or "C" for anti-sense strand) nucleotides identified by methylation interference in the monomer 8th are shown. Underlined-CpG dinucleotides.
(B) Mutations of contact guanine residues in the monomer 8th resulting in the lack of in vitro CTCF binding. Positions of mutated nucleotides in three sequence variants Mu1, Mu2 and Mu3 are marked in bold. Only a 55 bp core sequence of 100 bp fragments used for EMSA is shown.

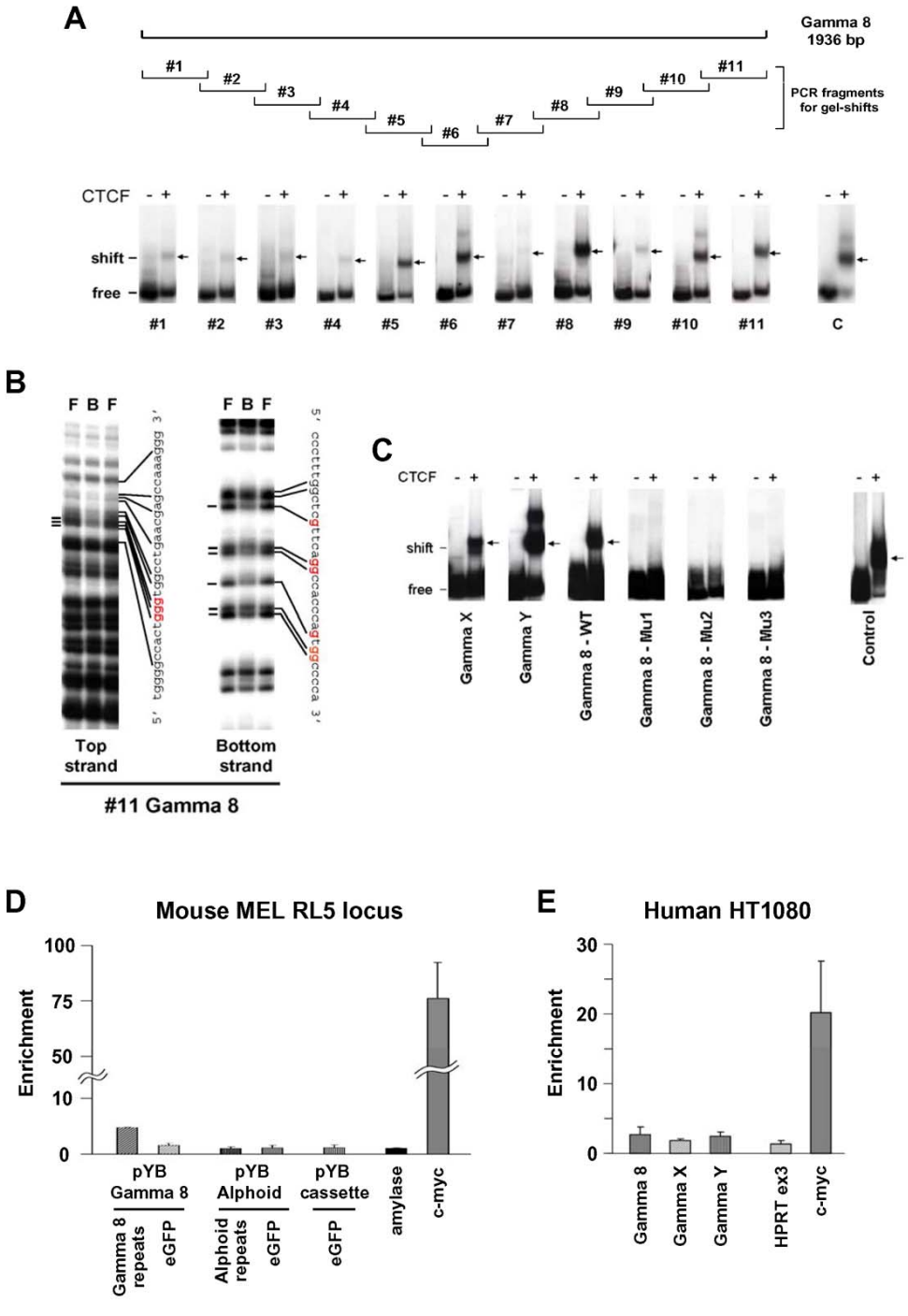


Figure S10. In vitro and in vivo interaction of CTCF with human gamma-satellite DNA
(A) A schematic representation of eleven overlapping fragments of 8-mer of the human gamma-satellite 8 DNA used in EMSA. EMSA was carried out with either control lysate (-) or lysate containing the in vitro translated 11 ZF DNA binding domain of CTCF protein (+). C; Positive control (c-myc promoter) of the EMSA reaction.

(B) Methylation interference of the CTCF binding to the human gamma-satellite 8 DNA. Top and bottom strands of the gamma-satellite DNA fragment #11 are shown. The corresponding DNA sequences are shown on the right. Seven methylated contact guanine residues are marked in red. Lane F, modified guanines of the free DNA probe. Lane B, methylated guanines of DNA molecules bound by CTCF.
(C) In vitro interaction of CTCF with mutated forms of gamma-satellite 8 DNA, Mu1, Mu2 and Mu3, with substitution of contact Gs to As and with gamma-satellite DNA repeats from chromosomes X and Y. Control, a fragment of the c-myc promoter.
(D) ChIP analysis of pYB gamma-satellite, pYB alpha-satellite DNA and pYB cassette were targeted into the RL5 locus in mouse MEL cells. ChIP analysis was performed three times; averages are shown with standard deviation bars.
(E) ChIP analysis of gamma-satellite DNA arrays in human HT1080 cells.

Protocols for Electrophoretic mobility shift (EMSA) and DMS methylation interference assays

The luciferase control as well as 11 ZF DNA binding domain of CTCF protein were synthesized from the Luciferase T7 control DNA and pET16b-11ZF construct, respectively (Filippova et al, 1996; Awad et al, 1999), with the TnT reticulocyte lysate coupled in vitro transcription-translation system (Promega, Madison, WI, USA). Overlapping ~250 bp fragments covering a 1.9 kb gamma-satellite 8 DNA unit or a 1.8 kb alphoid DNA unit from chromosome 21 were 32P-labeled, gel purified, and used as DNA probes for gel mobility shift assays with equal amounts of in vitro translated luciferase and CTCF proteins as described (Filippova et al, 1996; Awad et al, 1999). Gamma-satellite monomers from chromosome X and Y were PCR amplified from genomic DNA, cloned into a TA vector and sequenced before their analysis by EMSA. Corresponding primer sequences are presented in Table S9. Binding reactions were carried out in buffer containing standard PBS with 5 mM MgCl2, 0.1 mM ZnSO4, 1 mM DTT, 0.1% NP40, and 10% glycerol in the presence of poly-(deoxyinosinic-deoxy-CMP) and salmon sperm DNA. Reaction mixtures of 20 µl final volume were incubated for 30 minutes at room temperature and then analyzed on 5% nondenaturing PAGE run in 0.5x TBE buffer. For electrophoretic mobility gel-shift assay (EMSA) with in vitro methylated DNA probes, treatment with the SssI-methylase was done as previously described for CTCF-binding fragments DMD4 and DMD7 of the H19 ICR (Kanduri et al, 2000). The extent of methylation was verified by digestion overnight with Sau96I restriction endonuclease. Methylation interference analysis was carried out as described previously (Filippova et al, 1996).

References:

Awad TA, Bigler J, Ulmer JE, et al (1999) Negative transcriptional regulation mediated by thyroid hormone response element 144 requires binding of the multivalent factor CTCF to a novel target DNA sequence. J Biol Chem 274: 27092–27098

Filippova GN, Fagerlie S, Klenova EM, Myers C, Dehner Y, Goodwin G, Neiman PE, Collins SJ, Lobanenkov VV (1996) An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. Mol Cell Biol 16: 2802-2813

Kanduri C, Pant V, Loukinov D, et al (2000) Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive. Curr Biol 10: 853–856

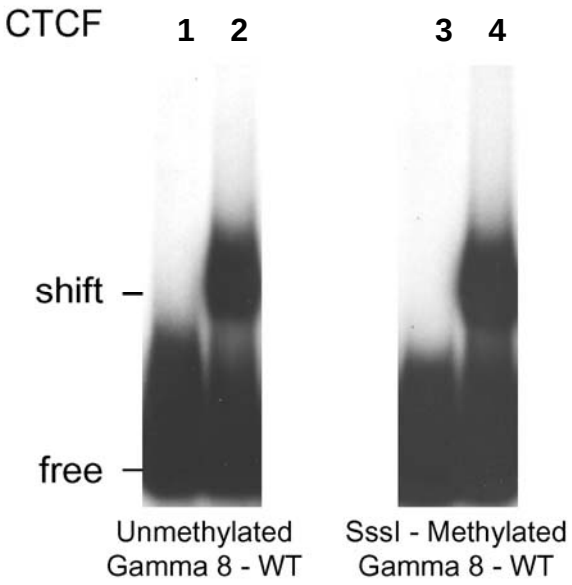


Figure S11. Binding of CTCF to Gamma Satellite is Independent of CpG Methylation
Control unmethylated (lanes 1, 2) or SssI-methylated (lanes 3, 4) fragments were analyzed by gel-shift assay (EMSA). (-) free probe; (+) CTCF-bound probe. There are two CpG dinucleotides within a 55 bp CTCF core sequence (monomer 8th), one of them corresponds to the contact nucleotides identified by methylation interference (Figure S9A).

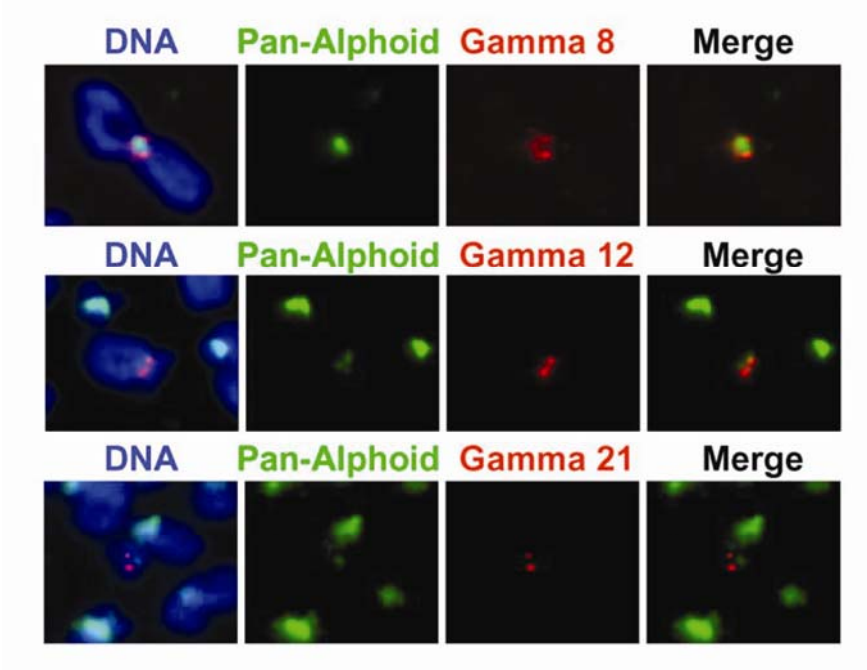


Figure S12. Detection of gamma-satellite repeats localization by FISH analysis.
Green signal is a pan-alphoid probe used to detect alpha-satellite repeats at the centromere. Red signal is gamma-satellite probes specific for chromosomes 8, 12 or 21, respectively. Blue is DAPI staining of DNA. Sequences of primers used for preparation of chromosome-specific probes are presented in Table S10.

A

```
GSATII       1  CATGCGTGGG GCCCA--GGG GACCCTGGGC ATCCCTGGTT CATGCCCACG GAGTGCCTCG
 GSATX       1  CCTGCGCCGG ACCCGGGGGG G-TCGT-GGA GTCCCTGGCT TTCACCCAGG GTGCGTGTCT
  GSAT       1  CCAGGGC-TG TCCCGGGCGG G-CTGT-AAA GCCCCAGGCT TTGGAGCAGG GTGCCTGTGT

GSATII      59  CGGGCCCAC- GGGGGCCAGC CC-------- --AAGGCGGC AGGAAGGC-- ----------
 GSATX      59  CTCTCCCAC- TGGGGGCA-C CC-------- -CAAAGCGGC AAGAAGTC-C CCCAGGGGAC
  GSAT      58  GTCTCTCGCG GAAGGCCCCC ACAAGCGAAA ACGGGGCCGC AGGGTGGCGT GTGCGGGCC-
                                                  C  GGG   CN    C
GSATII      96  ---------- ---------- TTGAAAGGGG AGGTCGAGGC ACCTGTGTTG TGGAAGGAAA
 GSATX     107  ACAGGGACAG GACGCCAGGC TTTCAGGGGG ACGTTGAGGC AGG-----CC GGGGAA---A
  GSAT     117  C-GCA----- -----GAAAC TCAG--GGGG ACATTGAGGC AGG------- ---CAG-AGG

GSATII     136  AAACACAAAC GGCGAG-GCA GA-------- -----GGTCC CCCCCACGG GCGAAAGTG-
 GSATX     159  -AAA---AGC GGCGAG-GCC AAAGAGG-AG GCTGGGGTCC TCCC--CCGG AGGTCAGTGC
  GSAT     153  GGGGAGAAGC GGCGAGACCT CAGGG--AAT GCTGGGAGCC TCCC--AAGG AGGCCTCTCC

GSATII     181  G-CCTCCCCA CCGCG
 GSATX     211  GCCTTCCCGG CAGCC
  GSAT     209  CCCATCCCAG AAGCC
```

B

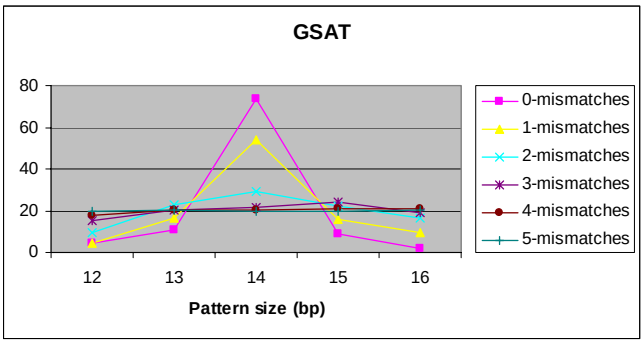
	GSATII	GSATX	GSAT
GSATII		68.79	53.8
GSATX	50.85		63.45
GSAT	39.32	52.74	

C

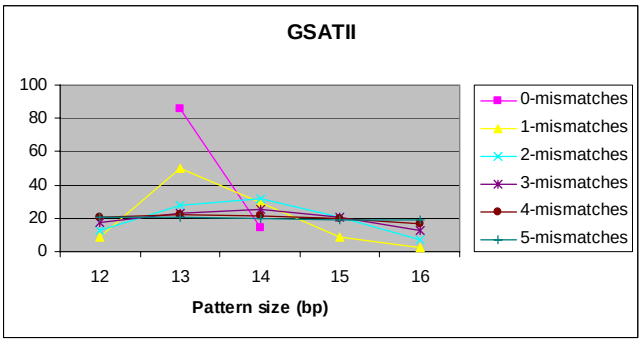
	GSATII	GSATX	GSAT
GSATII		26.07	26.92
GSATX			16.88
GSAT			

Figure S13. Alignment of GSAT, GSATX, GSATII Consensus Sequences
(A) The consensus sequences were reconstructed based on comparison of gamma-satellite monomers (Figure 8). The CTCF binding core identified by methylation interference is shown in red (Figure S10B). The region is relatively preserved between three gamma-satellite subfamilies.
(B) Mean identity between gamma-satellite families. In red are identities without insertions and deletions (indels), in blue are identities including indels.
(C) Proportions of insertions and deletions (indels) in pairwise aligments (%).

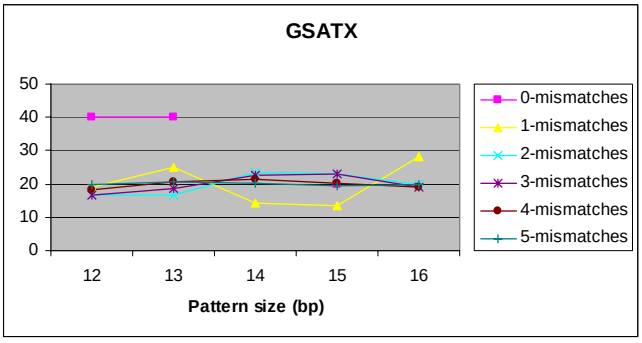
A



B



C

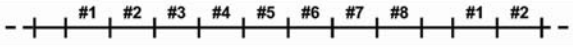
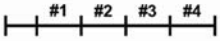
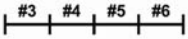
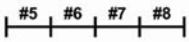
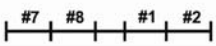


D

Satellite	Mismatches	Sequence number	Total length (bp)	Hits	Hit length (bp)	Hits/1kb
GSAT	0	67	239407	46	641	0.192
GSAT	2	67	239407	7218	101996	30.149
GSATII	0	164	163493	7	92	0.043
GSATII	2	164	163493	3037	41951	18.576
GSATX	0	52	103030	5	65	0.049
GSATX	2	52	103030	2005	28310	19.460

Figure S14. Analysis of Ikaros binding sites in the human gamma-satellite DNA.
(A, B, C) Frequency of two-core motifs with a high affinity to the largest Ikaros isoform, IK-H, in GSAT, GSATII and GSATX gamma-satellite repeats. This shows that motifs 14pb long (4 pb linked between monomers) are most frequent for GSAT. GSATII data suggest 3bp linker. No clear preference for GSATX., (D) Density of Ikaros two-core recognition sites with 0 and 2 mismatches from the consensus in three subfamily of the human gamma-satellite DNA repeats. Note that all motifs, including overlapping hits, are counted.

Table S1. Functional analysis of sub-fragments derived from the 8-mer gamma-satellite DNA.

Name	Repeat unit composition of the pYB Gamma vectors	Amplified repeats size	Relative eGFP intensity
pYB Gamma8 3 kb		3.1 kb	+++++
pYB G #1 - #4		4.2 kb	++++
pYB G #3 - #6		3.5 kb	+++++
pYB G #5 - #8		3.5 kb	++++
pYB G #7 - #2		4.2 kb	+++++

Each monomer of gamma 8 repeat unit size is around 220 bp. Between monomer #8 and monomer #1, there is partial 173 bp monomer.

Table S2. Primers Used for Amplification of Gamma and Alphoid-Satellite Fragments for Electrophoretic Mobility Shift Assay (EMSA)

Name	Forward primer	Reverse primer
Gamma 8 gel 1	5’ ctgggagtgacccaaagagg 3’	5’ gaatgggatgagaacgcaggg 3’
Gamma 8 gel 2	5’ ctatgagcttctgtgatggg 3’	5’ ggggtcttctgagatagaag 3’
Gamma 8 gel 3	5’ gggaagagtccagacggcag 3’	5’ tcccagcattccctgcggtc 3’
Gamma 8 gel 4	5’ atggccaggccgcagggac 3’	5’ cttctcctctatgcttgcc 3’
Gamma 8 gel 5	5’ ggctggatggcatgggccg 3’	5’ gcctcaacgtctccctgag 3’
Gamma 8 gel 6	5’ caaaaacagtgccgcagt 3’	5’ ccagcccacgccaccctgcgg 3’
Gamma 8 gel 7	5’ ccccaggctttggaacagcg 3’	5’ acagccctgggtgcttctggg 3’
Gamma 8 gel 8	5’ ggcaggcagagatgagaag 3’	5’ ctcccagcattccatgtgg 3’
Gamma 8 gel 9	5’ gggcagcagggactcacgg 3’	5’ ccccttttccgcttgtggg 3’
Gamma 8 gel 10	5’ gtgctgtaatgcttcaggttttg 3’	5’ caacctattccaaagcctggg 3’
Gamma 8 gel 11	5’ atgctgcgaccctcccaagg 3’	5’ gaattccttgtgggctcgc 3’
Alphoid gel 1	5’ aattcaaataaaaggtagac 3’	5’ aaaggttccactctgttagc 3’
Alphoid gel 2	5’ aaaagtaaatatcttccata 3’	5’ tctgttagttgaggacacac 3’
Alphoid gel 3	5’ tatcgttggaaaagggaata 3’	5’ gaatgcagatatcaccaagt 3’
Alphoid gel 4	5’ aaacgggaatatcatcatct 3’	5’ cttttagttgagtacacaca 3’
Alphoid gel 5	5’ tgcctttgttgaaaaggaaa 3’	5’ ttgaatggaaatatccgaaa 3’
Alphoid gel 6	5’ gattgctttgaggatttcgt 3’	5’ tatcaccaacaagtttctga 3’
Alphoid gel 7	5’ cgcctacggtgaaaaaggaa 3’	5’ tgaatgcagtcatcagaaag 3’
Alphoid gel 8	5’ ggatagcttggaggatttcg 3’	5’ tcacaaacttgtttctcaga 3’
Alphoid gel 9	5’ ggacatttggagcgctttga 3’	5’ tctgagagggcttctgtcta 3’
Gamma X gel 1	5’ ctgtggggacagacacacac 3’	5’ tttcaggggtacgttgaagc 3’
Gamma Y gel 1	5’ acctgacgtgctgtctcctt 3’	5’ gaggcctacttgcgactttg 3’
Major gel 1	5’ tatggcgaggaaaactg 3’	5’ tttcacgtcctaaagtgtg 3’
Major gel 2	5’ cagtggacatttctaaattt 3’	5’ ggaatatggtgagaaaactg 3’

Table S3. Location of gamma-satellite DNA repeats in the sequenced part of the human genome

Chromosome	Number of repeats	Size of the genomic region harboring gamma-satellite monomers (in bp)
1	6	6241
2	2	3708
3	1	166
4	8	9356
5	2	181
6	1	96
8	68	269539
9	8	21040
10	1	148
12	151	121456
13	1	1935
14	3	5495
15	3	642
17	1	155
18	1	461
21	1	2955
22	3	2779
X	16	39099
Y	6	20478

Table S4A. Location of gamma-satellite DNA repeats in the sequenced part of the chimpanzee genome

Chromosome	Number of repeats	Length in bp
1	5	1659
2a	1	178
2b	3	3397
3	1	166
4	22	37280
5	2	190
6	1	96
8	76	140400
9	1	1575
10	1	148
12	141	88110
13	2	927
14	1	2571
15	3	636
17	1	155
18	1	1763
21	1	2948
22	1	1467
X	15	16076
Y	14	20978

Table S4B. Gamma-satellite DNA repeats in the sequenced part of the chimpanzee genome by families

Chromosome	Family	Number of repeats	Length in bp
1	GSAT	1	92
1	GSATII	2	1371
1	GSATX	2	196
2a	GSAT	1	178
2b	GSATII	3	3397
3	GSATX	1	166
4	GSAT	2	302
4	GSATX	20	36978
5	GSATX	2	190
6	GSATX	1	96
8	GSAT	60	117379
8	GSATX	16	23021
9	GSATII	1	1575
10	GSAT	1	148
12	GSAT	1	135
12	GSATII	120	75553
12	GSATX	20	12422
13	GSAT	1	176
13	GSATX	1	751
14	GSATII	1	2571
15	GSAT	2	513
15	GSATX	1	123
17	GSAT	1	155
18	GSATX	1	1763
21	GSATX	1	2948
22	GSATX	1	1467
X	GSATX	15	16076
Y	GSAT	1	118
Y	GSATX	13	20860

Table S4C. Location of gamma-satellite DNA repeats in the sequenced part of the rhesus macaque genome

Chromosome	Number of repeats	Length in bp
2	1	170
3	15	27512
4	1	97
5	27	53361
6	1	81
7	4	1262
8	14	51275
9	7	2598
12	2	226
13	2	290
16	1	164
17	1	151
18	1	911
19	1	2376

Table S4D. Gamma-satellite DNA repeats in the sequenced part of the rhesus macaque genome by families

Chromosome	Family	Number of repeats	Length in bp
2	GSATX	1	170
3	GSATII	15	27512
4	GSATX	1	97
5	GSAT	3	448
5	GSATII	1	60
5	GSATX	23	52853
6	GSATX	1	81
7	GSAT	2	560
7	GSATX	2	702
8	GSAT	12	48854
8	GSATII	2	2421
9	GSAT	2	331
9	GSATII	5	2267
12	GSATII	2	226
13	GSAT	1	171
13	GSATII	1	119
16	GSAT	1	164
17	GSAT	1	151
18	GSATX	1	911
19	GSATX	1	2376

Table S5. Primers Used for PCR Amplification of Repeats

Name	Primer sequence	Size of PCR product	Number of repeat units
Mouse major F Mouse major R	5’ acgtgaattctggcgaggaaaactgaaaaaggtg 3’ 5’ gccagaattcacgtcctaaagtgtgtatttctca 3’	704 bp*	3
Gamma8 repeat F Gamma8 repeat R	5’ cgatgaaggcctctccgatcct 3’ 5’ gaaagtcctgggggcttctgga 3’	1,962 bp*	8
HS4 1copy F HS4 1copy R	5’ gatcactagtgagctcacggggacagcc 3’ 5’ gatctctagactctctttcagcctaaagct 3’	278bp	1
HS4 2copy F HS4 2copy R	5’ gatcggccggccagtgtgctggaattcgccct 3’ 5’ gatcggccggcctgtgatggatatctgcagaat 3’	572 bp	2

* All repeats were PCR amplified from genomic DNA and sequenced. Size of fragments cloned in TA vector is shown.

Table S6. Thio-Phosphate Linked Primers Used for RCA Amplification of Repeats

Name	Primer sequence
Mouse major	MRCA F1 : 5’ acttgacGA 3’ MRCA F2 : 5’ tgcacactGA 3’ MRCA R2 : 5’ cgccatatTC 3’
Human Gamma 8	GRCA F1 : 5’ aattctgGG 3’ GRCA R1 : 5’ ttaagacCC 3’ GRCA R2 : 5’ cctccacAG 3’

Each primer was linked by Thio-phosphate through the last two oligomers.

Table S7. Primers Used for pYB Vector Construction

Name	Forward primer	Reverse primer
Cla Rsr alphoid	5’ gttacctatcgatatcggaccgtctagacagaa gcattctcagaaactt 3’	5’ tgtcggatcgattacggaccgatgtgaagatatt cccgtttccaac 3’
Mlu Rsr	5’ atgactacgcgtaaacactctttttgtagaatctg caag 3’	5’ catggtaacgcgtctgctctatcaaaaggaaggtt caact 3’
Frt alphoid	5’ cctatactttctagagaataggaacttctggccg gccccggaccg 3’	5’ attctctagaaagtataggaacttcgacgtcagc ggccgcacggaccgatgtgaagatattcccgtttccaac 3’
Cla Frt	5’gttactatcgatagaagttcctatactttctagaga ataggaacttcg 3’	5’tgtcggatcgatagctagcaaccgcggtgaagttcctattc tctagaaagtataggaacttcg 3’
YAC cassette	5’ cgcagcggccgcatctgtgcggtatttcacacc gc 3’	5’ atgcgcggccgccgaaaagtgccacctgggtcc 3’
BAC cassette	5’ tatgtcgacatcggatgcagcccggttaa 3’	5’ ttgtggtttgtccaaactcatcaatg 3’
Kan^R gene	5’ gatcgtcgactgaaagccacgttgtgtctc 3’	5’ gatcgggccctcccgtcaagtcagcgtaat 3’
PUC linker	5’ gatcacgcgtactgatgcatgatccgggtt 3’	5’ gatcacgcgtactgatgcatgatccgggtt 3’

Table S8. Targeting Hook Sequences

Name of hook	Hook sequence	Product size
Mouse major 5’	5’gatccggaccgatggcgaggaaaactgaaaaaggtggaaaatttagaaatgtcca ctgtaggacgtggaatatggcaagaaaactgaaaatcatggaaaatgagaaacatcc acttgacgaacgcgtgatc 3’	131 bp
Mouse major 3’	5’gatcacgcgttgaaaaatgacgaaatcactaaaaacgtgaaaaatgagaaatgca cactgaaggacctggaatatggcgagaaaactgaaaatcacggaaaatgagaaatac acactttaggacgtgcggaccggatc 3’	138 bp
Gamma8 repeats 5’	5’gatccggaccgactatggtggacattgtggtcaggcagaggtgagaagacagtga gaccgcagggaatgctgggagcctcctagggatgtctctcccaccccagaagcttac catngttgtttcggatgggctgtaataccccatgctttggtacgcgtgatc 3’	163 bp
Gamma8 repeats 3’	5’gatcacgcgtgtagagggaagaattggcaagactgcagggtaatgctgcgaccct cccaaggagagcctctcccatcctagaagccccccaggtctgtcacggataggctgt agtgtcggaccggatc 3’	128 bp
Human alpha satellite * 5’ and 3’	5’atgcatcgataagagtgtttcaaaactgctctatcaaaaggaatgttcaACGCG Tgagttgaatgcaaacttcacaaagaagtttctgagaatgctcgaggcatgcat 3’	108 bp

Restriction sites were introduced into 5’ ends of the primers to simplify cloning into pYB vector.
* Alpha satellite sequence contains a native *Mlu*I site (marked in capital). Digestion with *Mlu*I enzyme produces two hooks of ~40 bp each.

Table S9. Primers Used for Cloning of the RL5 Locus and Determination of Orientation of the Insert

Name	Primer sequence	Name	Primer sequence
DW-ACP 1*	5’ ACP-aggtc 3’	RL5 Cen 1	5’ gtgaagaccaggcatggaggct 3’
DW-ACP 2*	5’ ACP-tggtc 3’	RL5 Cen 2	5’ aagctctccccactggtttgctc 3’
DW-ACP 3*	5’ ACP-gggtc 3’	RL5 Cen 3	5’ catgagcctgtggggagatgtcc 3’
DW-ACP 4*	5’ ACP-cggtc 3’	RL5 Cen 4	5’ gtctcactctgtagtgcagaaca 3’
DW-ACPN*	5’ ACPN-ggtc 3’	RL5 Cen 5	5’ catggtcatactttccctagct 3’
Uni-primer*	5’ tcacagaagtatgccaagcga 3’	RL5 Cen 6	5’ cacagcagtaaaaccctaacta 3’
5’HyTk TSP 1	5’ gggtaccgagctcgaattcact 3’	seq 1	5’ aacggagtaacctcggtgtg 3’
5’HyTk TSP 2	5’ gccgtcgttttacaacgtcgtgac 3’	seq 2	5’ agctgctgagtgggagagag 3’
5’HyTk TSP 3	5’ gggaaaaccctggcgttacccaact 3’	seq 3	5’ gctgtacaagtaaagcggcc 3’
3’HyTk TSP 1	5’ gttatccgctcacaattccaca 3’	seq 4	5’ caagacgtttcccgttgaat 3’
3’HyTk TSP 2	5’ ccacacaacatacgagccggaagc 3’	D1	5’ aacgccagggttttcccagtcacg 3’
3’HyTk TSP 3	5’ cctggggtgcctaatgagtgagc 3’	D2	5’ gggcagtgagcgcaacgcaatta 3’
3’HyTk TSP 4	5’ cgtttttccataggctccgcc 3’	D3	5’ ggcggtaatgttggacatgagcgaat 3’
3’HyTk TSP 5	5’ cggtaagacacgacttatcgcca 3’	D4	5’ ctgaagcttcccgggggtaccgaat 3’
3’HyTk TSP 6	5’ cacagcagtaaaaccctaacta 3’	D5	5’ cggccgctgacgtcgaagttcctat 3’
RL5 Tel 1	5’ tggaggcccctctccactca 3’	D6	5’ tcactctcggcatggacgagctgta 3’

* These primers were provided by the manufacturer (Seegene).

Table S10. Primers Used for Chromatin Immunoprecipitation (ChIP) Assay and FISH probe

Name	Forward primer	Reverse primer
C mBG 4	5’ atggcctgaatcacttggac 3’	5’ ttctcaggatccacatgcag 3’
C mAmylase 2	5’ ttctgctgctttccctcatt 3’	5’ cgaacaggtggacaatagca 3’
C RL5 Cen 1	5’ cagggagccaacagtctttc 3’	5’ ccacacaaggagtccaaggt 3’
C YAC cassette 2	5’ tcaccaatgcactcaacgat 3’	5’ cagtagcagaacaggccaca 3’
C BAC cassette 1	5’ ctggggaagcatggttctaa 3’	5’ caccagttgaagagcgttga 3’
C eGFP 6	5’ agaacggcatcaaggtgaac 3’	5’ tgctcaggtagtggttgtcg 3’
C RL5 Tel 1	5’ ttatggcatggcgatttgta 3’	5’ tgaccttcccagtcttgctt 3’
Gamma 8	5’ gtgctgtaatgcttcaggttttg 3’	5’ caacctattccaaagcctggg 3’
Gamma X	5’ ctgtggggacagacacacac 3’	5’ tttcaggggtacgttgaagc 3’
Gamma Y	5’ acctgacgtgctgtctcctt 3’	5’ gaggcctacttgcgactttg 3’
Mouse c-myc	5’ aagtaagtgtgccctctactgg 3’	5’ aaggaagcatcttcccagaa 3’
Human c-myc	5’ agaataacaaggaggtggctggaaacttg 3’	5’ ttgcaaattactcctgcctccaggcctt 3’
Gamma FISH 8	5’ gaattctgggagtgacccaa 3’	5’ gaattccttgtgggctcgc 3’
Gamma FISH 12	5’ tcactccctgggcacgaacc 3’	5’ gcagaggtcacccccaacga 3’
Gamma FISH 21	5’ gcccacgtaattcaattcact 3’	5’ aaggagtgtgaccaaaactca 3’
Gamma X Fiber	5’ ttcaacgtacccctgaaagcctgg 3’	5’ ctattttgtcccaagcctgcc 3’
Ikaros RT-insert	5’ tgagcccatgcctgtccctgag 3’	5’ ggtcttctgccatctcgttgtggtta 3’
IKCa1 promoter	5’ tctacacgtattgggtttcg 3’	5’ gcacacaacacaacctacac 3’
Ikaros g#3+g#4	5’ atcgtaggaacccagggaga 3’	5’ ccaaagtcttgggccttaca 3’
Ikaros g#4+g#5	5’ ctcttggaagacccccctca 3’	5’ gatagtttggggagcttctgg 3’
Ikaros g#5+g#6	5’ gcaaaatactcccacaagcaa 3’	5’ ataaaagcttgccgggaca 3’