Hot spots of DNA double-strand breaks and genomic contacts of human rDNA units are involved in epigenetic regulation

Nickolai A. Tchurikov,[*] Daria M. Fedoseeva, Dmitri V. Sosin, Anastasia V. Snezhkina,

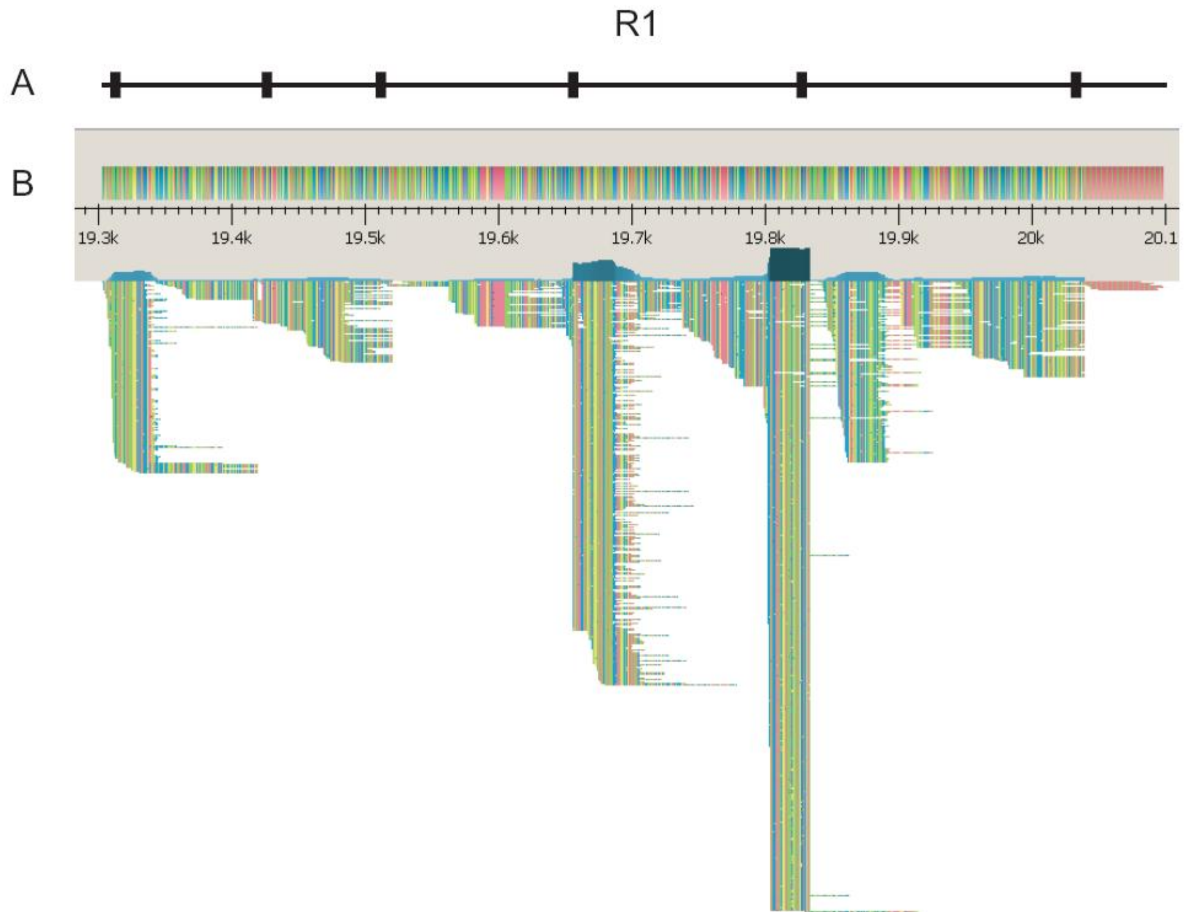
Nataliya V. Melnikova, Anna V. Kudryavtseva, Yuri V. Kravatsky, Olga V. Kretova

***To whom correspondence may be addressed:** E-mail: tchurikov@eimb.ru

Supplementary material:

Supplementary Figures S1-S18
Supplementary Tables S1-S4
Supplementary Methods
Supplementary References

Supplementary Figures



Supplementary Figure S1. Overview of reads mapped in R1

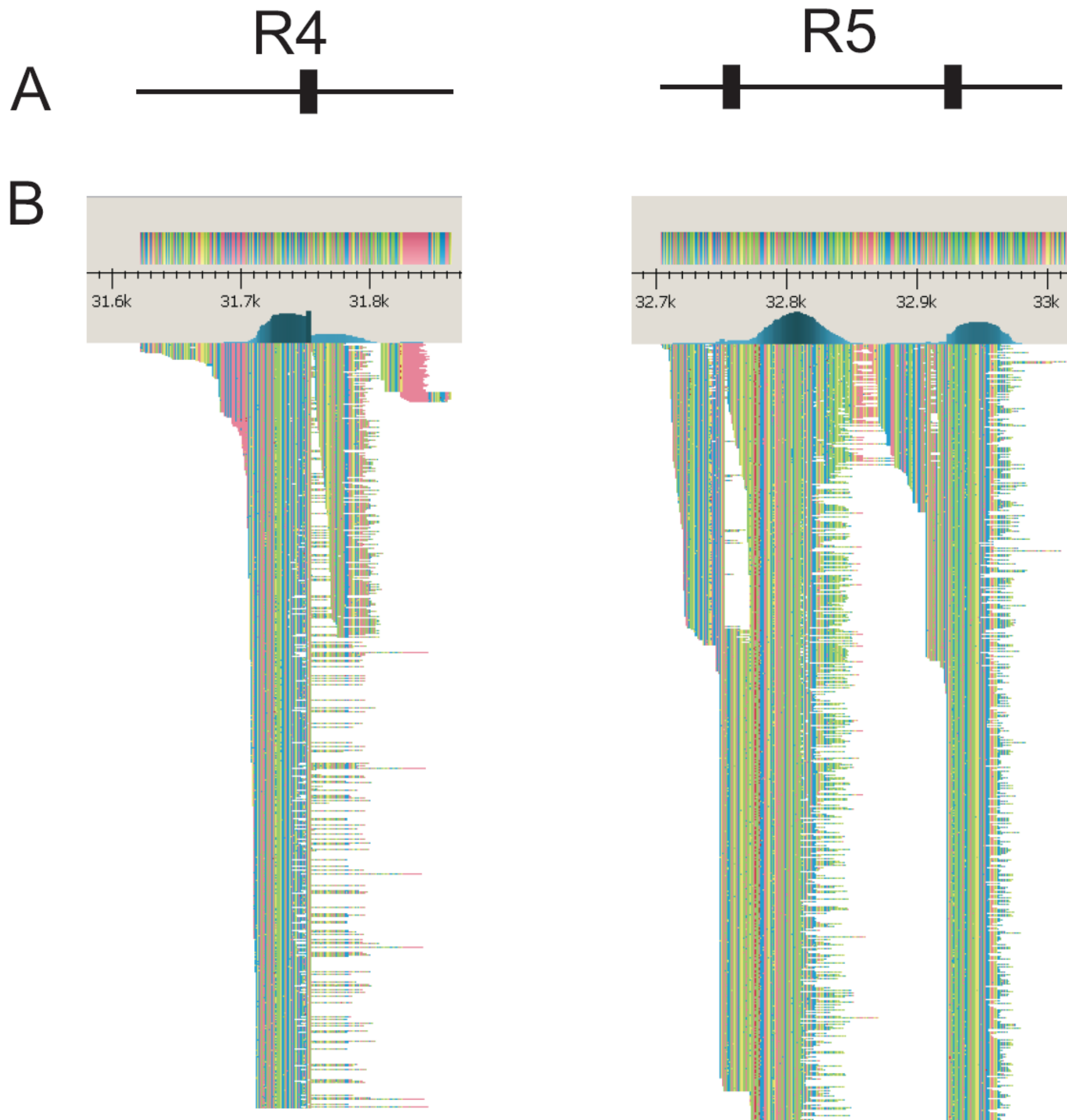
(A) Scheme showing the distribution of Sau3A sites (black bars), which together with sites at DSBs delimit the reads.

(B) The corresponding reads schematically are visualized by UGENE software (http://ugene.unipro.ru/). The numberings are according to the human ribosomal DNA complete repeating unit sequence (Accession number U13369). The reads in R1 are aligned in 474 rows.

Supplementary Figure S2. Overview of reads mapped in R2 and R3

(A) Scheme showing the distribution of Sau3A sites (black bars) or a mixture containing the diverged sites (blue bar), which together with sites at DSBs delimit the reads.

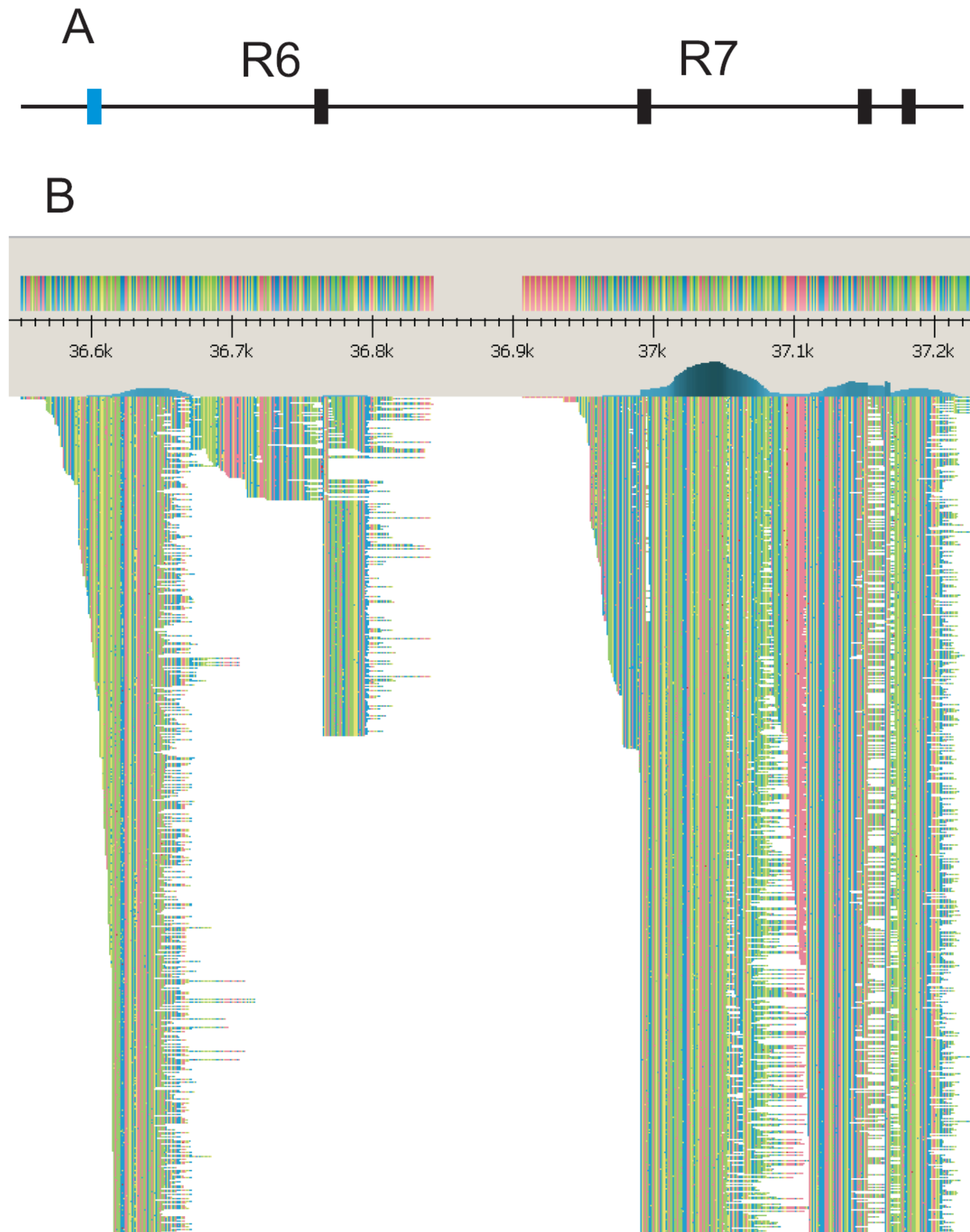
(B) The corresponding reads schematically are visualized by UGENE software (http://ugene.unipro.ru/). The numberings are according to the human ribosomal DNA complete repeating unit sequence (Accession number U13369). The reads in R2 are aligned in one site presented in 1198 rows, and the reads in R3 are presented in two adjacent sites possessing 325 and 334 rows.



Supplementary Figure S3. Overview of reads mapped in R4 and R5

(A) Scheme showing the distribution of Sau3A sites (black bars), which together with sites at DSBs delimit the reads. Blue bar shows a mixture of normal and diverged Sau3A sequences.

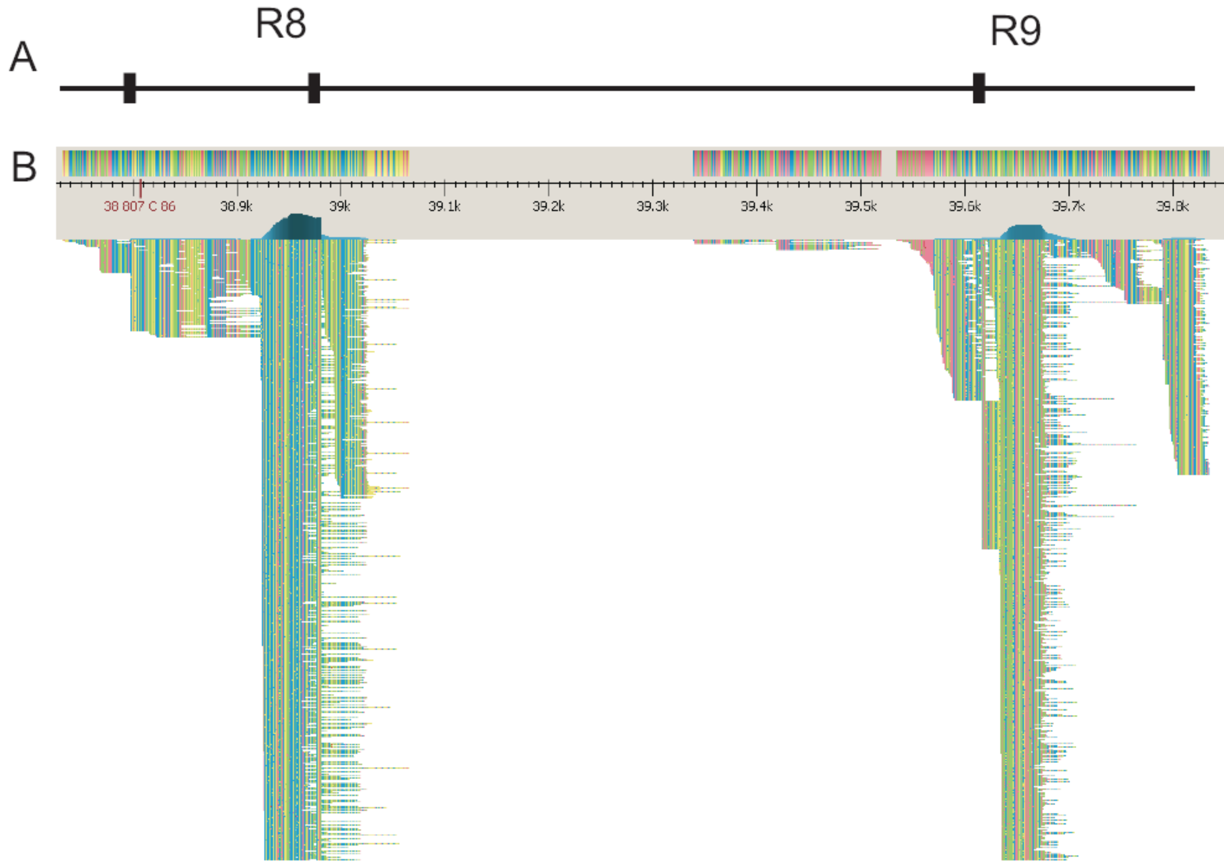
(B) The corresponding reads schematically are visualized by UGENE software (http://ugene.unipro.ru/). The numberings are according to the human ribosomal DNA complete repeating unit sequence (Accession number U13369). The reads in R4 are aligned in one site presented in 2743 rows, and the reads in R5 are presented in two adjacent sites possessing 3962 and 2722 rows. Only the top 594 reads are shown on the scheme.



Supplementary Figure S4. Overview of reads mapped in R6 and R7

(A) Scheme showing the distribution of Sau3A sites (black bars) or a mixture containing the diverged sites (blue bar), which together with sites at DSBs delimit the reads.

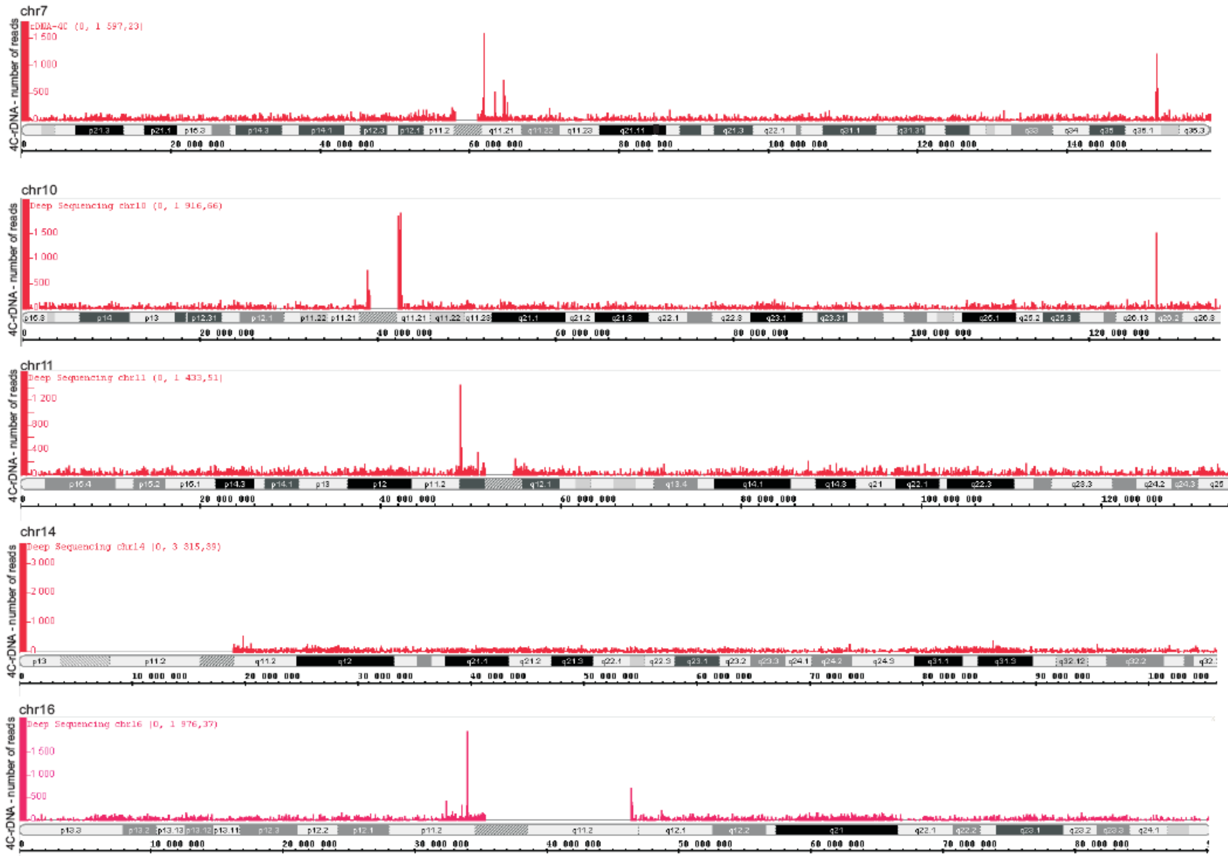
(B) The corresponding reads schematically are visualized by UGENE software (http://ugene.unipro.ru/). The numberings are according to the human ribosomal DNA complete repeating unit sequence (Accession number U13369). The reads in R6 are aligned in two sites presented in 1815 and 242rows, and the reads in R7 also are presented in two adjacent sites possessing 7065 and 3010 rows. Only the top 594 reads are shown on the scheme.



Supplementary Figure S5. Overview of reads mapped in R8 and R9

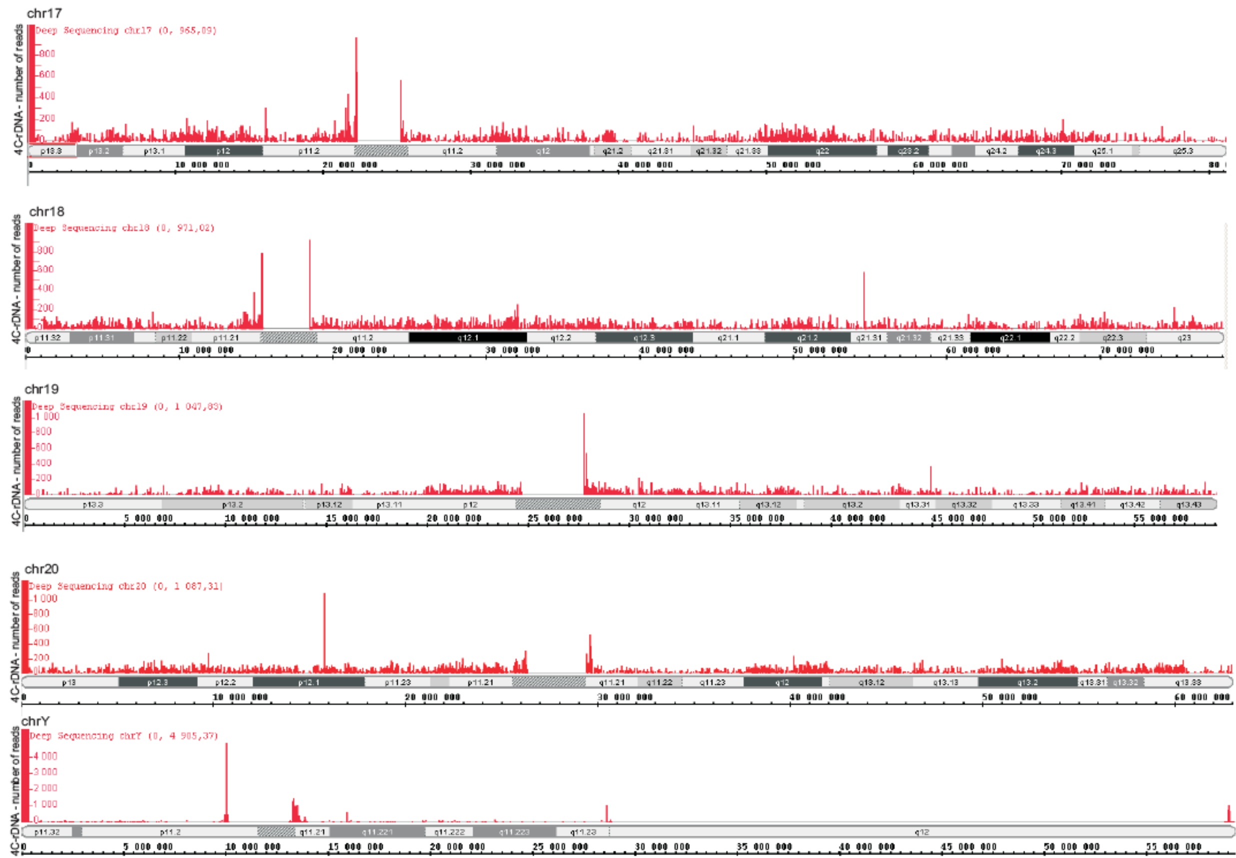
(A) A scheme showing the distribution of Sau3A sites (black bars), which together with sites at DSBs delimit the reads.

(B) The corresponding reads schematically are visualized by UGENE software (http://ugene.unipro.ru/). The numberings are according to the human ribosomal DNA complete repeating unit sequence (Accession number U13369). The reads in R6 are aligned in two sites. The left site is presented in 3230 rows. The reads in R9 also are presented in two adjacent sites. The left site is presented in 1856 rows. Only the top 594 reads are shown on the scheme.



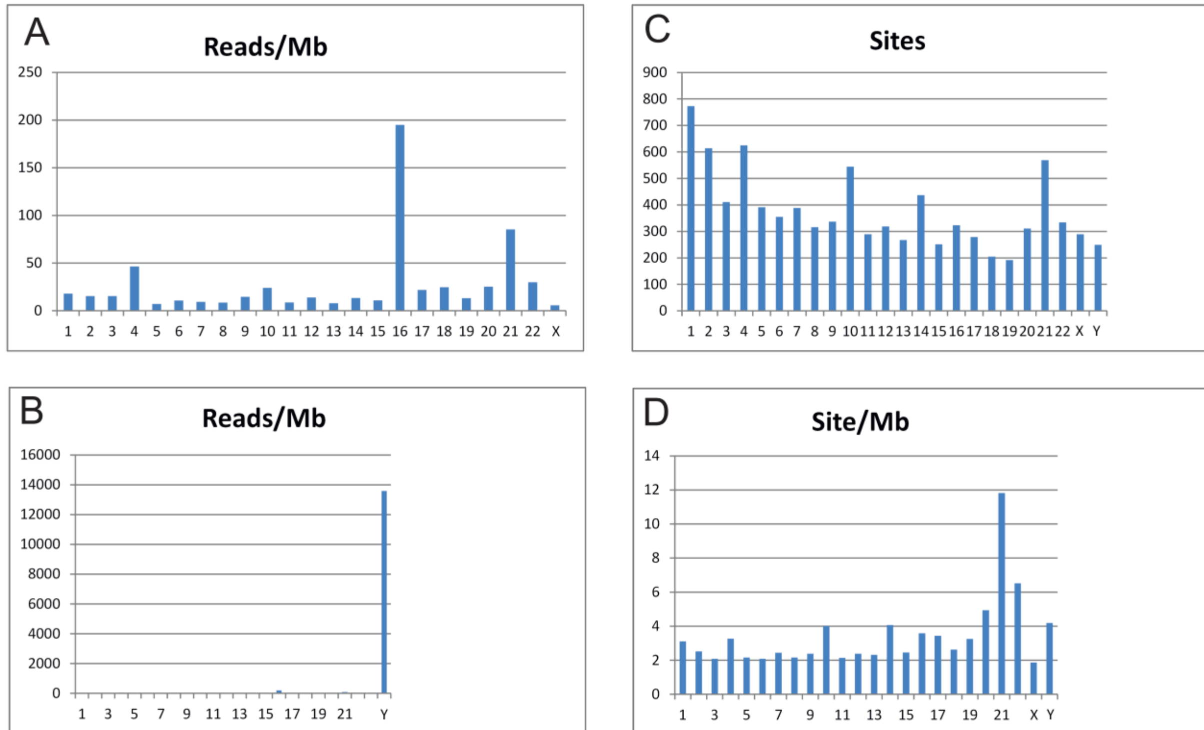
Supplementary Figure S6. Overviews of chr7, chr10, chr11, chr14, and chr16, and contact regions of rDNA.

Integrated Genome Browser (Affymetrix) was used. The profile of the 4C-rDNA reads is shown in red. Scale presents the number of mapped 4C-rDNA reads.



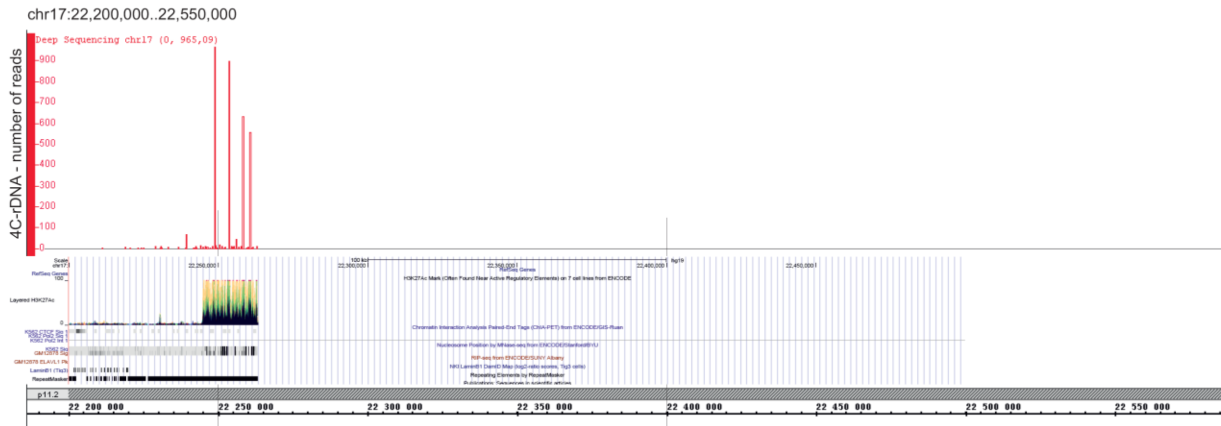
Supplementary Figure S7. Overviews of chr17, chr18, chr19, chr20, and Y, and contact regions of rDNA.

Integrated Genome Browser (Affymetrix) was used. The profile of the 4C-rDNA reads is shown in red. Scale presents the number of mapped 4C-rDNA reads.



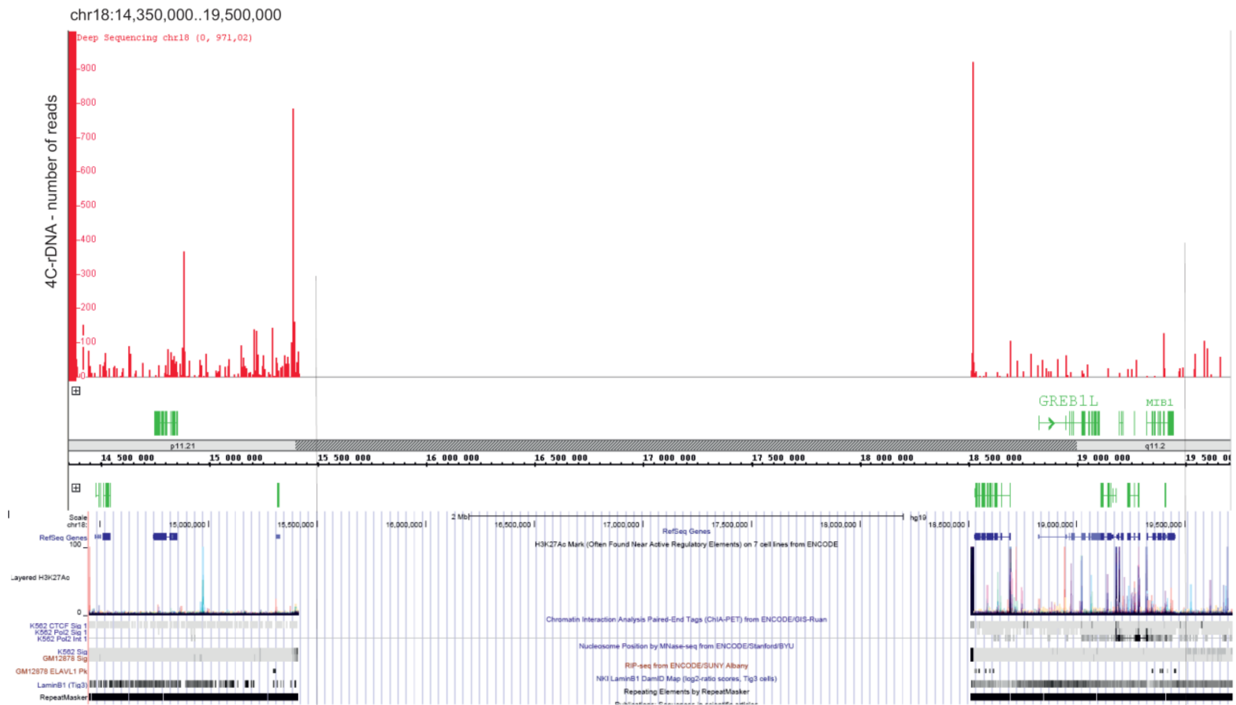
Supplementary Figure S8. Diagrams showing the number of 4C reads mapped in all chromosomes in HEK293T cells.

(A) and (B)—density of reads per Mb in different scales. (C) and (D) present the number of the rDNA contact sites in all chromosomes and the site density, respectively.



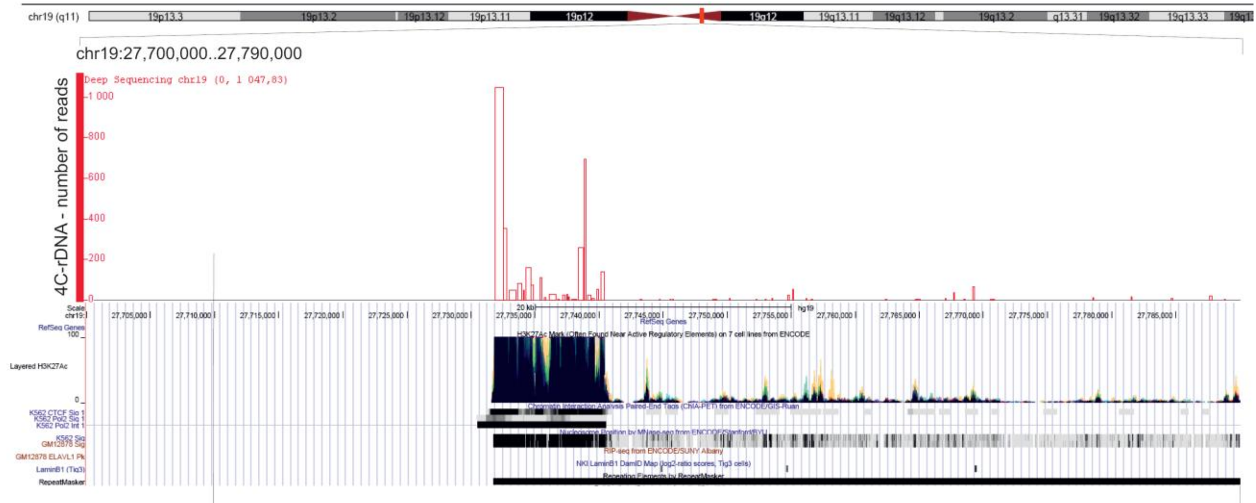
Supplementary Figure S9. Features of the chromosomal region where rDNA contacts were observed

The Integrated Genome Browser or UCSC Genome Browser was used with the human genome assembly of Feb. 2009 (GRCh37/hg19). Distribution of layered H3K27Ac marks, ChIA-PET signals (ENCODE/GIS-Ruan), nucleosome position (MNase-Seq from ENCODE/Stanford/BYU), and RIP-Seq data (ENCODE/SUNY Albany) are shown. Pericentromeric region of chr17 possesses overlapping profiles of rDNA contacts (shown in red) and H3K27Ac marks.



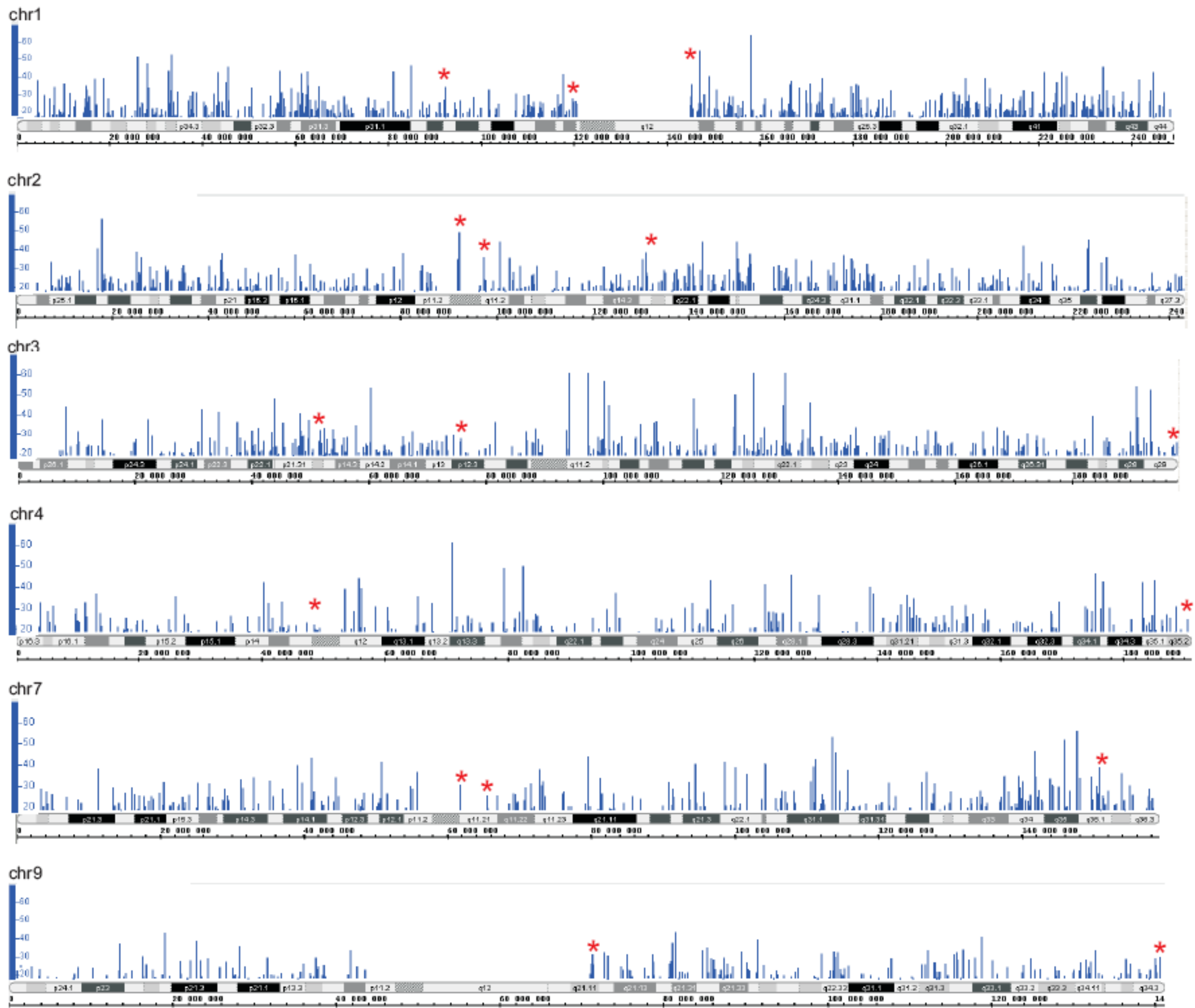
Supplementary Figure S10. Features of the chromosomal regions where rDNA contacts were observed

The Integrated Genome Browser or UCSC Genome Browser was used with the human genome assembly of Feb. 2009 (GRCh37/hg19). Distribution of layered H3K27Ac marks, ChIA-PET signals(ENCODE/GIS-Ruan), nucleosome position (MNase-Seq from ENCODE/Stanford/BYU), and RIP-Seq data (ENCODE/SUNY Albany) are shown. Both pericentromeric regions of chr18 possess overlapping profiles of rDNA contacts (shown in red), but only one, at “q”, coincides with the region possessing H3K27Ac marks.



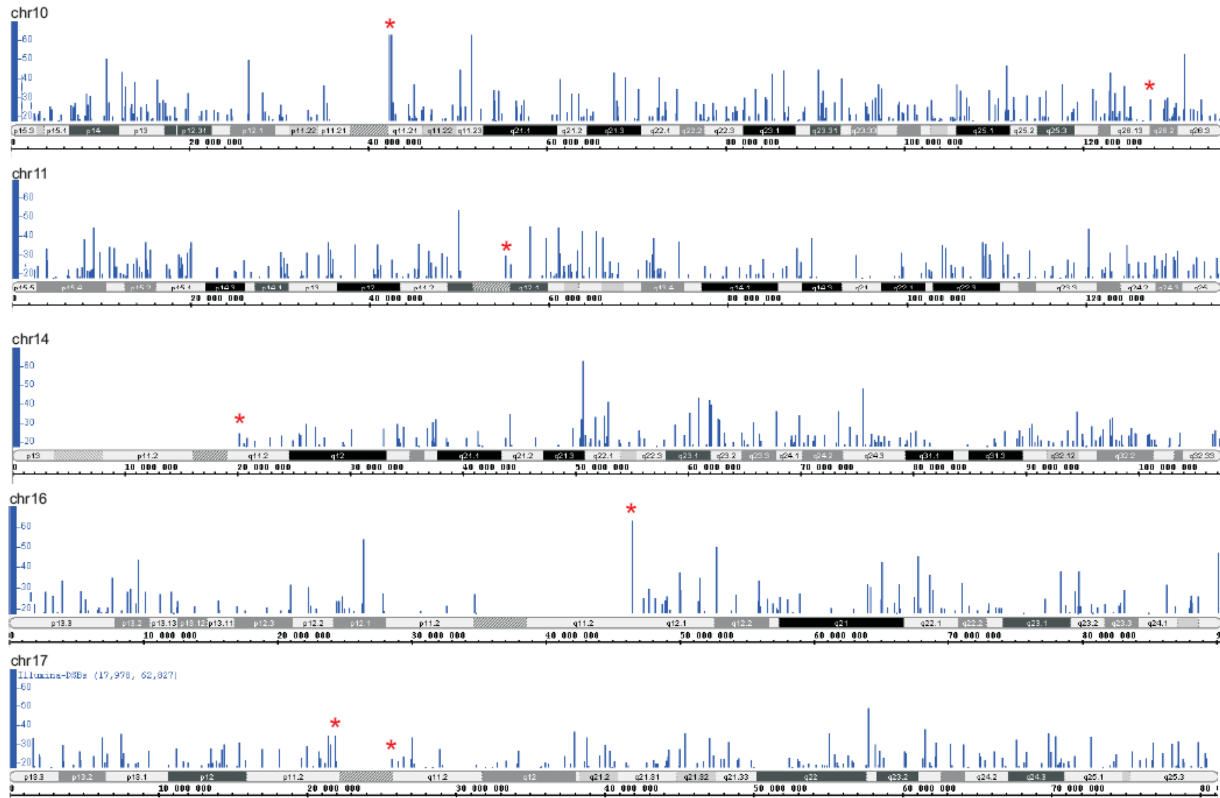
Supplementary Figure S11. Features of the chromosomal region where rDNA contacts were observed

The Integrated Genome Browser or UCSC Genome Browser was used with human genome assembly of Feb. 2009 (GRCh37/hg19). Distribution of layered H3K27Ac marks, ChIA-PET signals ChIA-PET (ENCODE/GIS-Ruan), nucleosome position (MNase-Seq from ENCODE/Stanford/BYU), and RIP-Seq data (ENCODE/SUNY Albany) are shown. Pericentromeric region of chr19 possess overlapping profiles of rDNA contacts (shown in red), with the region possessing H3K27Ac marks, higher nucleosome density, and high ChIA-PET signals).



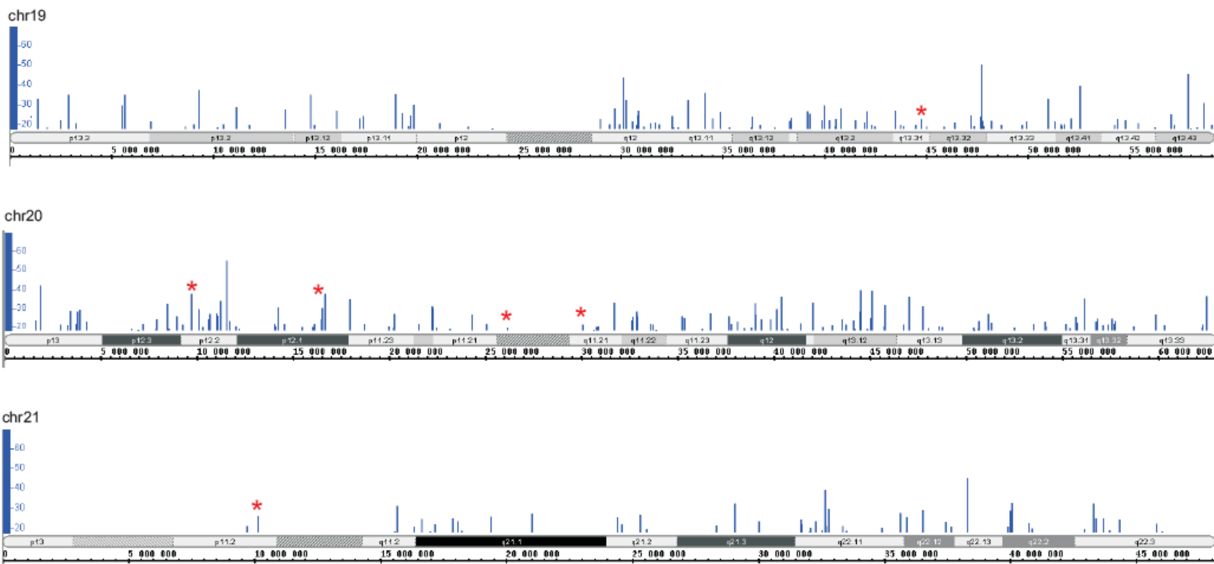
Supplementary Figure S12. Profiles of rDNA contacts along cytobands in chr1, chr4, chr7, and chr9.

Integrated Genome Browser (Affymetrix) and human Feb. 2009 (GRCh37/hg19) Assembly were used. The profiles of DSBs are shown along the cytobands. The scale presents amount of mapped DSBs possessing >17 reads. Red stars indicate the regions where DSBs coincide with rDNA contacts.



Supplementary Figure S13. Profiles of rDNA contacts along cytobands in chr10, chr11 chr14, chr16, and chr17.

Integrated Genome Browser ([Affymetrix]) and human Feb. 2009 (GRCh37/hg19) Assembly were used. The profiles of DSBs are shown along the cytobands. The scale presents amount of mapped DSBs possessing >17 reads. Red stars indicate the sites where hot spots of DSBs coincide with rDNA contacts. Mostly these sites are located at pericenromeric regions.

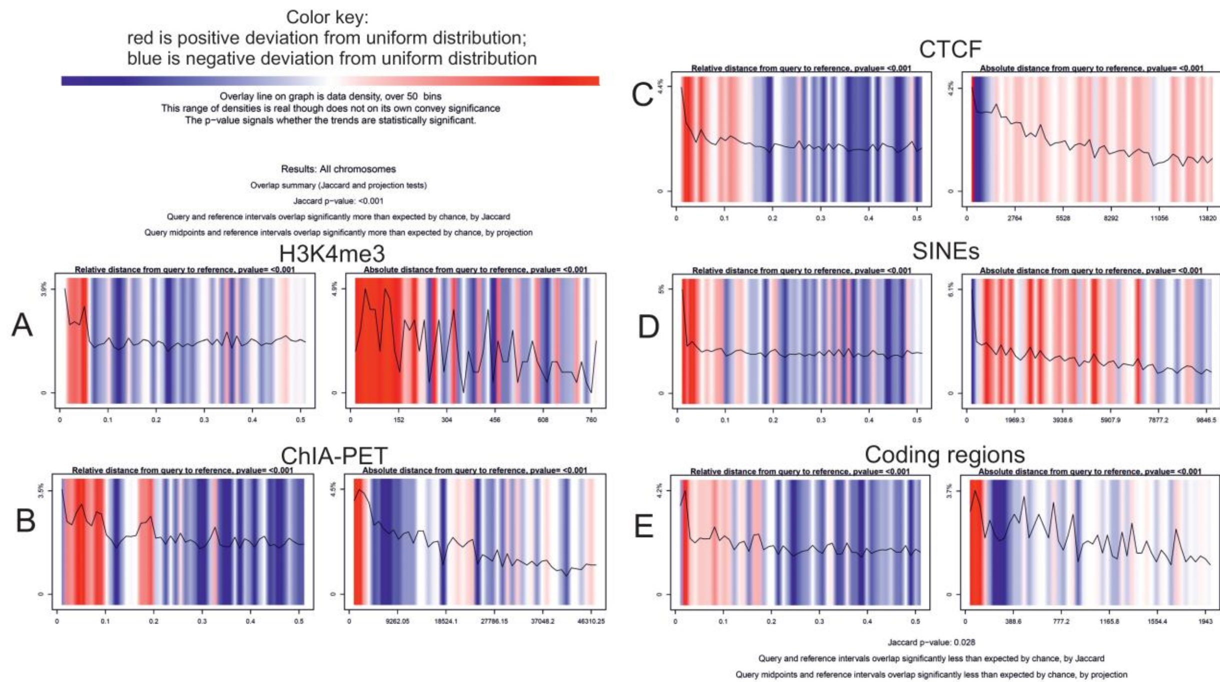


Supplementary Figure S14. Profiles of rDNA contacts along cytobands in chr19, chr20, and chr21.

Integrated Genome Browser (Affymetrix) and human Feb. 2009 (GRCh37/hg19) assembly were used. The profiles of DSBs are shown along the cytobands. The scale presents amount of mapped DSBs possessing >17 reads. Red stars indicate the sites where hot spots of DSBs coincide with rDNA contacts.

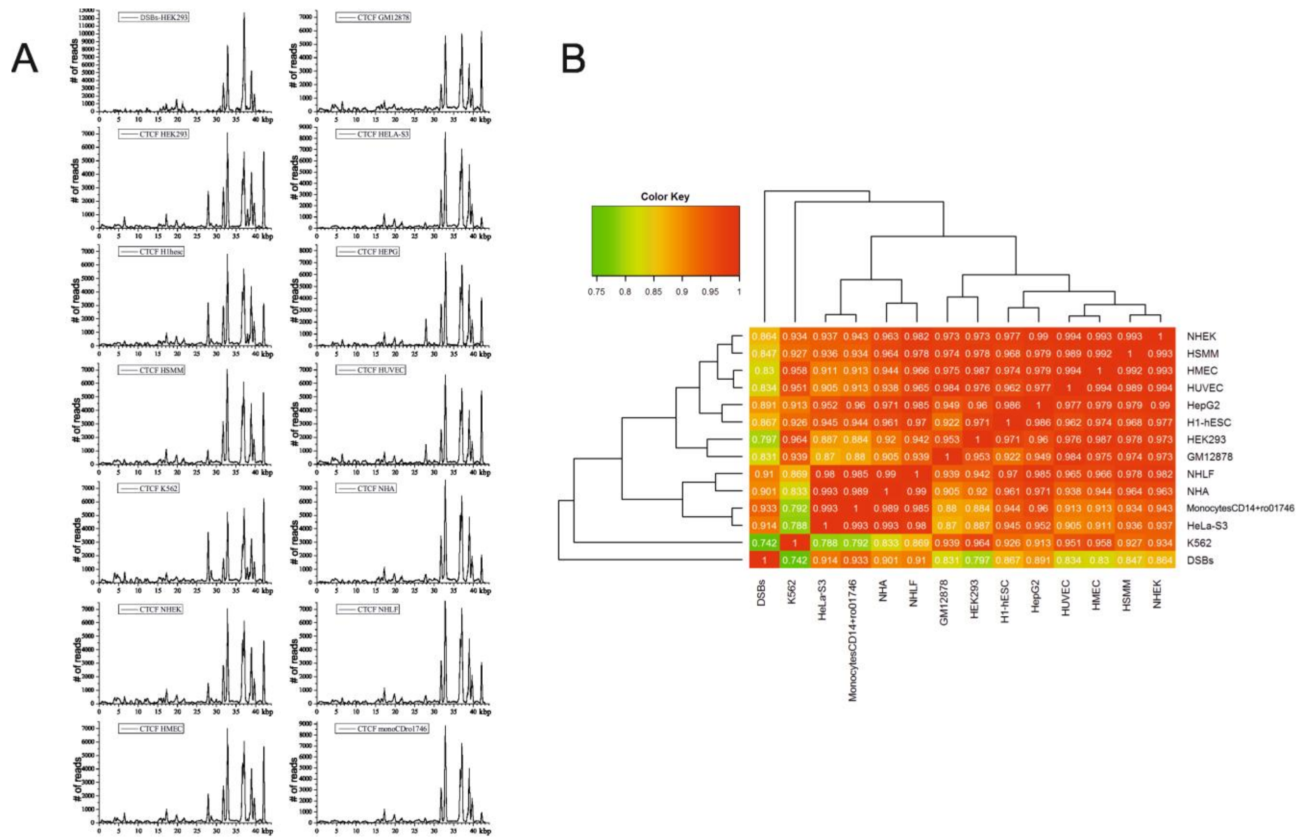
Supplementary Figure S15. The correspondence between profiles of DSBs and profiles of rDNA contacts at pericentromeric regions.

The Integrated Genome Browser was used with human genome assembly of Feb. 2009 (GRCh37/hg19). Distribution of DSBs and rDNA contacts are shown in chr4, chr10, and chr17.



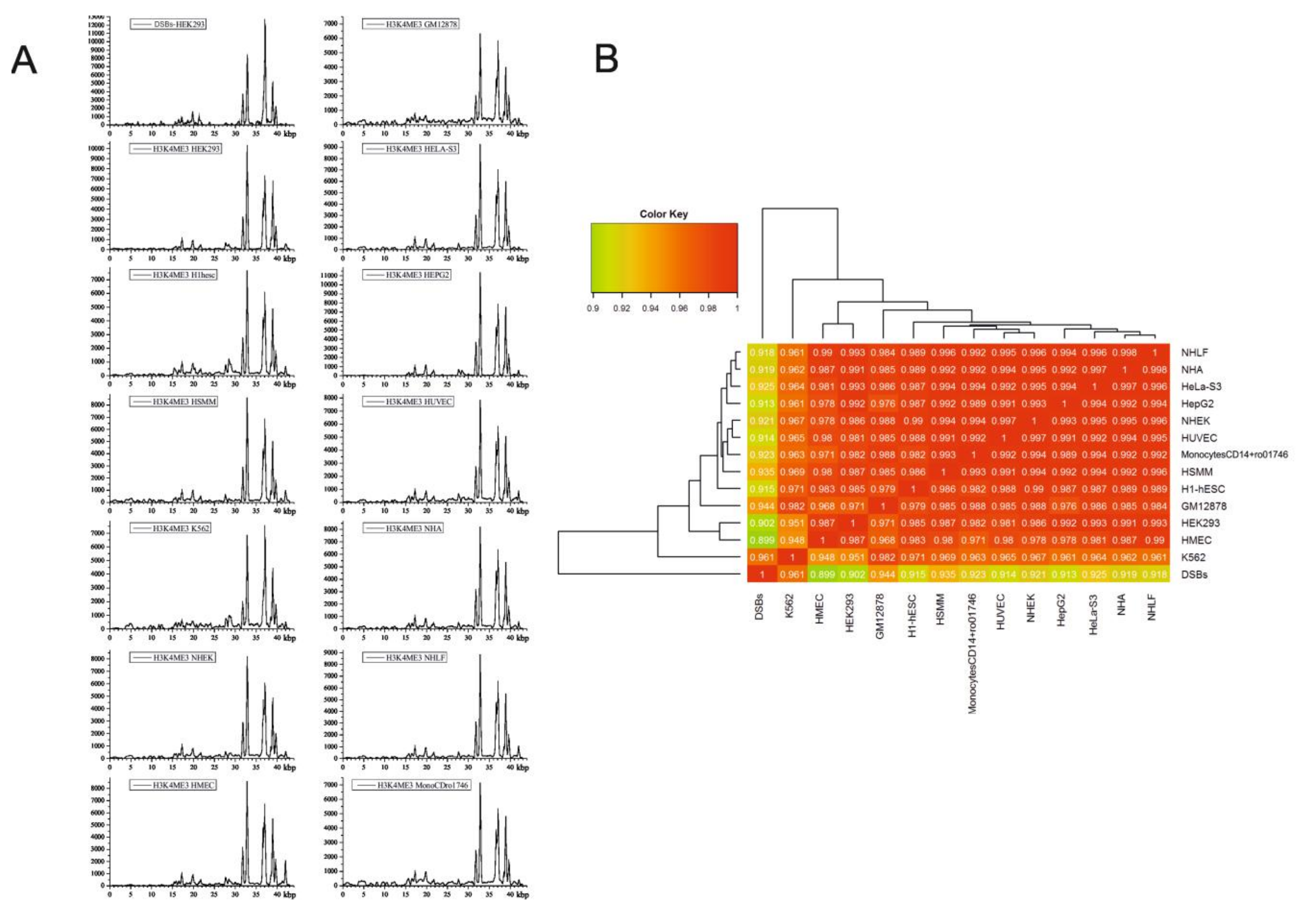
Supplementary Figure S16. Statistical analysis of spatial correlations between rDNA contact regions, H3K4me3 binding sites (A), ChIA-PET signals (B), CTCF binding sites (C), SINEs (D), and coding regions (E).

The statistical analysis was performed by the GenometriCorr package. Summary results for all chromosomes together are displayed in two panels. Color-coded density plots in each panel represent deviation from the expected uniform distribution. Red color indicates positive deviation from the expected uniform distribution; blue indicates negative deviation from the expected uniform distribution. The overlay line indicates the density of the data at each absolute or relative distance. Query corresponds to a set of intervals corresponding to the genomic contacts of rDNA units (4C-rDNA data).



Supplementary Figure S17. Comparison of profiles of DSBs and CTCF-binding sites that were detected in the rDNA of HEK293T cells with CTCF profiles detected in the rDNA of twelve additional cell lines.

(**A**) The profiles along the 43-kb rDNA unit. Publicly available ChIP-Seq ENCODE data were used. (**B**) Correlation heatmap of pairwise comparisons between median signals for DSBs and CTCF. All correlations in this heatmap are high (>0.74) and significant at $p < 0.05$.



Supplementary Figure S18. Comparison of profiles of DSBs and H3K4me3 marks that were detected in the rDNA of HEK293T cells with H3K4me3 profiles detected in the rDNA of twelve additional cell lines.

(**A**) The profiles along the 43-kb rDNA unit. Publicly available ChIP-Seq ENCODE data were used. (**B**) Correlation heatmap of pairwise comparisons between median signals for DSBs and H3K4me3. All correlations in this heatmap are high (>0.9) and significant at $p < 0.05$.

Supplementary Tables

Supplementary Table S1. Statistics on DSBs density inside the 43-kb rDNA unit and in the whole human genome.

Name	Mean	Std. Deviation	Median	[2.5%, 97.5%]	[25%, 75%]
rDNA	1234439		1234439		
Genome	221.66		203.833		
chr1	341.579	409.55	311	[0.000, 843.000]	[191.000, 452.000]
chr2	219.601	127.12	204	[15.000, 500.000]	[133.000, 286.000]
chr3	211.592	123.87	195	[16.000, 506.000]	[125.000, 278.000]
chr4	178.228	110.55	163	[0.000, 429.000]	[99.000, 239.000]
chr5	216.53	131.08	196	[12.000, 528.000]	[125.000, 286.000]
chr6	241.178	135.17	224	[13.000, 551.000]	[145.000, 318.000]
chr7	225.915	142.38	204	[0.000, 545.000]	[129.000, 300.000]
chr8	194.316	116.66	180	[0.000, 449.000]	[114.000, 258.000]
chr9	194.121	144.53	182	[0.000, 516.000]	[87.000, 282.000]
chr10	272.905	358.40	248	[0.000, 606.000]	[164.000, 352.000]
chr11	221.822	138.86	204	[0.000, 527.000]	[129.000, 295.000]
chr12	244.523	151.24	220	[0.000, 612.000]	[138.000, 323.000]
chr13	173.964	133.79	164	[0.000, 465.000]	[67.000, 261.000]
chr14	190.687	161.71	181	[0.000, 508.000]	[75.000, 282.000]
chr15	215.319	182.95	196	[0.000, 642.000]	[65.000, 320.000]
chr16	253.762	308.50	233	[0.000, 655.000]	[127.000, 346.000]
chr17	315.097	175.80	300	[0.000, 686.000]	[193.000, 422.000]
chr18	169.268	124.13	145	[0.000, 475.000]	[80.000, 234.000]
chr19	356.186	394.60	331	[0.000, 794.000]	[211.000, 467.000]
chr20	249.634	148.13	233	[0.000, 585.000]	[147.000, 334.000]
chr21	123.029	129.40	102	[0.000, 434.000]	[0.000, 186.000]
chr22	270.855	242.64	268	[0.000, 782.000]	[0.000, 439.000]
chrX	227.51	141.60	208	[0.000, 540.000]	[133.000, 301.000]
chrY	12.422	82.44	0	[0.000, 125.000]	[0.000, 0.000]

The whole genome with mapped DSBs was scanned with a 43-kb window and 1-kb steps. The data on DSBs density for the 43-kb window in the 43-kb rDNA unit, in the whole sequenced genome, and in individual chromosomes are shown. Even assuming that about 300 copies of rDNA are present per genome (Stults et al., 2009), the average DSBs density per single rDNA unit is significantly higher than the average DSBs density for the 43-kb window in any chromosome (1234439/300 = 4114).

Supplementary Table S2. Statistical tests of correlations between DSBs, CTCF binding sites, H3K4me3 marks, DNase-I-hypersensitive sites (DNase-I-HSS), and DNA methylation sites inside rDNA units.

Tracks	Projection test	Jaccard test	Scaled absolute minimum distances test
DSBs-H3K4ME3	Passed, $p = 1.315e^{-5}$	Passed, $p = 0.001$	Passed, $p = 0.019$
DSBs-CTCF	Passed, $p = 1.552e^{-7}$	Passed, $p = 0.001$	Passed, $p = 0.061$
DSBs-DNAseI-HSS	Failed, $p = 0.116$	Passed, $p = 0.005$	Passed, $p = 0.037$
DSBs-MethylRRBS	Passed, $p = 0.040$	Failed, $p = 0.053$	Failed, $p = 0.201$

Statistical tests were performed using GenometriCorr package (Favorov et al., 2012. Exploring massive, genome scale datasets with the GenometriCorr package. PLoS Comput. Biol. *8*, (5):e1002529). The DSBs track was used as the reference, and all other tracks were used as query. GenometriCorr can process only segmented tracks (not profiles) in formats like BED, so we defined peak regions for H3K4ME3, CTCF, DNAseI-HSS, and Methyl-RRBS from their alignments with the rDNA sequence by MACS2 2.1.0 peak caller (Zhang et al., 2008) Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137. doi: 10.1186/gb-2008-9-9-r137] with parameters -nomodel -g=42999. DSBs were preliminary filtered at the threshold of 10.0 coverage. Results that are presented in this table are consistent with the visual analysis and reveal more information than simple correlation values in the case of partial success tests. In the case of the DNAseI-HSS track, the projection test failed, whereas the Jaccard and scaled absolute minimum distances tests were passed, which means that DSBs and DNAseI-HSS do not overlap more than by chance (which we can also see from a visual analysis in Figure 3A), but in the cases in which they are close to each other (at the IGS in Figure 3A), they are closer than by chance.

Supplementary Table S3. Distribution of 4C-rDNA reads mapped in chromosomes of HEK293T cells

Chromosomes	Reads	Reads/Mb	Sites	Sites /Mb
chr1	4486	17.9979	773	3.1013
chr2	3726	15.3208	614	2.5247
chr3	3024	15.271	411	2.0755
chr4	8878	46.4442	625	3.2696
chr5	1281	7.0807	391	2.1612
chr6	1837	10.7355	355	2.0746
chr7	1480	9.3001	388	2.4381
chr8	1251	8.5472	316	2.159
chr9	2065	14.6233	337	2.3865
chr10	3248	23.9643	544	4.0137
chr11	1182	8.7551	289	2.1406
chr12	1878	14.0304	319	2.3832
chr13	900	7.8145	267	2.3183
chr14	1439	13.4048	437	4.0708
chr15	1114	10.865	251	2.448
chr16	17618	194.987	323	3.5748
chr17	1779	21.9102	279	3.4362
chr18	1925	24.6551	205	2.6256
chr19	773	13.0731	192	3.2471
chr20	1587	25.1803	311	4.9345
chr21	4106	85.3108	569	11.8222
chr22	1529	29.8024	334	6.5101
chrX	886	5.7062	289	1.8613
chrY	806364	13581.1954	249	4.1938

Number of reads per Mb illustrates the frequencies of rDNA contacts with a particular chromosome. Sites indicate the group of reads covering the same region in a chromosome.

Supplementary Table S4. Sequences at the most frequent contacts of rDNA units in HEK293T cells

chr #	# of reads	Sequence	Location
chr16	2744	ctcgccttctctctggctctcgctatctcccaccctctctttcttTGCAAAATAAGTTCAAGTACGTCTAATCTAATCCATTACCACGGCCTGAATTCTTAACTTTAGACATCCCAGATTTGATCTCCCTACAGAATGCTGTACAGAACTGGCGAGTTGATTTCTGGACTTGGATTCCTCAT	223670 bp at 5' side: splicing factor, arginine/serine-rich 19-like
chr16	13208	TctttgaaaaaaaaatcccagaagtggctttgactttttggctaggaggcctaaatctgctgagaactttcctgcccaggattctgtgtgaacaaaaTTGCCTCTGCTGGGAGCTgggatcctcgggaccatgcttgctagcgctggatga	190549 bp at 5' side: Ig heavy chain V-I region V35-like
Y	802960	CtggccctttgaaaaaaatcccagaggtggctttggctttttggctaggaggcctaaacctgctgagaactttcctgcccaggatcctgtgtgaacaaaagtGCCTCTGCTGGGAGCTGGGATCCTTGGGACCATGCTTGCT	208863 bp at 3' side: uncharacterized protein LOC648761

The first two rows present the sequences from pericentromeric regions in chr16 that correspond to the most prominent peaks in Figure S6. The third row presents the sequence at the most frequent contact of rDNA units in the human genome, which is located in the pericentromeric region of the Y chromosome, and corresponds to the most prominent peaks in Figure S7.

Supplementary Methods

Isolation of DNA preparations used for genome-wide mapping of spontaneous DSBs

Preparation of DNA-agarose plugs

DNA-agarose plugs were prepared as described by Tchurikov and Ponomarenko, 1992. About 6 million HEK293T cells in 2 ml of culture medium were pelleted by centrifugation at 2000 rpm, resuspended in 0.3 ml of the same medium, gently mixed at 42 ℃ with an equal volume of 1% agarose L (LKB) in DMEM, and distributed in a mold containing 100-μL wells. The mold was placed on ice for 2–5 min and covered with Parafilm. The agarose plugs were then placed in Petri dishes with 5 ml of solution containing 0.5 M EDTA (pH 9.5), 1% sodium laurylsarcosine, and 1–2 mg of proteinase K solution per ml for 40–48 h at 50 ℃. They were stored at 4 ℃ in the same solution. Each DNA-agarose plug usually contained about 15 μg of DNA corresponding to about one million cells.

To test the quality of isolated DNA, fractionation in PFGs was performed as described previously (Tchurikov and Ponomarenko, 1992). Portions of the original agarose-DNA plugs (5–50 μl) containing 1–10 μg of DNA were used for electrophoresis without any restriction enzyme digestion. The DNA samples were run in 0.8% agarose gels on an LKB Pulsaphor system using a hexagonal electrode and switching times of 25, 100, or 450 sec.

For elution of DNA preparations, fractionation in a 1% agarose conventional mini-gel was performed. One-half of a DNA-agarose plug was washed in 1×TE three times (for 15 min each) followed by washing three times in the same solution containing 17.4 μg/ml PMSF in ethanol. After fractionation in the mini-gel, the ethidium-bromide-stained DNA band was excised and electroeluted inside a cellulose membrane dialysis bag. After overnight dialysis without stirring against 1 L of 0.01×TE at 4 ℃, the DNA was concentrated with PEG (4 ℃) and re-dialyzed.

Isolation of DNA for PCR amplification

DNA termini at DSBs were isolated as described by Tchurikov et al., 2011. About 1.5 μg of isolated DNA (see above) were treated by Klenow fragment of *E. coli* DNA polymerase I and then ligated with 70 ng of double-stranded oligonucleotide (25-bp-long 5'-phosphorylated 5' pCCCCTGCAGTATAAGGAGAATTCGGG 3' oligonucleotide annealed with 26-bp-long 5' biotinylated 5' bio-CCGAATTCTCCTTATACTGCAGGGG 3' oligonucleotide) in 150 μl of

solution containing 0.1 M NaCl, 50 mM Tris-HCl (pH 7.4), 8 mM $MgCl_2$, 9 mM 2-mercaptoethanol, 7 μM ATP, 7.5% PEG, and 40 units of T4 DNA ligase at 20 °C for 16 h. After heating at 65 ℃ for 10 min, the DNA preparation was digested with Sau3A enzyme to shorten the forum domain to the termini attached to the ligated oligonucleotide. The selection of such termini was performed in 0.5 ml Eppendorf tubes using 300 μl of suspension containing streptavidin paramagnetic particles (SA-PMP, Promega, Madison, WI, USA) according to the manufacturer's recommendations. After extensive washing with 0.5×SSC to remove DNA fragments corresponding to internal parts of DNA domains, the termini (T) attached at sites of DSBs were eluted from the SA-PMP using digestion with *Eco*RI enzyme in a final volume of 50 μl (double-stranded T). The T was then ligated with 100× molar excess of double-stranded Sau3A adaptor (5'-phosphorylated 5' pGATCGTTTGCGGCCGCTTAAGCTTGGG 3' oligonucleotide annealed with 5' CCCAAGCTTAAGCGGCCGCAAAC 3' oligonucleotide). In some experiments, the T DNA preparations were eluted from the SA-PMP using heating via incubation at 100 ℃ for 3 min in 50 μl of 0.01×TE (single-stranded T). Before heating, the T preparation was ligated with 100× molar excess of double-stranded Sau3A adaptor in suspension with SA-PMP (see above). Both final DNA samples (double-stranded T or single-stranded T) were used for PCR amplifications. Forty cycles of PCR amplification in 30 μl of a solution containing 67 mM Tris-HCl (pH 8.4); 6 mM MgCl2; 10 mM 2-mercaptoethanol; 16.6 mM ammonium sulfate; 6.7 μM EDTA; 5 μl 5 mg/ml BSA; 1 mM dNTPs; 1 μg of primer corresponding to the Sau3A adaptor (5' CCCAAGCTTAAGCGGCCGCAAAC 3'); 1 μg of primer corresponding to the biotinylated oligonucleotide (5' CCGAATTCTCCTTATACTGCAGGGG 3'), and 1 u of Taq polymerase were performed using Eppendorf Mastercycler Personal. Amplification conditions were 90 ℃ for melting, 65 ℃ for annealing, and 72 ℃ for extension, for 1 min each.

Library preparation

After purification in 2% agarose gels, DNA was eluted using QIAquick gel extraction kit (Qiagen). The libraries were prepared using TruSeq RNA Sample Preparation Kit v2 (Illumina) using adapter AR004. The samples were sequenced on a Genome Analyzer IIx machine using TruSeq SR Cluster Kit v2–cBot-GA kit (Illumina).

4C procedure

DNA samples were prepared as described previously (Dekker et al., 2002; Osborne et al., 2004; http://www.protocol-online.org/cgi-bin/prot/view_cache.cgi?ID=3978) and schematically is illustrated in Figure 5A.

Cross-linking

To a suspension containing about 20×10^6 HEK293T cells in 40 ml of DMEM, formaldehyde solution was added to final concentration 1.5%. After mixing, incubation of cell suspension was performed for 10 min at room temperature with mixing. Quenching with 2.75 ml of 2 M glycine (final concentration 0.125 M) was performed. After incubation at room temperature for 5 min the suspension was cooled for 15 min in an ice bath and then cells were collected by centrifugation for 15 min at 3500 rpm at 2 ℃.

Lysis

The pellet of cells was resuspended at 0 ℃ in 1 ml of buffer containing 10 mM Tris-HCl buffer, pH 8, 10 mM NaCl, 0.2% NP-40, and freshly added protease inhibitors (0.1 mM PMSF and 1:500 protease inhibitor cocktail (Sigma)). After incubation for 15 min, cells were homogenized by passing through a syringe about 50 times into an Eppendorf. Then nuclei were spun down by centrifugation for 5 min at 5000 rpm in an Eppendorf centrifuge 5415 R at 2 ℃.

Digestion with EcoRI

The nuclei pellet was resuspended in 756 µl of solution containing 40 mM Tris-HCl buffer, pH 7.4, 50 mM NaCl, 10 mM MgCl2 and 10 mM 2-mercaptoethanol. Then 20% SDS was added to final concentration 0.3% and incubation with shaking was performed for 1 h at 37 ℃. To sequester SDS, 180 µl of 10% Triton X-100 was added and the solution was incubated for 1 h at 37 ℃. Before digestion, 1 µl of BSA (5 mg/ml) was added with mixing. Then 50 µl of EcoRI (10 u/µl) was added and after mixing digestion was performed overnight at 37 ℃.

Ligation

To inactivate the restriction enzyme, 35 µl of 20% SDS was added (final concentration 0.7%) and the probe was heated to 65 ℃ for 30 min. Then the mixture was transferred into a 15 ml

Nunc tube and consequently 375 μl of 20% Triton X-100 (to final concentration 1%), 750 μl of 10× ligase buffer, 7.5 μl of BSA (5 mg/ml), 80 μl of 100 mM ATP, and 5241 μl of milliQ water were added and the final 7.5 ml solution was well mixed. Then 10 μl of T4 DNA ligase (200 u/μl) was added and after mixing incubation was performed for 5 h at 16 ℃ and then for 30 min at room temperature. During ligation, DNA concentration was equal to 2 ng/μl.

DNA purification

For isolation of DNA, 50 μl of proteinase K (10 mg/ml) was added (final concentration 50 μg/ml) and, after mixing, incubation was performed at 55 ℃ overnight. For RNA digestion, 40 μl of RNase A (10 mg/ml) was added (final concentration 0.5 μg/ml) and, after mixing, incubation was performed for 30 min at 37 ℃. After extensive extraction with phenol-chloroform extraction (three times with 7 ml each) DNA was precipitated by 2.5 vol. of ethanol after addition of 40 μl 10 mg/ glycogen and 175 μl 4 M NaCl. The final DNA pellet was washed twice with 70% ethanol and then dissolved in 0.1×TE.

Digestion with FaeI

About 15 μg of DNA was digested in 250 μl solution with 75 u of FaeI overnight at 37 ℃. Then the enzyme was inactivated by heating at 65 ℃ for 30 min. DNA was isolated after phenol-chloroform extraction, precipitated by ethanol, and dissolved in 100 μl of 0.1×TE.

Ligation

For circularization, 15 μg of DNA was incubated in 8 ml of T4 DNA ligase buffer containing 400 u of T4 DNA ligase for 5 h at 16 ℃. DNA was isolated after phenol-chloroform extraction, precipitated by ethanol, and dissolved in 50 μl of 0.1×TE.

PCR and library preparation

One or two rounds of PCR were used as described below. There are about 300-400 copies of rDNA, which is why a single round of PCR (with up to 35 cycles) was found to be sufficient. DNA concentration was titrated and finally about 30 ng of DNA was used for PCR with primers 5’ TCTTTGAAAAAAATCCCAGAAGTGGT 3’ and 5’ AAGTCCAGAAATCAACTCGCCAGT 3’ (for PCR-1), and 5’

GCCTAAGCCTGCTGAGAACTTTC 3’ and 5’ CAGCATTCTGTAGGGAGATCAAATC 3’ (for PCR-2). After separation in 2% agarose gels, two DNA fractions (200-400 bp and higher than 400 bp) were eluted using a QIAquick gel extraction kit (Qiagen). The libraries were prepared using TruSeq RNA Sample Preparation Kit v. 2 (Illumina) using adapter AR006 for 200-400 bp DNA fraction and adapter R007 for DNA fractions higher than 400 bp. The samples were sequenced using MiSeq (Illumina).

Computer treatments

The data on DSBs were processed in the following way: the raw data were obtained using an Illumina Genome Analyzer IIx machine. The data then were decoded to FASTQ format using native Illumina Casava software, v. 1.8. The next step was to cut off two primers: primer_HindIII_NotI—CCCAAGCTTAAGCGGCCGCAAAC and primer_EcoRI_PstI—CCGAATTCTCCTTATACTGCAGGGG. We have tested different software designed for such tasks (including our own scripts/wrappers around FASTA, FASTA/Q clipper from FASTX-toolkit, cutadapt and trimmomatic), and found that the best solution for our dataset is the cutadapt tool (http://code.google.com/p/cutadapt), because it has an option to control the alignment error rate. Cutadapt version 1.2.1 (http://code.google.com/p/cutadapt) was used for this task with the parameters "--minimum-length=30 --error-rate=0.02 --trimmed-only --quality-base=33 quality-cutoff=3 -n 2 -g CCCAAGCTTAAGCGGCCGCAAAC -g CCGAATTCTCCTTATACTGCAGGGG". As a result, the primer cut-off was performed with the assumption that any of the primers should be at either end of a read and only sequences with the found and clipped primers are selected for further analysis.

All sequences shorter than 30 bp were removed from the dataset (parameter—minimum-length=30 for cutadapt). We have inspected our clipped data by FastQC 0.10.1 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) to be sure that they are good enough to obtain reliable mappings and that they do not contain over-represented sequences. We have tested the set of programs for the mapping of the reads: bowtie (http://bowtie-bio.sf.net), bowtie2 (http://bowtie-bio.sourceforge.net/bowtie2/), BWA (http://bio-bwa.sourceforge.net) methods bwasw and mem. The results comparison revealed that the best result in our case were produced by the BWA method mem—it produced more unique and non-unique alignments. Thus the final mappings were performed using the BWA 0.7.5a method mem (http://bio-bwa.sourceforge.net)

and samtools 0.1.19 (http://samtools.sourceforge.net) with the human ribosomal DNA repeating unit GenBank accession number U13369 (masked by RepeatMasker http://www.repeatmasker.org with Repbase http://www.girinst.org), and *Homo sapiens* masked genome (assembly GRCh37p10/hg19) taken in the form of MFA files from ftp://ftp.ncbi.nih.gov/genomes/H_sapiens/Assembled_chromosomes/seq.

For Circos presentation, the complete rDNA unit was included at a tip of chromosome 14 of *Homo sapiens* genome from the very beginning of the sequence. The alignments were converted to a convenient table format and to BED/WIG files by *ad hoc* Perl scripts. The same procedure was applied to 4C-rDNA data obtained by the Illumina MiSeq machine. The only difference is the primers that we have to remove—in this case they were CTGGTGAGGGCATG and GAATTCCCTGCTTCCCATT. Both raw data and mapping results were deposited into the GEO database.

Circular presentations of 4C-rDNA data were created by the means of Circos 0.64 software (http://www.circos.ca). We converted our tables to source data for Circos by means of *ad hoc* Perl scripts, thresholds have been chosen to obtain unobstructed pictures to display the most prominent results.

Quantitative real-time PCR analysis of DSBs inside rDNA units

To estimate the amount of damaged DNA stretches that possess the mapped DSBs at R5 (Figure 4), we performed real-time PCR across the R5 using Applied Biosystems' 7500 Real-Time PCR System. DNA preparations from HEK 293T cells were isolated in solution by a procedure that simulates the isolation of DNA in DNA-agarose plugs. About 1.5×10^6 cells in 500 µl of DMEM containing 10% FBS were precipitated in 1.5 ml Eppendorf tubes in a minispin centrifuge (1500 rpm for 2 min) at room temperature. They were then suspended in 50 µl of the same medium at 42 ºC. For isolation of control DNA ("intact"), the suspension of cells was immediately mixed with 100 µl of solution containing 0.5 M EDTA (pH 9.5), 1% sodium laurylsarcosine, and 2 mg of proteinase K solution per ml. After incubation of the mixture for 48 h at 50 ℃, the control DNA was isolated by phenol-chloroform extractions followed by isopropanol precipitation. To isolate "damaged" DNA (*i.e.,* possessing spontaneous DSBs), the suspension of cells in 50 µl of DMEM containing 10% FBS after 2–3 min incubation at 42 ℃ was additionally incubated for 5 min at 0 ℃, and then DNA was isolated as described above for the control DNA. In the

experiments, HEK293T cells were incubated with 2 mM hydroxyurea for 18 h, as described by Gagou et al. (2010), and then used for DNA preparations as“intact” DNA. Heat shock treatment of HEK293T cells was performed by incubation at 43 ℃ for 20 min followed by incubation at 37 ℃ for 2.5 h, as described by Sonna et al. (2002).

Supplementary References

Gagou ME, Zuazua-Villar P, Meuth M. Enhanced H2AX phosphorylation, DNA replication fork arrest, and cell death in the absence of Chk1. Mol Biol Cell. 2010;21:739-752.

Sonna LA, Gaffin SL, Pratt RE, Cullivan ML, Angel KC, Lilly CM. Effect of acute heat shock on gene expression by human peripheral blood mononuclear cells. J Appl Physiol. 2002;92:2208-2220.

Tchurikov NA, Ponomarenko NA. Detection of DNA domains in *Drosophila*, human and plant chromosomes possessing mainly 50- to 150-kilobase stretches of DNA. Proc Natl Acad Sci USA. 1992;89:6751-6755.

Tchurikov NA, Kretova OV, Sosin DV, Zykov IA, Zhimulev IF, Kravatsky YV. Genome-wide profiling of forum domains in *Drosophila melanogaster*. Nucl Acids Res. 2011;39:3667-3685.

Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. Science. 2002;295:1306-1311.

Osborne CS, Chakalova L, Brown KE, Carter D, Horton A, Debrand E, Goyenechea B, Mitchell J, Lopes S, Reik W, Fraser P. Active genes dynamically colocalize to shared sites of ongoing transcription. Nature Genetics. 2004;36:1065-1071.