Angewandte
Eine Zeitschrift der Gesellschaft Deutscher Chemiker
Chemie

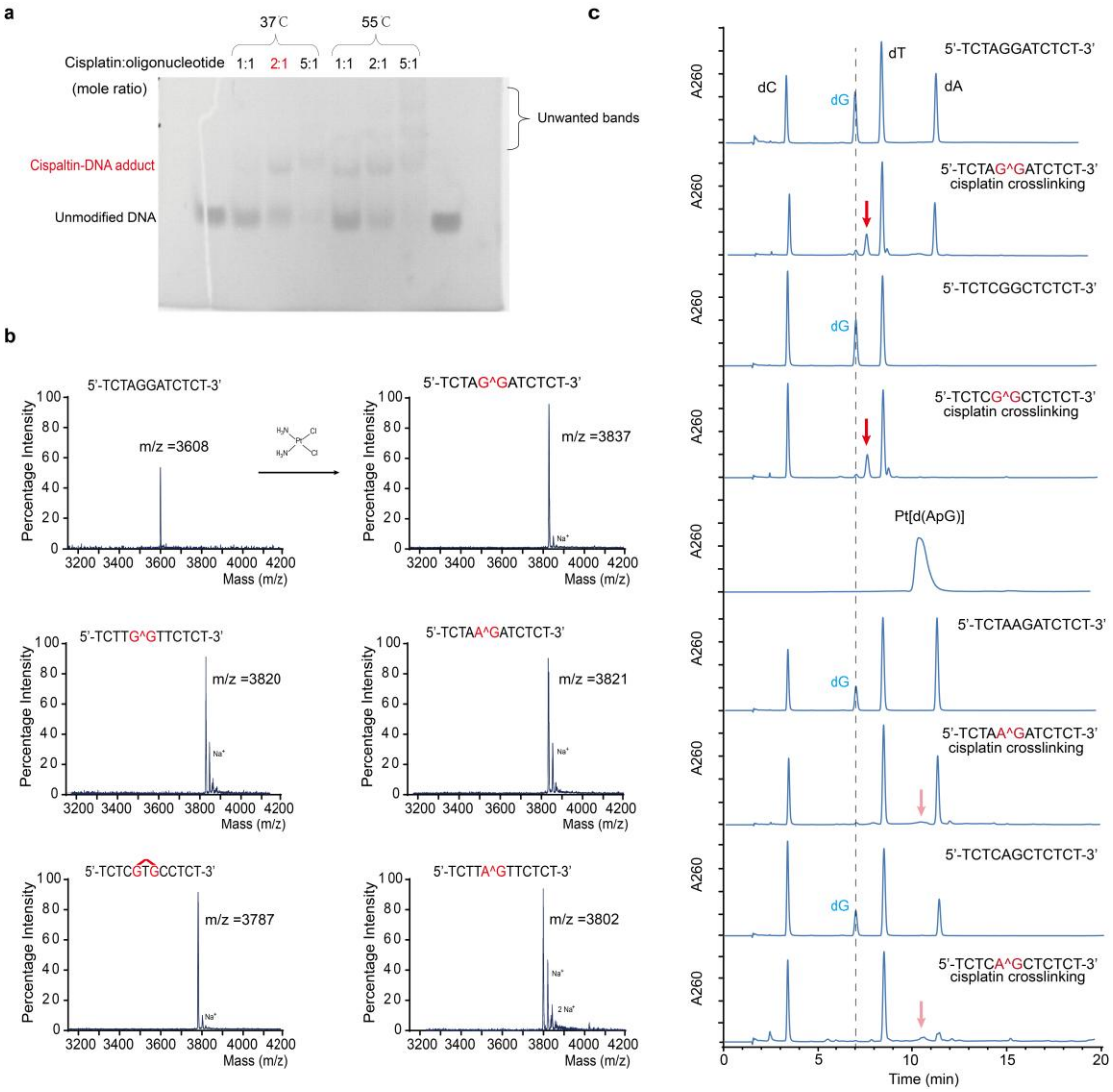
Supporting Information

Base-Resolution Analysis of Cisplatin–DNA Adducts at the Genome Scale

Xiaoting Shu[+], *Xushen Xiong*[+], *Jinghui Song, Chuan He,* and Chengqi Yi**

anie_201607380_sm_miscellaneous_information.pdf

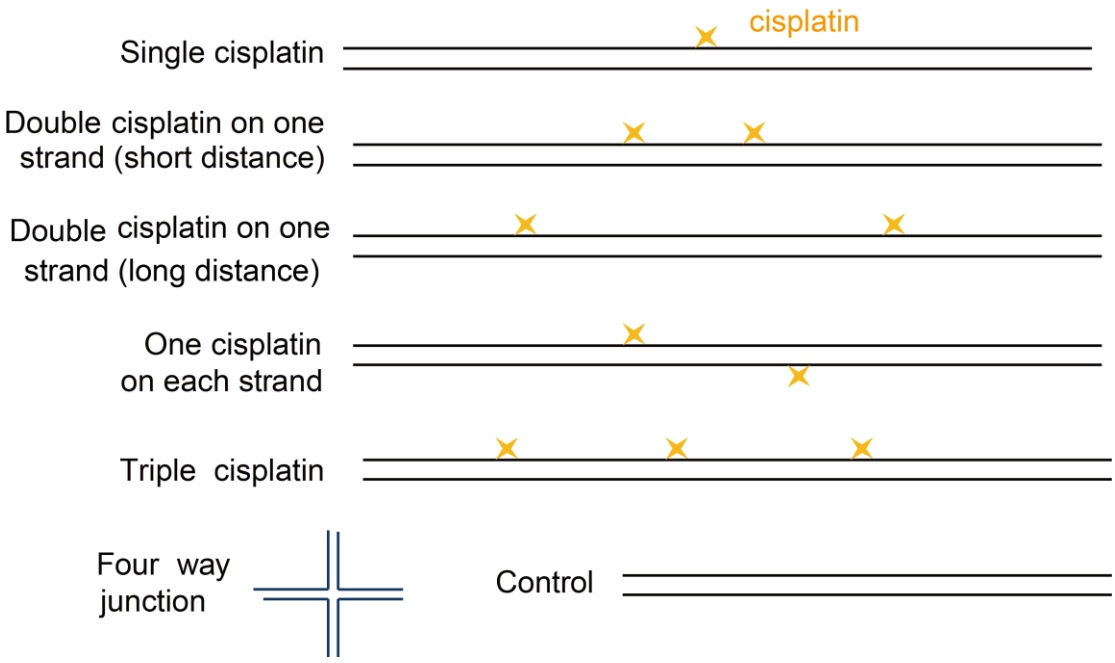
Supplementary Figures



Supplementary Figure 1. Preparation and characterization of 12 mer cisplatin-modified model DNA sequences.

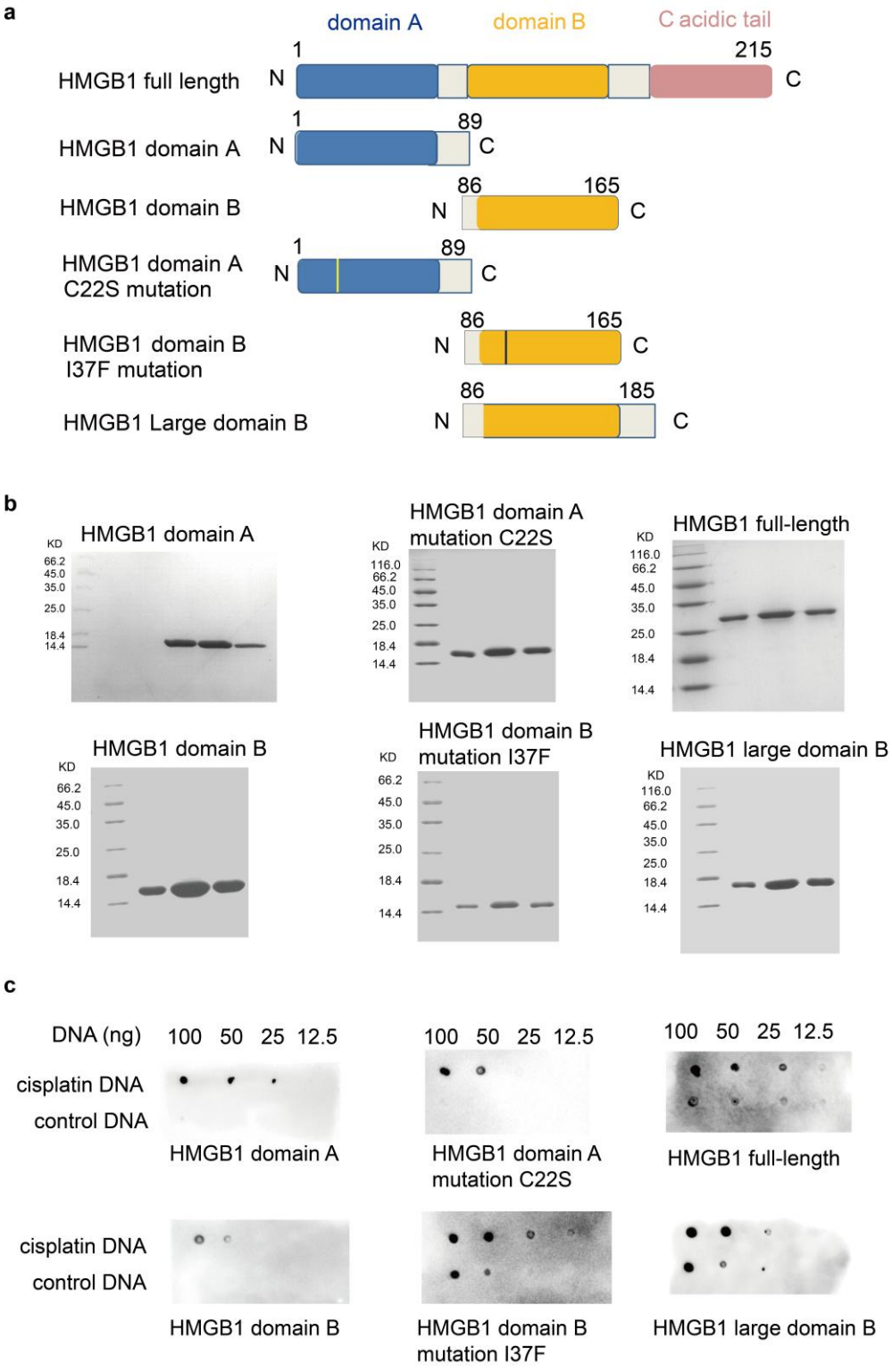
(**a**) DNA oligonucleotides reacted with cisplatin at different conditions were examined on 20% native-PAGE. The reaction condition was the best when the molar ratio of cisplatin:oligonucleotide was 2:1 at 37℃, taking both efficiency and specificity into account. (**b**) The molecular weights of cisplatin-modified oligonucleotides determined by

MALDI-TOF confirmed that each oligonucleotide was modified by one cisplatin. (**c**) HPLC analysis of purified cisplatin-modified DNA. Oligonucleotides were digested to nucleosides, and analyzed with a C-18 column of HPLC (260 nm). For *cis*-GG adducts, the peaks for dG disappear and the peaks for Pt[d(GpG)] are indicated by red arrow. For *cis*-AG adducts, the peaks for dG disappear and the peaks for dA drop significantly, which demonstrates the main product is an intrastrand crosslink between adjacent A and G. The peaks for Pt[d(ApG)] indicated by pink arrows were identified by the same elution time as the synthesized standard.



Supplementary Figure 2. Cisplatin-modified model DNA sequences.

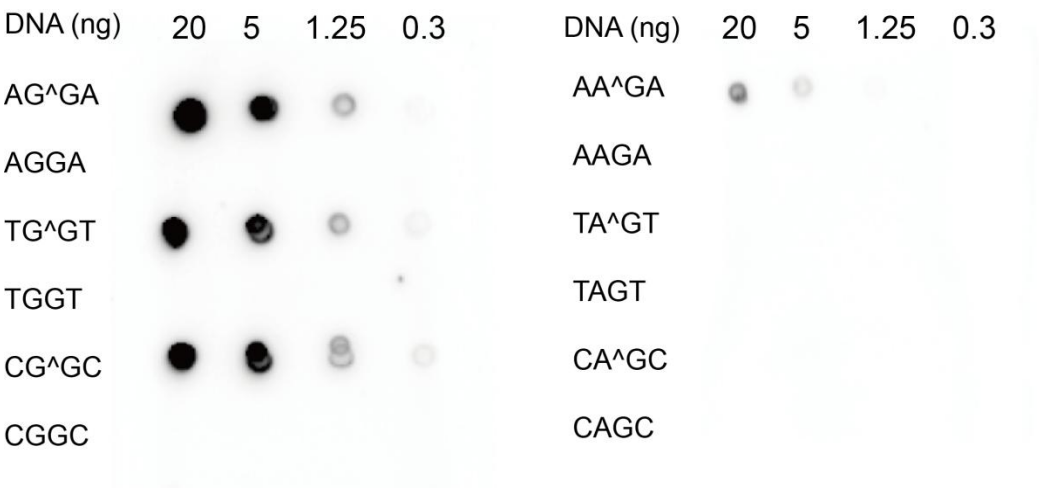
Schematics of cisplatin-modified model DNA sequences used in dot blotting, pull-down and high-throughput sequencing. Detailed sequence information can be found in **Supplementary Table**.



Supplementary Figure 3. Expression and purification of different HMGB1 constructs and analysis of their specificity to cisplatin-modified DNA.

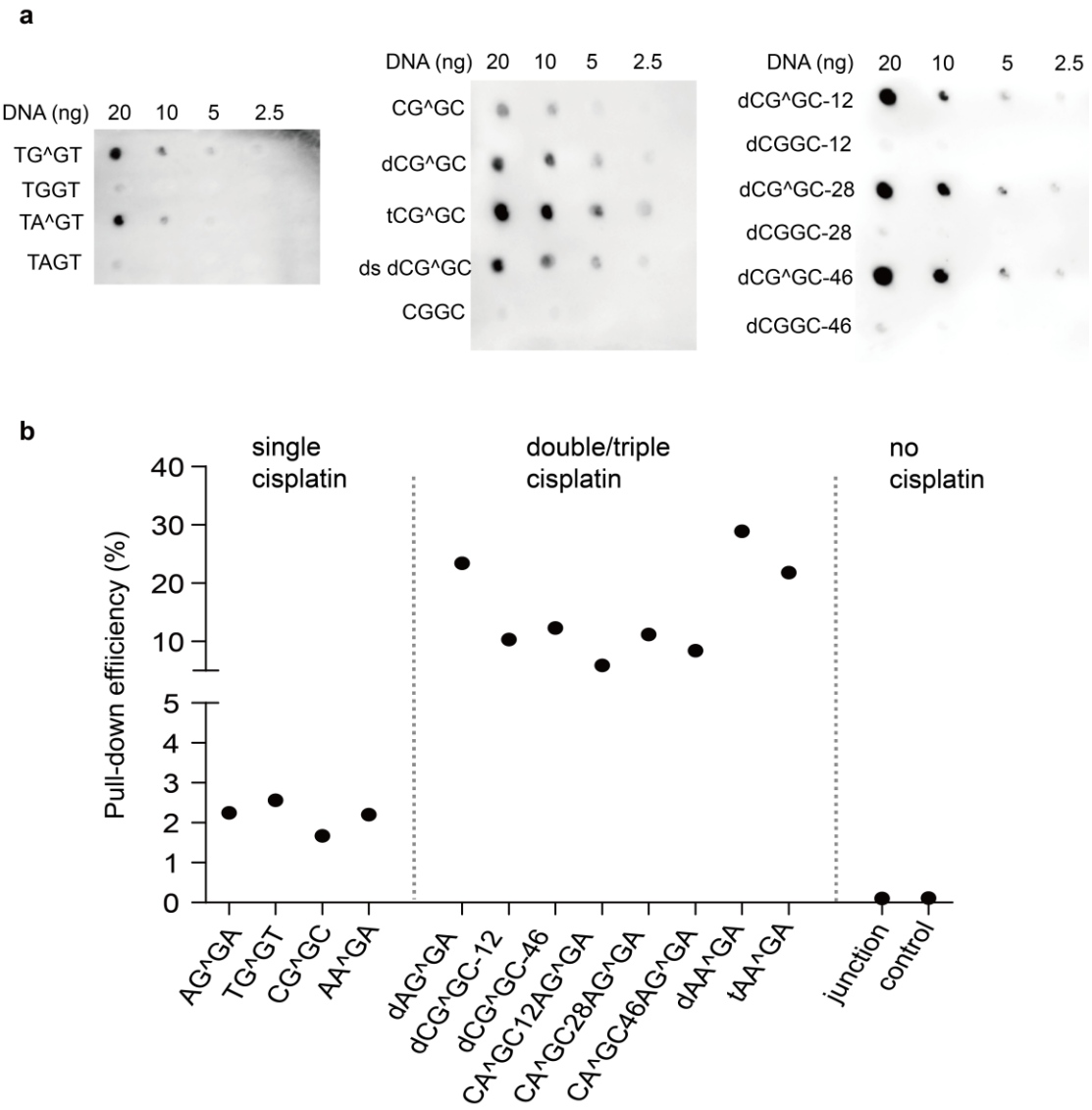
(**a**) HMGB1 contains two tandem DNA binding domains: domain A and domain B. The

gray boxes represent loop regions. (**b**) SDS-PAGE analysis of HMGB1 proteins after purification. (**c**) Specificity and Sensitivity of HMGB1 constructs to cisplatin-modified DNA, tested using dot blotting. DNA with single TG^GT (G^G denoting cisplatin crosslinking two guanosines) was used as cisplatin DNA; TGGT with same sequence was used as control DNA.



Supplementary Figure 4. The performance of Anti-Cisplatin modified DNA antibody in dot blot assays.

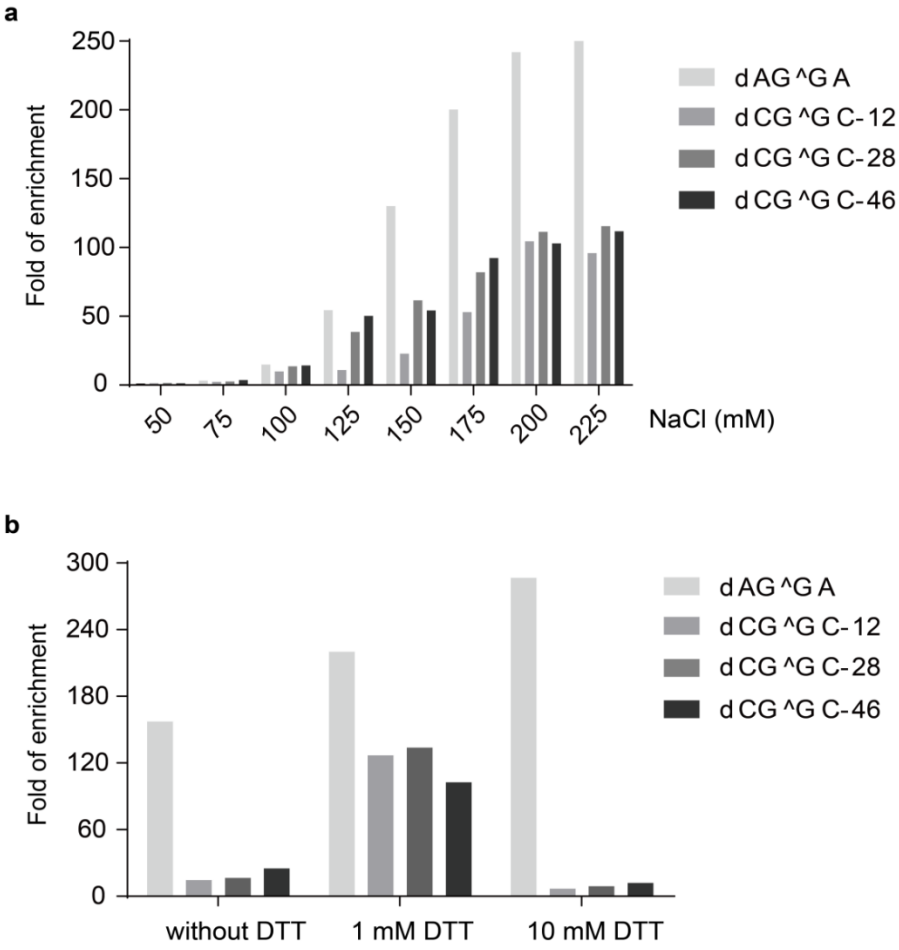
The commercial Anti-Cisplatin modified DNA antibody (abcam; ab103261) has good specificity to *cis*-GG adducts; however, the antibody has little immunoreactivity to *cis*-AG adducts. Model DNA sequences with cisplatin crosslinking or the same sequences without cisplatin modifications were spotted on the membrane. The antibody was 1:1000 diluted to detect cisplatin modification. Experiment details can be found in “Dot blotting” of Experimental Section.



Supplementary Figure 5. Preferential binding of HMGB1 domain A to model DNA sequences with different cisplatin-DNA adducts.

(**a**) Both G^G and A^G can be selectively recognized by HMGB1 domain A (upper left). Double and triple cisplatin modifications increased the binding capacities of HMGB1 domain A, no matter they were on the same strand or on different strands (bottom left). (ds dCG^GC means one CG^GC on each strand; G^G denoting cisplatin crosslinking two guanosines). The distance between two cisplatin had little effect on HMGB1 domain A binding ability (right). (**b**) Pull-down efficiency of cisplatin-modified model DNA sequences

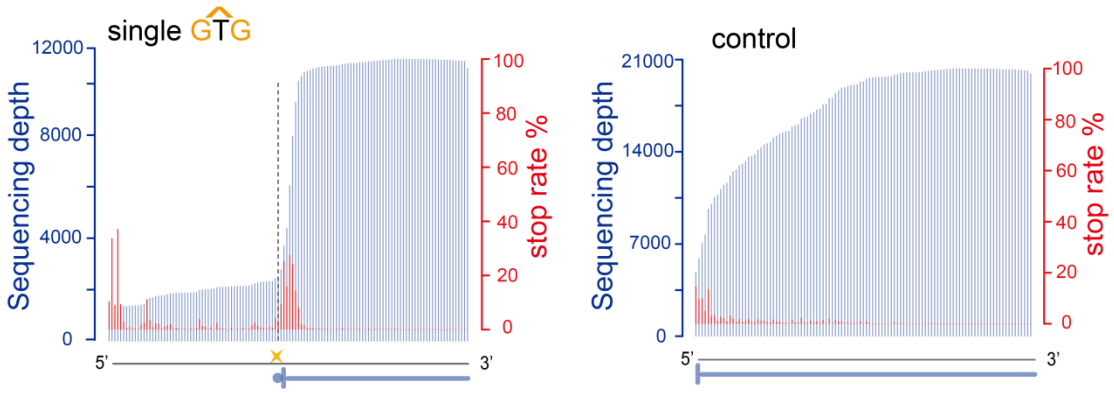
by HMGB1 domain A. For the sequence named “dCG^GC-12”, “d” stands for “double” cis-GG adducts, and “12” means a 12-bp distance between two cis-GG adducts. Similarly, in other sequences, “t” stands for “triple” and the digit represents the distance between two nearby cisplatin adducts; “junction” represents a sequence with four-way junction structure, which was used as a control. The pull-down efficiency of model sequences with one single cisplatin is ~1-3% of using HMGB1 domain A; yet the recovery efficiency goes up to ~20% for those with double or triple cisplatin. The commercial Anti-Cisplatin modified DNA antibody (abcam; ab103261) has an efficiency of ~0.1-0.3% for model sequences with one cisplatin and ~0.2%-1% for those with double or triple cisplatin.



Supplementary Figure 6. Optimization of cisplatin-modified DNA pull-down conditions.

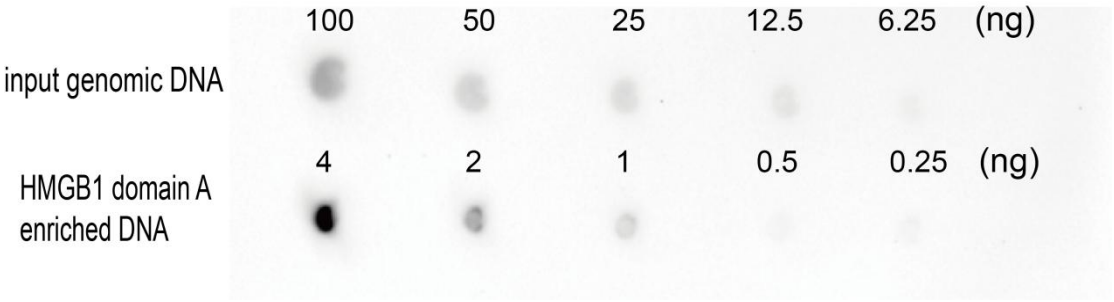
(**a**) The effect of NaCl concentration on the enrichment of cisplatin-modified DNA. The "enrichment" here represents the fold change in cisplatin-modified DNA/unmodified DNA relative to the input sample; it is calculated by comparing pull-down sample to input sample using the results from real-time PCR assay. The fold of enrichment increased when NaCl concentration was increased from 50 mM to 200 mM. No significant improvement occurred when NaCl concentration increased from 200 mM to 225 mM. (**b**) The effect of DTT concentration on the enrichment of cisplatin-modified DNA. Compared to DTT-free conditions, 1 mM DTT notably increased the enrichments of all cisplatin

modified DNA. Further increasing DTT concentration to 10 mM affected the overall enrichment of DNA with different cisplatin crosslinking adducts. Experiment details can be found in “Cisplatin-seq” and “Real-time PCR for calculating enrichments of cisplatin model sequences” of Experimental Section.



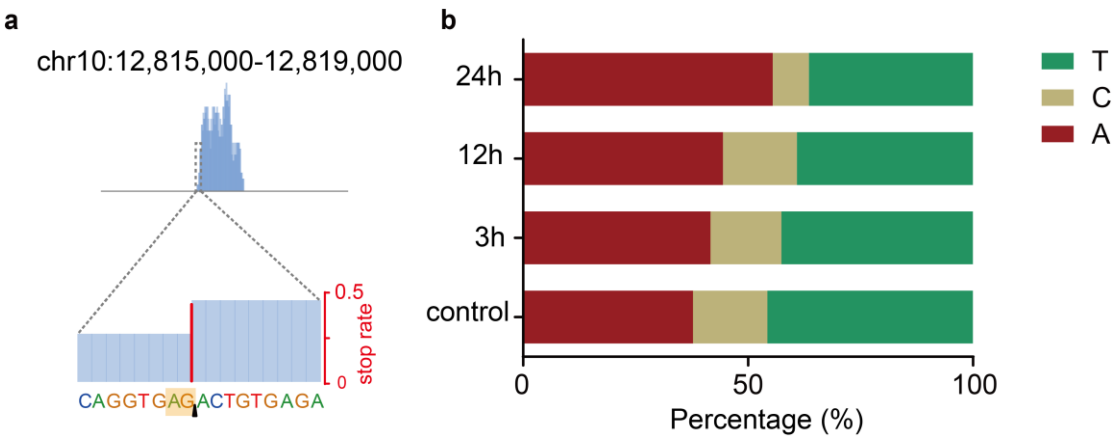
Supplementary Figure 7. Primer extension stops at the cisplatin crosslinking site

The stop site during primer extension is used to detect cisplatin crosslinking site in a model sequence. For *cis*-GTG 1,3-intrastrand adducts, sites of high stop rates occur ~2-8nt 3' to the *cis*-GTG adduct; hence only near base-resolution could be achieved. The blue vertical lines are sequencing depths while the red lines are calculated stop rates (stop rate is defined in the "Experimental Section"). The dashed line represents the cisplatin crosslinking site.



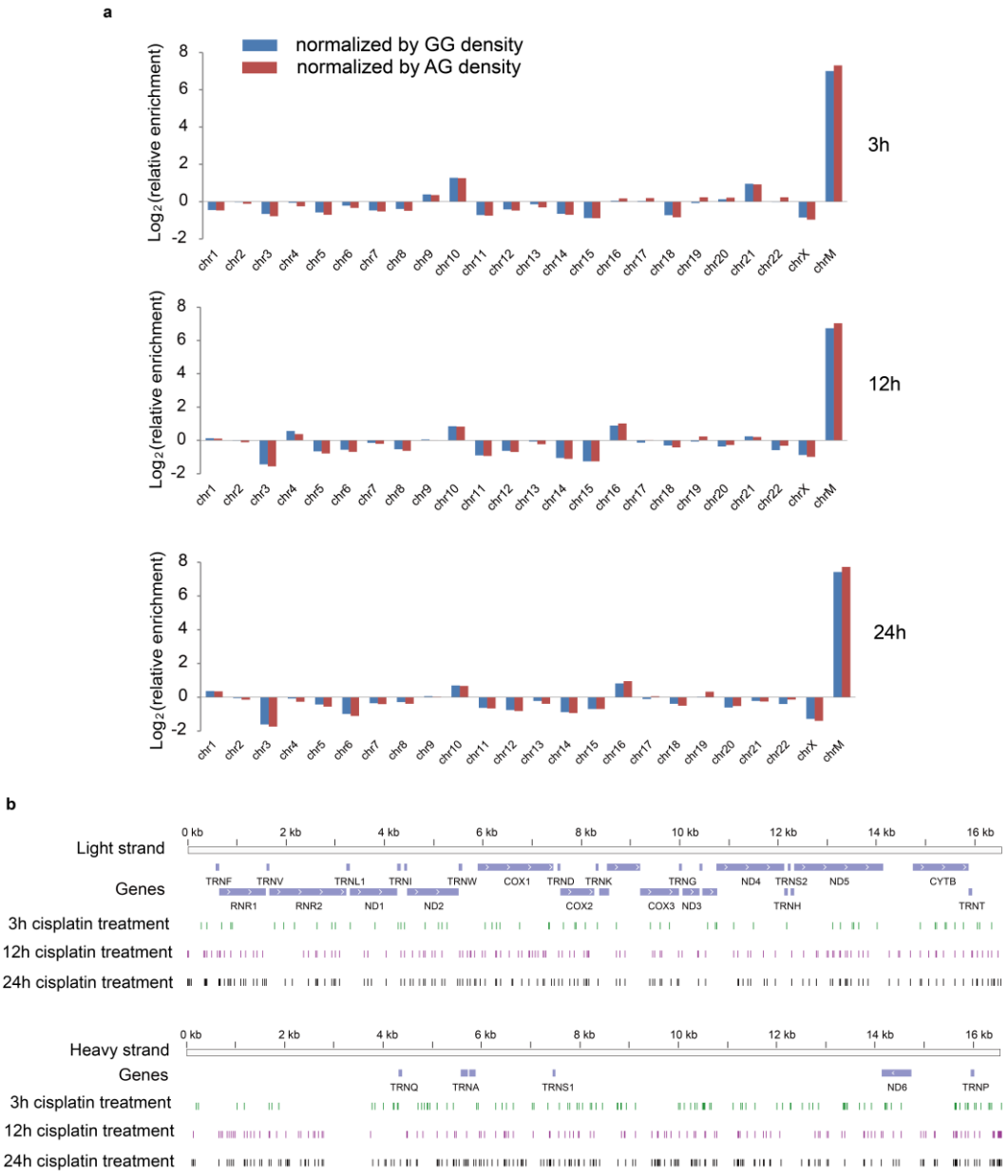
Supplementary Figure 8. HMGB1 domain A effectively enriched cisplatin-modified genomic DNA in a pull-down experiment.

Genomic DNA from cells treated with cisplatin was pulled-down by HMGB1 domain A coupled beads. In the membrane, the upper dots are DNA before pull-down. The lower dots are eluted DNA after HMGB1 domain A pull-down experiment. Anti-Cisplatin modified DNA antibody (abcam; ab103261) was used at 1:1000 dilution in dot blot assay to estimate the content of cisplatin-modified DNA before and after HMGB1 domain A pull-down. Experiment details can be found in “Cisplatin-seq” and “Dot blotting” of Experimental Section.



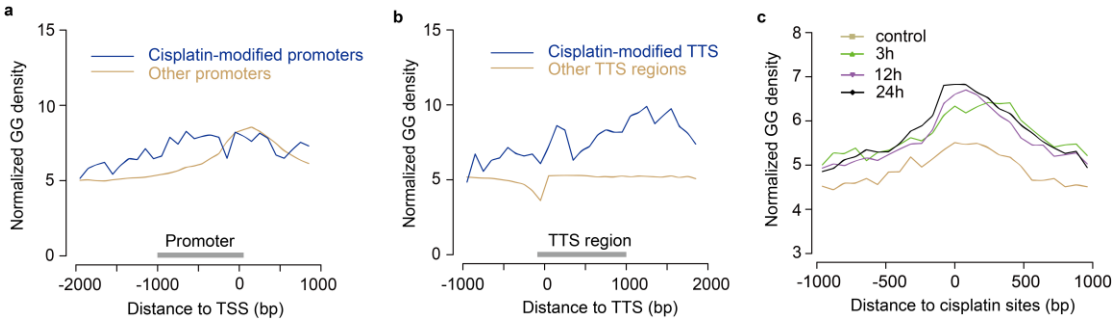
Supplementary Figure 9. Cisplatin-seq is capable of detecting *cis*-AG crosslinking adducts.

(**a**) IGV views of a representative *cis*-AG cisplatin site identified in the genome. Blue lines are sequencing depths. In the zoom-in view, the identified *cis*-AG site is shown. (**b**) Nucleotide frequency of position -1 when position 0 is a guanine ("Position 0" corresponds to the high stop-rate sites). Cases with guanines in both Position 0 and Position -1 (corresponds to *cis*-GG adducts) were excluded for analysis. The frequency of A is elevated with cisplatin treatment, indicating *cis*-AG adducts were enriched. There are 183, 394, 345 *cis*-AG cisplatin sites identified for cells with 3h, 12h and 24h cisplatin treatment, respectively (all cisplatin sites are available in Supplementary-dataset).



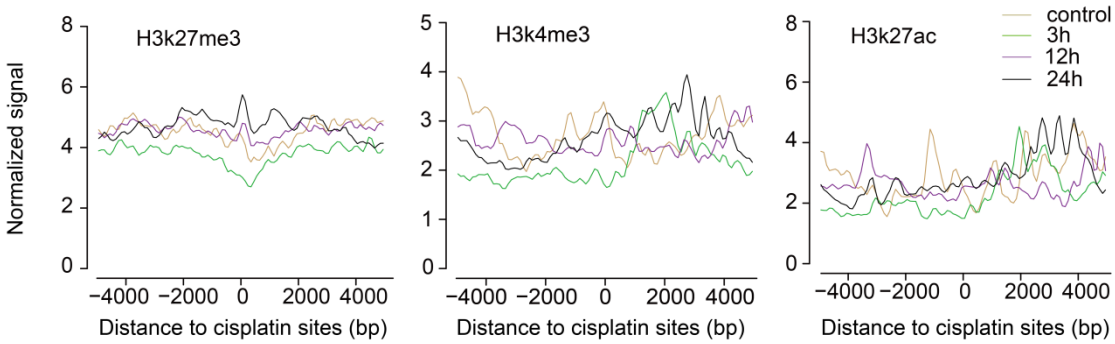
Supplementary Figure 10. Relative enrichment of cisplatin sites on each chromosome and dynamics of cisplatin crosslinking in mitochondrial DNA.

(**a**) Relative enrichment of cisplatin sites on each chromosome. Enrichments were normalized first by the sequencing coverage of input sample and further by the GG dinucleotide density or AG dinucleotide density of each chromosome. (**b**) Locations of cisplatin crosslinking sites on light and heavy strand of mitochondrial DNA after 3h, 12h or 24h cisplatin treatment, respectively.



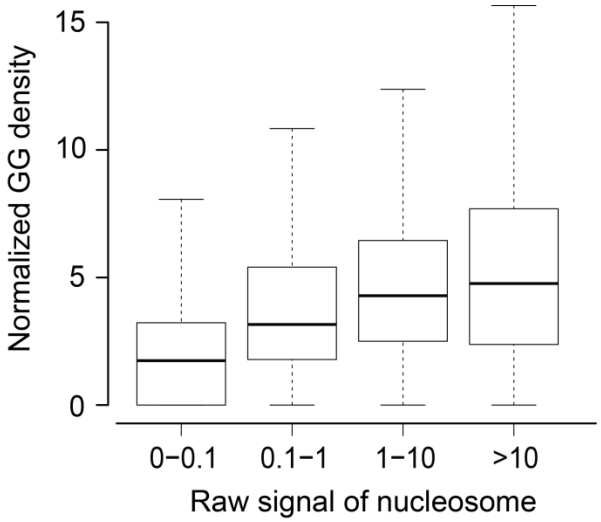
Supplementary Figure 11. The distribution of cisplatin adducts is positively affected by GG dinucleotide density.

(**a**) The normalized GG dinucleotide density of cisplatin-modified promoters (the blue line) is higher than that of other promoters (the orange line). A promoter is defined as 1,000 bp upstream and 100 bp downstream of the transcription start site. (**b**) The normalized GG dinucleotide density of cisplatin-modified transcription termination sites (TTS) is higher than that of other TTS. TTS regions are defined as 100 bp upstream and 1,000 bp downstream of the translation termination sites. (**c**) GG dinucleotide density (the average occurrence of GG dinucleotide in 100 bp sequence) across the identified cisplatin sites.



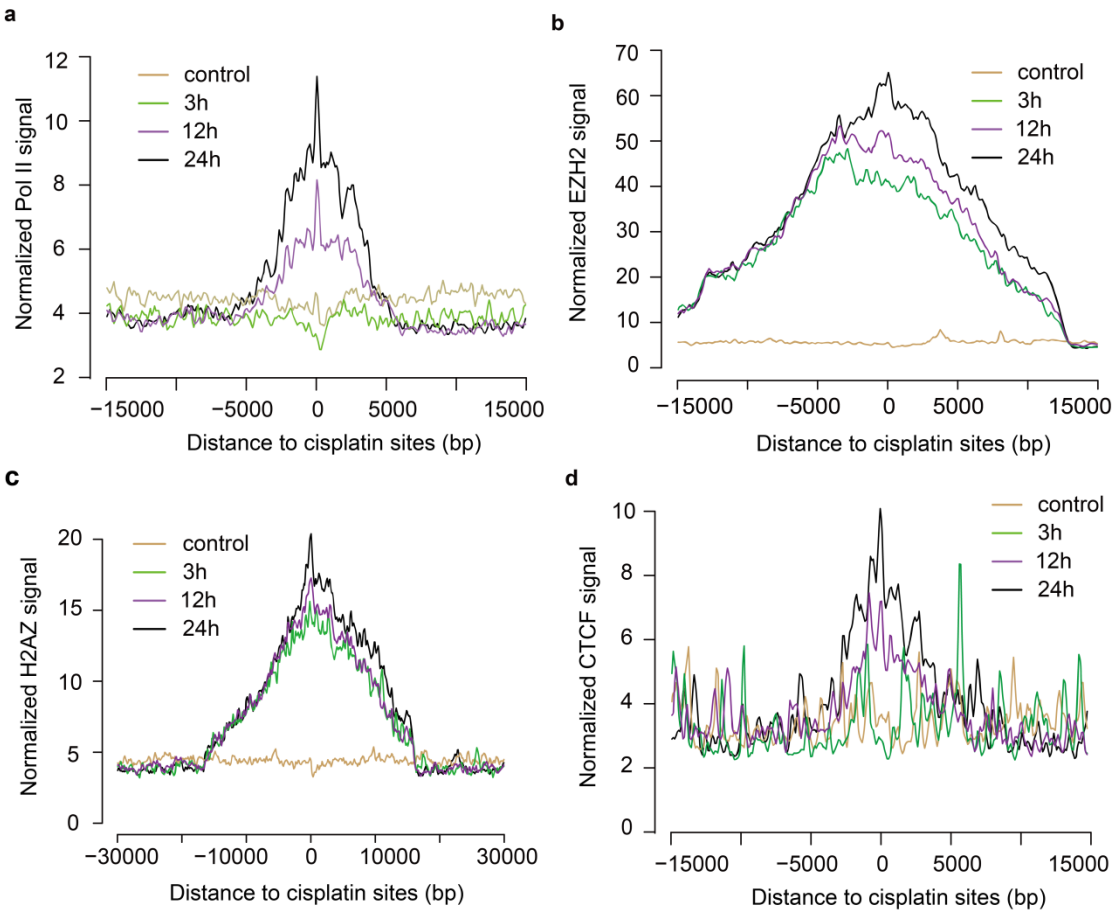
Supplementary Figure 12. No significant correlation was observed between the distribution of cisplatin crosslinking sites and several histone modification signals.

Normalized sequencing signals of H3K27me3, H3K4me3 and H3K27ac ChIP-seq across the identified cisplatin crosslinking sites.



Supplementary Figure 13. GG dinucleotide density is higher in the regions with stronger nucleosome signals.

Regions with stronger nucleosome signals tend to have higher GG dinucleotide density.



Supplementary Figure 14. Signals of Pol II, EZH2, H2AZ and CTCF are enriched near cisplatin crosslinking sites.

Normalized ChIP-seq sequencing signals of Pol II (**a**), EZH2 (**b**), H2AZ (**c**) and CTCF (**d**) across the identified cisplatin crosslinking sites.

Supplementary Table

Adaptors for cisplatin-seq library preparation

Name	Sequences
3'block adaptor up	5'P-GATCGGAAGAGCACACGTCTGAACTCCAGTC
3'block adaptor down	5'-GCTCTTCCGATCT-3'P
Bio-elong primer	5'biotin-GACTGGAGTTCAGACGTGT
ssDNA Ligase	5'P-AGATCGGAAGAGCGTCGTGT-3NH2

12 mer cisplatin oligonucleotides (^ indicates cisplatin crosslinking site)

TCTAG^GATCTCT
TCTTG^GTTCTCT
TCTCG^GCTCTCT
TCTAA^GATCTCT
TCTTA^GTTCTCT
TCTCA^GCTCTCT
TCTCGTGCCTCT
TCTTGTGTCTCT

Cisplatin model sequences for Dot Blot and IP-qPCR enrichment tests (^ indicates cisplatin crosslinking site)

Name	Sequences (^ indicates cisplatin crosslinking site)
AG^GA	CGATGGGCAACAATCTACCTAGTAACTGACCAGATCTCTAG^GAT CTCTTGAGGCTACTGAGTTAGAAAGGACTTGTGCACAGCA
TG^GT	AATGTGGATGCCGCAGTTGCAGTAACTGACCAGATCTCTTG^GTT CTCTTGAGGCCTACTGAGTTACAACCGGGACATCACGGAT
CG^GC	CTACGCAAACTGGCTGTCAAAGTAACTGACCAGATCTCTCG^GCT CTCTTGAGGCTACTGAGTTATCATGGACGCTACCTCACAG
AA^GA	TATAACCCGACGACTCGACCAGTAACTGACCAGATCTCTAA^GAT CTCTTGAGGCTACTGAGTTAAGTGCAACATTGGGGCTAAC
TA^GT	AGGCCAACATACATGCCTTCAGTAACTGACCAGATCTCTTA^GTTC TCTTGAGGCTACTGAGTTAGAATGGCAGAGTCAAGGAGC
CA^GC	TGGGAGATGGTTGCCAGGGTAGTAACTGACCAGATCTCTCA^GC TCTCTTGAGGCTACTGAGTAGGCCATTGATGCATCTTTCCGAC
CGTGC	CATGAGTGCCCTCAGCAGTAAGTAACTGACCAGATCTCTCGTGC CTCTTGAGGCTACTGAGTTATCCAACCTTTAGGAGCCATG
TGTGT	ATTCACCCCCACTGAGACTGAGTAACTGACCAGATCTCTTGTGTC TCTTGAGGCTACTGAGTCATGCACGCAGAAAGAAATAGC
Double AG^GA	CAGTTGATGCGAGAGATGGACTCTAG^GATCTCTCTTAGTAGGGG

	ATTCTTAAGGCATTTTGAGAATTTCTAG^GATCTCTCGTAACGGAG CAAAAACTCGC
Double 12 CG^GC	CTGCGTTTCCAAAGGAAGAGATACTCTGAGGCTCTCG^GCTCTC TTCTCG^GCTCTCTCAGATCATAGTATAGCAGCAAGAGGATCCGA C
Double 28 CG^GC	TTATCTGGCGGTGGAGATGCTGTATCTCG^GCTCTCTGACCACA GTAGTAGCTTCTCG^GCTCTCTTGATGTAATGCTTGCGTGGCAAA C
Double 46 CG^GC	GAAAGATTGAAGAGCTTCTGCTCTCG^GCTCTCTTGGTTGACACC TGTTCTTATTAAAGTAATGGTACTCTCG^GCTCTCTGAGTGTTTCT TGGTTTGAGGT
Double AA^GA	ACCTGGGGTTTGTAACATGCAGTCTAA^GATCTCTGCTATACAATA TGACAAAAGTCTTCGTAACAGGTTCTAA^GATCTCTCTCAAACACA TTGCCAAACGTC
Double CA^GC	GCACCACTTTGGACAAAGACACTCTCA^GCTCTCTTTATTTTAGC AATTGCTGTGTGAGGCACTGACCTTCTCA^GCTCTCTGCTGAAAT CCTCGATGAACAGG
CA^GC12AG^G A	GCTAGGTGAGGATTTGGGACAACAGGTGGTATTCTCA^GCTCTCT TCTAG^GATCTCTCTGTCACTATGGATTCATCATTGAGGCAAATT
CA^GC28AG^G A	CCTCCGTTGTGATGTAACTGGATATCTCA^GCTCTCTTTCGTAAC AGGTGCCATCTAG^GATCTCTCCGATGATGTTACTTTCCACCAAT
CA^GC46AG^G A	TTCACTCGATCTTCGCCATTGTTCTCA^GCTCTCTCTCATGAGGAT CAATACATGTATAAACTTACCATTCTAG^GATCTCTCCTAAAAGCA TGGTGGATGTGA
Double strand 20 AG^GA	ATAAGGCCATTGAAAGGCAAACTCTTGTCCGTTCTAG^GATCTCTA AAGATGTAGAGATCCTAGAGCTGAATATCGGTGCCTTCATGGAG
Double strand 35 AG^GA	CCCTGTATGGGCATCTTTGGAACTTCTAG^GATCTCTGTCGTTCAT CAACTATTAGATCAAGAGATCCTAGAAAGAGAGCATATGTGAGC AAATCA
Double strand 20 CG^GC	CAACAACTGGACTCTCCTCACCTTTTAGATAATCTCG^GCTCTCT ATTTCGATAGAGAGCCGAGAAATAGATAATGGGGAAGTGGAAGG A
Double strand 35 CG^GC	CACACCCCTTTAACAATCTTCACATCTCG^GCTCTCTTTCAGATGT TGAAATTGTCTTACAGAGAGCCGAGAATGGAGAAGTAGAAGGAT CATCGC
Triple CG^GC	GCGTGGTGTACTACTATTGAATCTCG^GCTCTCTAGGATTTTACAT AAAGGACATTTGGTCACTTTGATCTCG^GCTCTCTTTGGGGTCTG TCTTATGGAAGTTTCTAGGTGCTGTCTCG^GCTCTCTAATAGTCC ACGCCATGTTAG
Triple CA^GC	CCAGTTTGGAATGCCAGGACTCTCA^GCTCTCTAATTCTTAATGT GGAGGTCTGTTAGTGATTAACATCTCA^GCTCTCTCGGGAGGCT GAAGCAGGAAGATTGCTGTAAGTTG TCTCA^GCTCTCTAGCCCGCTAGGGTTTCTCC
Triple AA^GA	CAAATGGGAAGGGAAGAGAAGCTCTAA^GATCTCTTATTTGAAGA

	ATACGTTACTGATAAATTGAACTTTCTAA^GATCTCTCAACCAGCT GTTAATGGTACTAGAAATGTCTTGATCTAA^GATCTCTACATCCTT CTTGCCAGGTCGAT
Control	GCACACAGGTGTTCACTTTGCTAAGTAACTGACCAGATGTCATG TGCTCTTGAGGACTGAGCTGGTACGCAAATCGCAGCATTG
Four-way junction	
Four way seq1	AGAATTTGCTCATGATTTTCAAGCGAATTCAGCACGAGTCCTAAC GCCAGATCTGAGCTAAGGGTGATGGAAAAACA
Four way seq2	TGTTTTTCCATCACCCTTAGCTCAGATCTGGCGTTAGGTGA TACCGATGCATC
Four way seq3	GATGCATCGGTATCAGGCTTACGACTAGTG
Four way seq4	CACTAGTCGTAAGCCACTCGTGCTGAATTCGCTTGAAAATCATG AGCAAATTCT

Long cisplatin-modified DNA model sequences as spike_in for Next-generation sequencing (^ indicates cisplatin crosslinking site)

NNNNNNCGATGGGCAACAATCTACCTCTCTCTTTACGCGTCACACACGTATAAACTAG ATCTAG^GATCTCTTGACTGAGTTAGACTTGTGCACAGCACAAATATTCGGAAAGGAC TTGTGCACAGCAAGA
NNNNNNCATGAGTGCCCTCAGCAGTAAGTAACTGACCAGATCCACAGTTCTCATGGG AATCTCGTGCCTCTATGCTAAACCGTTTTGAGGCTACTGAGTTATCCAACCTTTAGGA GCCATGATGTGAATA
NNNNNNTGCTGTAAGTTGTTTCAACAGTTGATGCGAGAGATGGACTCTAG^GATCTC TCTTAGTAGGGGATTCTTAAGGCATTTTGAGAATTTCTAG^GATCTCTCGTAACGGAG CAAAAACTCGC
NNNNNNCCAACCAGCTGTTAATGCACCACTTTGGACAAAGACACCATATCTCA^GCT CTCTAGCAACTGCTGTGGCATCTCA^GCTCTCTTAGCTAAATTGCTAGCTGAAATCCT CGATGAACAGG
Double strand double CG^C (20bp between): Forward: NNNNNNTGTCAAGCAACAACTGGACTCTCCTCACCTTTTAGATAATCTCG^GCTCTC TATTTCGATAGAGAGCCGAGAAATAGATAATGGGGAAGTGGAAGGAAAGGTGCTGG TTAT Reverse: ATAACCAGCACCTTTCCTTCCACTTCCCCATTATCTATTTCTCG^GCTCTCTATCGAAAT AGAGAGCCGAGATTATCTAAAAGGTGAGGAGAGTCCAGTTGTTGCTTGACANNNNNN
NNNNNNTGGCTCCATCTTCCCCTGTTCTCTAG^GATCTCTCAATCTCGTCCAGCCATT GACCATCGTCTCTAG^GATCTCTGAAGCCGACGACACTACCGTGCCGCCTTTCTAG^ GATCTCTGGATTAGGCGTCAACACCGAG
NNNNNNCCACAGAGGACTGAGGTGGTTCTCGTGCCTCTACTGCACACCCTCTA ATACTGTTACTCGTCTCGTGCCTCTGACGCGGCTACGACAAGAGTTGGCCTCATCT CGTGCCTCTCTCAAACGGGGATTCGATTA

Primers for real-time PCR

Name	Sequences
AGGA forward	CGATGGGCAACAATCTACCT
AGGA reverse	TGCTGTGCACAAGTCCTTTC
TGGT forward	AATGTGGATGCCGCAGTTG
TGGT reverse	ATCCGTGATGTCCCGGTTG
CGGC forward	CTACGCAAACTGGCTGTCAA
CGGC reverse	CTGTGAGGTAGCGTCCATGA
AAGA forward	TATAACCCGACGACTCGACC
AAGA reverse	GTTAGCCCCAATGTTGCACT
TAGT forward	AGGCCAACATACATGCCTTC
TAGT reverse	GCTCCTTGACTCTGCCATTC
CAGC forward	TGGGAGATGGTTGCCAGGGT
CAGC reverse	GTCGGAAAGATGCATCAATGGCCT
CGTGC forward	CATGAGTGCCCTCAGCAGTA
CGTGC reverse	CATGGCTCCTAAAGGTTGGA
TGTGT forward	ATTCACCCCCACTGAGACTG
TGTGT reverse	GCTATTTCTTTCTGCGTGCAT
Double AGGA forward	CAGTTGATGCGAGAGATGGA
Double AGGA reverse	GCGAGTTTTTGCTCCGTTAC
Double 12 CGGC forward	CTGCGTTTCCAAAGGAAGAG
Double 12 CGGC reverse	GTCGGATCCTCTTGCTGCTA
Double28 CGGC forward	TTATCTGGCGGTGGAGATG
Double28 CGGC reverse	GTTTGCCACGCAAGCATTA
Double 46 CGGC forward	GAAAGATTGAAGAGCTTCTGC
Double 46 CGGC reverse	ACCTCAAACCAAGAAACACTC
Double AAGA forward	ACCTGGGGTTTGTAACATGCAG
Double AAGA reverse	GACGTTTGGCAATGTGTTTGAG
Double CAGC forward	GCACCACTTTGGACAAAGACAC
Double CAGC reverse	CCTGTTCATCGAGGATTTCAGC
CAGC12AGGA forward	GCTAGGTGAGGATTTGGGACAA
CAGC12AGGA reverse	AATTTGCCTCAATGATGAATCCA
CAGC28AGGA forward	CCTCCGTTGTGATGTAACTGG
CAGC28AGGA reverse	ATTGGTGGAAAGTAACATCATCG
CAGC46AGGA forward	TTCACTCGATCTTCGCCATTGT
CAGC46AGGA reverse	TCACATCCACCATGCTTTTAGG
Double strand 20 AGGA forward	ATAAGGCCATTGAAAGGCAAAC
Double strand 20 AGGA reverse	CTCCATGAAGGCACCGATATTC
Double strand 35 AGGA forward	CCCTGTATGGGCATCTTTGG
Double strand 35 AGGA reverse	TGATTTGCTCACATATGCTCTCTT

Double strand 20 CGGC forward	CAACAACTGGACTCTCCTCACCTT
Double strand 20 CGGC reverse	TCCTTCCACTTCCCCATTATCTATT
Double strand 35 CGGC forward	CACACCCCTTTAACAATCTTCACA
Double strand 35 CGGC reverse	GCGATGATCCTTCTACTTCTCCAT
Triple CGGC forward	GCGTGGTGTACTACTATTGAA
Triple CGGC reverse	CTAACATGGCGTGGACTATT
Triple CAGC forward	CCAGTTTGGAATGCCAGGAC
Triple CAGC reverse	GGAGAAACCCTAGCGGGCT
Triple AAGA forward	CAAATGGGAAGGGAAGAGAAGC
Triple AAGA reverse	ATCGACCTGGCAAGAAGGATGT
Reference forward	GCACACAGGTGTTCACTTTGC
Reference reverse	CAATGCTGCGATTTGCGTACC
Four war junction forward	AGAATTTGCTCATGATTTTCAAGC
Four war junction reverse	TGTTTTTCCATCACCCTTAGCTC

Experimental Section

Synthesis and characterization of 12-mer site-specifically modified cisplatin oligonucleotides.

50 μL cisplatin (2 mM, Sigma-Aldrich) and 50 μL $AgNO_3$ (4 mM) were mixed in the dark at 37°C for 6h and the AgCl was then removed by centrifugation. A 12 mer DNA oligonucleotide was added to the supernatant in a 1:2 (DNA: Pt) molar ratio and incubated in the dark at 37°C for 12h. The platinated oligonucleotides were purified by 20% native-PAGE. The molecular weight of purified platinated oligonucleotides were characterized by matrix-assisted laser desorption/ionization–time of flight (MALDI-TOF) (**supplementary Fig. 1b**). The cisplatin crosslinking sites were validated by HPLC (**supplementary Fig.1c**) following DNA digestion into nucleosides. 1 nmole oligonucleotide was mixed with 1 μL nuclease P1 (Sigma, N8630) in 100 μL 10 mM NH_4Ac at 42°C for 6h and then 11 μL 1 M NH_4HCO_3 and 5μL alkaline phosphatase (Sigma, P4252) were added and incubated at 37°C for 12h. The HPLC programs for separating and analyzing the nucleosides follow the procedure described previously[1]. The detection wavelength was set at 260 nm.

Preparation of site-specifically modified, long duplex DNA as cisplatin-containing model sequences.

Long duplex DNA were prepared through ligation of short duplexes with sticky overhangs[2]. In brief, the ligation-site oligonucleotides were phosphorylated with T4

polynucleotide kinase (NEB) and then annealed with the corresponding complementary strands. Annealed duplexes with sticky overhangs were ligated with T4 DNA ligase (NEB) for 4 h at 16 °C, followed by purification with 10% PAGE. The Pt: DNA ratios were determined by ICP-MS.

Expression and purification of various truncated/mutated HMGB1 proteins

The rat HMGB1 cDNA was synthesized by Shanghai Sangon Biotech and cloned into pET-28a Plasmid (Novagen). Truncated HMGB1 proteins were PCR amplified from wild-type HMGB1 cDNA and point mutations were achieved by Fast Mutagenesis System (TransGen Biotech), and the modified plasmids were transformed into BL21 (DE3) cells. The transformed cells were grown in LB medium with 50 μg/mL kanamycin at 37 °C and 200 rpm. When OD600 was 0.5-1.0, IPTG was added at a final concentration of 0.5 mM to induce gene expression. The expressed HMGB1 proteins contain a histag at the N terminal. Then cells were incubated at 30 °C for 4 h and harvested by centrifugation and resuspended in 40 mL cold lysis buffer (10 mM Tris, 1 mM DTT, 300 mM NaC1, and 1 mM PMSF, pH 7.4). All subsequent steps were conducted either on ice or at 4°C. Cells were lysed by sonication and cleared by centrifugation for 30 min at 15,000g. The supernatant was purified by HiLoad HisTrap (Ni-NTA) column (GE Healthcare), Superdex 75 PG size-exclusion column (GE Healthcare) using an ÄKTA pure system (GE Healthcare).

Dot blotting

DNA were spotted on the membrane (Amersham Hybond-N+, GE) and air dry for 5 min before UV-crosslink (2×auto-crosslink, 1800 UV Stratalinker, STRATAGENE). After crosslinking, membranes were blocked in 5% skim milk (dissolved in PBST) for 1-2 h at room temperature. For HMGB1 construct proteins, the membranes were incubated in 5% milk containing HMGB1 domain A or other truncated/mutated proteins for 1h at room temperature; for Anti-Cisplatin modified DNA antibody (abcam; ab103261), the membrane was incubated with Anti-Cisplatin antibody (1:1000 dilution by 5% milk) for 1h at room temperature. After 3 times wash with PBST, membranes were incubated with HRP linked anti-histag antibody (CWBiotech) for 1h at room temperature. Signals were detected with ECL Plus Chemiluminescent reagent (Thermo Pierce).

Coupling HMGB1 domain A to magnetic beads

Purified protein was covalently fixed to magnetic beads[3]. 7 mg M-270 epoxy beads (Invitrogen) was suspended in 1 mL PBS, rotated in room temperature for 10 min and washed again. The beads were then mixed with 300 μg purified HMGB1 domain A protein in PBS with 1 M $(NH_4)_2SO_4$, 1 mM DTT and rotated for 48 h at 4°C. After coupling HMGB1 domain A to the beads, the beads were washed 30 min with pull-down buffer (10 mM Tris, 200 mM NaCl, 1 mM DTT, 0.15% NP-40) to block active groups on the beads. Beads were resuspended in 300 μL pull-down buffer and separated to 20 tubes, which can be stored in -20°C for 1month.

Cisplatin treatment and genomic DNA extraction

Hela cells were maintained in DMEM medium (Gibco) supplemented with 10% FBS and 1% penicillin & streptomycin. Mycoplasma contamination tests were performed routinely using the GMyc-PCR Mycoplasma Test Kit from YEASEN (cat. #40601). 50 μM cisplatin were used to treat Hela cells for 3 h, 12 h or 24 h. Genomic DNA was extracted by DNAzol (Invitrogen).

Cisplatin-seq

Ultrasonic fragmented genomic DNA (50 μg, 150-700 bp) with 1ng of each cisplatin-modified DNA model sequence was rotated with HMGB1 domain A coupled beads in pull-down buffer (10 mM Tris, 200 mM NaCl, 1 mM DTT, 0.15% NP-40) for 1h at 4°C. After 4 times wash, bound DNA was recovered by phenol: chloroform extraction followed by microspin-6 (Bio-rad) purification.

Purified DNA was used for End Repair/ dA –Tailing and adaptor ligation (New England Biolabs, NEBNext® Ultra™ DNA Library Prep Kit E7370) following the manufacturer's guidelines. The adaptor was self-designed (1^{st} 3'block adaptor) to ensure only the 3' end of DNA can be ligated with adaptor. After the adaptor ligation reaction, DNA was purified using 1 volumes of Ampure XP. Bio-elong primer (10 μM, 0.5 μL) was used for primer extension by NEBNext® High-Fidelity 2X PCR Master Mix. PCR machine program for primer extension is 98°C 40s, 54°C 30s, 72°C 5 min. Then 1 μL exonuclease I

was added to remove free primers in 37 °C for 1 h and heat inactivated by 80 °C 20 min. Extended DNA was purified by streptavidin C1 beads (Invitrogen) and the non-biotinylated templates were removed by 0.15 M NaOH. Biotinylated DNA was eluted at 95 °C for 3 min using deionized water. Subsequently, a single-stranded ligation adaptor (100 μM, 0.6 μL) was ligated by circLigase II ssDNA Ligase (EpiCenter) in 20 μL reaction with $MnCl_2$ and Betaine according to the instruction for 16 h at 60°C. After ligation, biotinylated DNA fragments were captured on streptavidin C1 beads (Invitrogen) and redundant ligation adaptor was removed by wash three times. Eluted DNA was subjected to PCR amplification. Sequencing was performed by Illumina Hiseq X Ten. The parameters in library preparation procedures were optimized. These parameters include primer extension time and cycles, beads washing and enzymes for the second adaptor ligation. The optimized procedure is as above.

Real-time PCR for calculating enrichments of cisplatin model sequences

100 pg each cisplatin-modification model sequence (**Supplementary table**) was spiked into 50 μg fragmented genomic DNA. 1% of the mixed DNA was separated as input. Cisplatin modified DNA enrichment was performed as procedures in Cisplatin-seq above. After DNA extraction and purification, SYBR® Premix Ex Taq™ II (Takara) was used to perform real-time PCR with LightCycler® 480 Real-Time PCR System (Roche). Three biological replicates were repeated to validate the results. The “enrichment” represents the fold change in cisplatin-modified DNA/unmodified DNA relative to the input sample;

calculated by comparing pull-down sample to input sample using real-time PCR assay.

Data processing and analysis

Raw reads were firstly sent for adapter and quality trimming using trim_galore, and reads shorter than 25 nt after trimming were excluded. For cisplatin model sequence analyses, processed reads were aligned to the model sequences using bowtie2. “Stop rate” for position *i* is calculated using the equation: $Stop^i = N^i_{stop}/(N^i_{stop} + N^i_{through})$, where N^i_{stop} is the count of reads whose 5’ most end mapped to position *i+1*, and $N^i_{through}$ is the count of reads covered at position *i*. For genome-wide cisplatin crosslinking sites identification, processed reads mapped to model sequence were discarded, and then mapped to human reference genome (hg19) using bowtie2, with parameters --end-to-end and --no-discordant. Duplicated reads were discarded using picard tools to remove artificial signals. Reads mapped to the forward strand and reverse strand were separated for the following analyses. Two algorithms were then adopted: 1) stop rate for each nucleotide in the genome is calculated following the equation mentioned above; 2) peaks were called using model-based analysis of ChIP-Seq (MACS2). Cisplatin crosslinking sites were then identified when all the following criteria are fulfilled: 1) stop rate > 10%; 2) stop reads >=3; 3) sites located within the peaks; 4) sites were not located in the peaks of control samples. Genomic annotations were performed using Homer software. Reads visualization was done by IGV. The histone marks, nucleosome and other ChIP-Seq signal data were downloaded from ENCODE database. Nucleosome signal data is from K562 cell line. For the analysis of the distribution of ChIP-Seq signals around cisplatin crosslinking sites, we

excluded the sites located in repeat regions. For the analysis of GG dinucleotide density, it was calculated as the GG dinucleotide occurrence per 100 bp. For the covariate analysis of nucleosome signal and GG dinucleotide density for cisplatin enrichment, regions were divided into 9 parts depending on their GG dinucleotide densities as well as their nucleosome signals; for GG dinucleotide density, 0-3 is defined as “low”, 3-7 is defined as “medium”, >7 is defined as “high”; for nucleosome signal, 0-5 is defined as “low”, 5-50 is defined as “medium”, >50 is defined as “high”. For each part of the regions, the number of cisplatin crosslinking sites per 10^6 bp were calculated, and then subjected to log change for the ballon plot.

[1] T. J. Meza, M. N. Moen, C. B. Vagbo, H. E. Krokan, A. Klungland, P. E. Grini, P. O. Falnes, *Nucleic Acids Res.* **2012**, *40*, 6620.

[2] D. Wang, R. Hara, G. Singh, A. Sancar, S. J. Lippard, *Biochemistry* **2003**, *42*, 6747.

[3] A. B. Robertson, J. A. Dahl, R. Ougland, A. Klungland, *Nat. Protoc.* **2012**, *7*, 340.