Genome-wide TOP2A DNA cleavage is biased toward translocated and highly transcribed loci

Xiang Yu,[1,8] James W. Davenport,[2,8] Karen A. Urtishak,[2] Marie L. Carillo,[2,9] Sager J. Gosai,[1,10] Christos P. Kolaris,[2,11] Jo Ann W. Byl,[3] Eric F. Rappaport,[4] Neil Osheroff,[3,5,6] Brian D. Gregory,[1] and Carolyn A. Felix[2,7]

[1]Biology Department, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA; [2]Division of Oncology, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA; [3]Department of Biochemistry, Vanderbilt University, Nashville, Tennessee 37232, USA; [4]NAPCore, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA; [5]Department of Medicine (Hematology/Oncology), Vanderbilt University, Nashville, Tennessee 37232, USA; [6]VA Tennessee Valley Healthcare System, Nashville, Tennessee 37212, USA; [7]Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

Type II topoisomerases orchestrate proper DNA topology, and they are the targets of anti-cancer drugs that cause treatment-related leukemias with balanced translocations. Here, we develop a high-throughput sequencing technology to define TOP2 cleavage sites at single-base precision, and use the technology to characterize TOP2A cleavage genome-wide in the human K562 leukemia cell line. We find that TOP2A cleavage has functionally conserved local sequence preferences, occurs in cleavage cluster regions (CCRs), and is enriched in introns and lincRNA loci. TOP2A CCRs are biased toward the distal regions of gene bodies, and TOP2 poisons cause a proximal shift in their distribution. We find high TOP2A cleavage levels in genes involved in translocations in TOP2 poison–related leukemia. In addition, we find that a large proportion of genes involved in oncogenic translocations overall contain TOP2A CCRs. The TOP2A cleavage of coding and lincRNA genes is independently associated with both length and transcript abundance. Comparisons to ENCODE data reveal distinct TOP2A CCR clusters that overlap with marks of transcription, open chromatin, and enhancers. Our findings implicate TOP2A cleavage as a broad DNA damage mechanism in oncogenic translocations as well as a functional role of TOP2A cleavage in regulating transcription elongation and gene activation.

[Supplemental material is available for this article.]

α β

p- –

′ KMT2A KMT2A

KMT2A KMT2A- r

r r p r

p

′

P

KMT2A

TOP2A cleavage events are SSNs more often than DSBs

Results

Optimization of procedure to detect TOP2 cleavage complexes

Genome-wide characterization of TOP2A cleavage cluster regions

′

KMT2A

′

p

p

r r r p

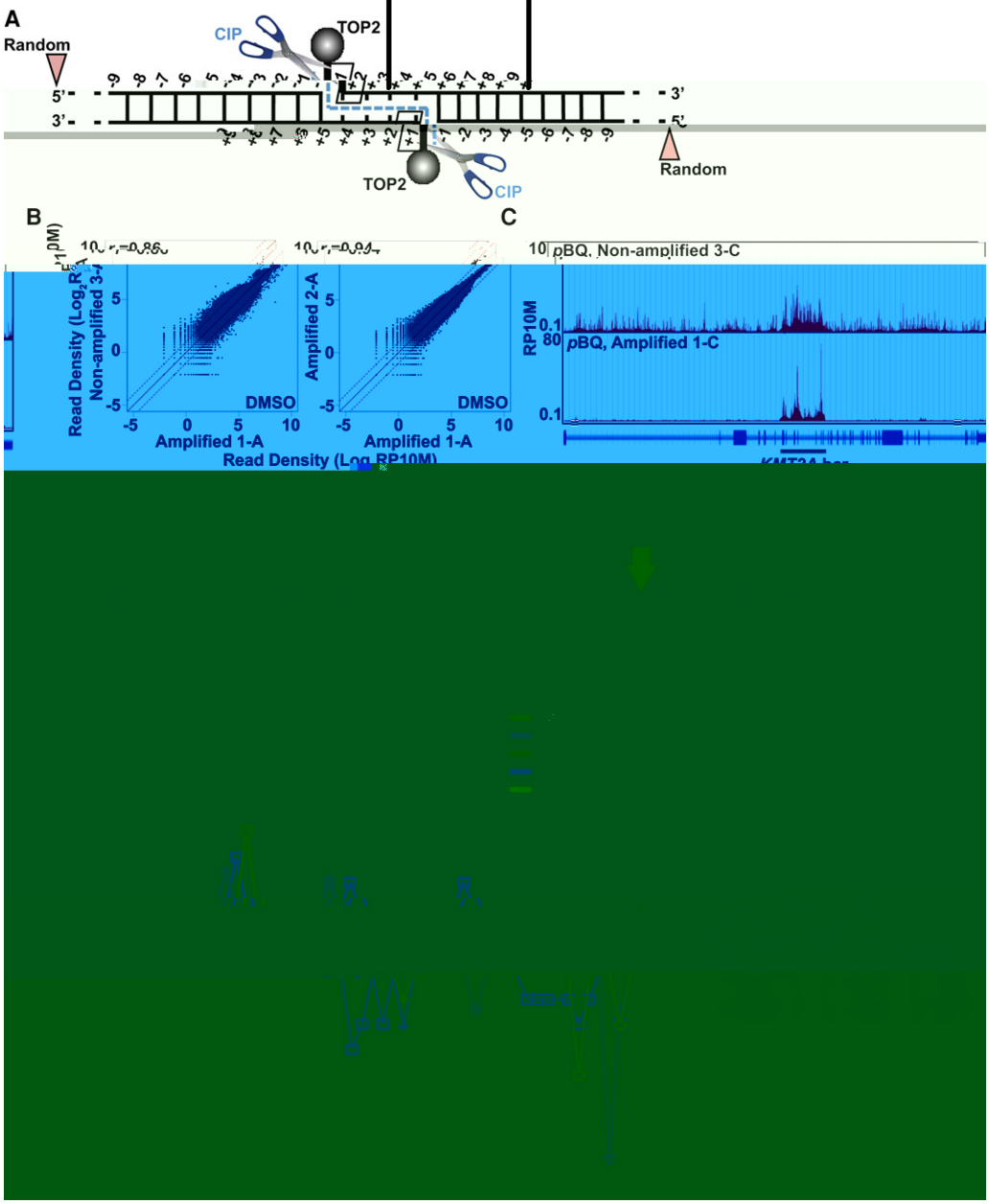


Figure 1. Approach, reproducibility, and assay validation. (*A*) TOP2 cleavage complexes detected by sequencing. After TOP2 immunocapture, CIP releases covalently attached TOP2 subunits from DNA at the +1 positions relative to the cleavage, which (+/− preamplification) become 5′ adapter-ligated ends; 5′ ends from sonication give random signals. Input control (data not shown) is sonicated lysate with random 5′ ends created by sonication. (*B*) Strong read count correlations in 10-kb windows between DMSO-treated biological replicates +/− preamplification (Supplemental Table S1). (*C*) UCSC Genome Browser images and *KMT2A* gene model (http://genome.ucsc.edu/) (Kent et al. 2002) showing similar read distribution in nonamplified and amplified *p*-benzoquinone (*p*BQ)–treated replicates. Black bar indicates bcr. (*D*) Overlap of TOP2A cleavage sites detected by sequencing with cleavage sites from the in vitro assay. Vertical line beneath the red arrow in the *KMT2A* bcr schematic is a translocation breakpoint hotspot from the TOP2A high-throughput sequencing assay (*bottom*, *left*) and autoradiograph *inset* from TOP2A in vitro cleavage assay of sense strand of same sequence (*bottom*, *right*). Colors indicate different treatments; symbols, different replicates. Arrows at peaks in sense strand (*bottom*, *left*) indicate +1 positions of cleavage sites also found in vitro (*bottom*, *right*, dashes). Connecting lines indicate sites with cleavage detected at +1 positions of both strands by sequencing (*bottom*, *left*). Coordinates, NC_000011.10 (GRCh38/hg38). (VP16) Etoposide. Bars beneath *KMT2A* bcr schematic, regions from both assays in Supplemental Figures S3 and S4.

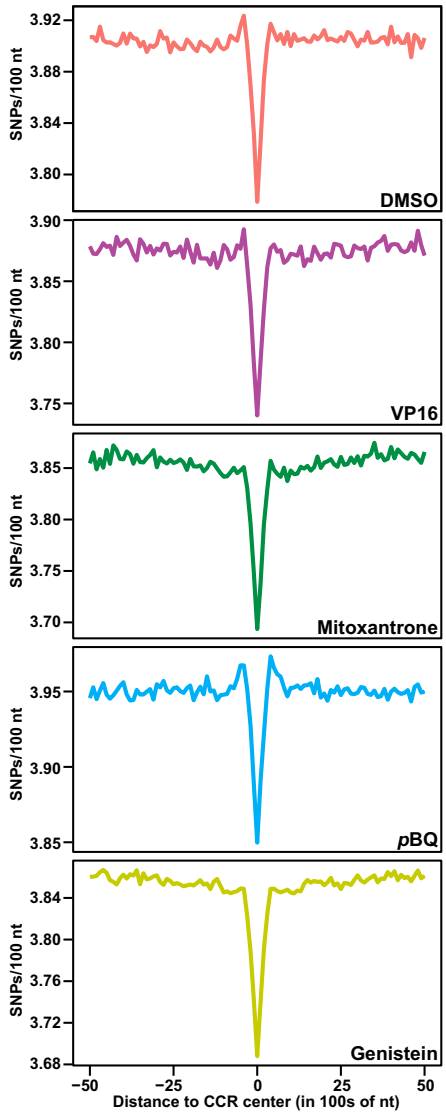


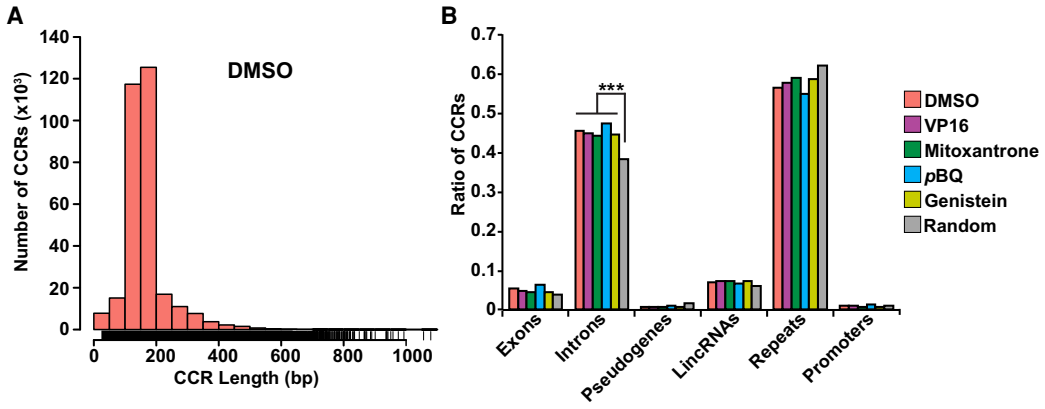
Figure 2. Regional TOP2A CCR sequence conservation. Lower SNP density within 100 nt at CCR centers compared with the surrounding 10 kb in 100-nt sliding windows. $P < 2.2 \times 10^{-16}$; Kruskal–Wallis test. Amplified samples; same treatments merged where applicable (Supplemental Table S1).

KMT2A

CCR profiles in coding genes support role of TOP2A cleavage in transcription elongation

Genes involved in oncogenic translocations show TOP2A CCR enrichment

KMT2A



– – ≥

r –

p- –

, ,

,

TOP2A cleavage of coding genes is independently associated with gene length and transcript abundance

TOP2A cleavage functions in lincRNA transcription

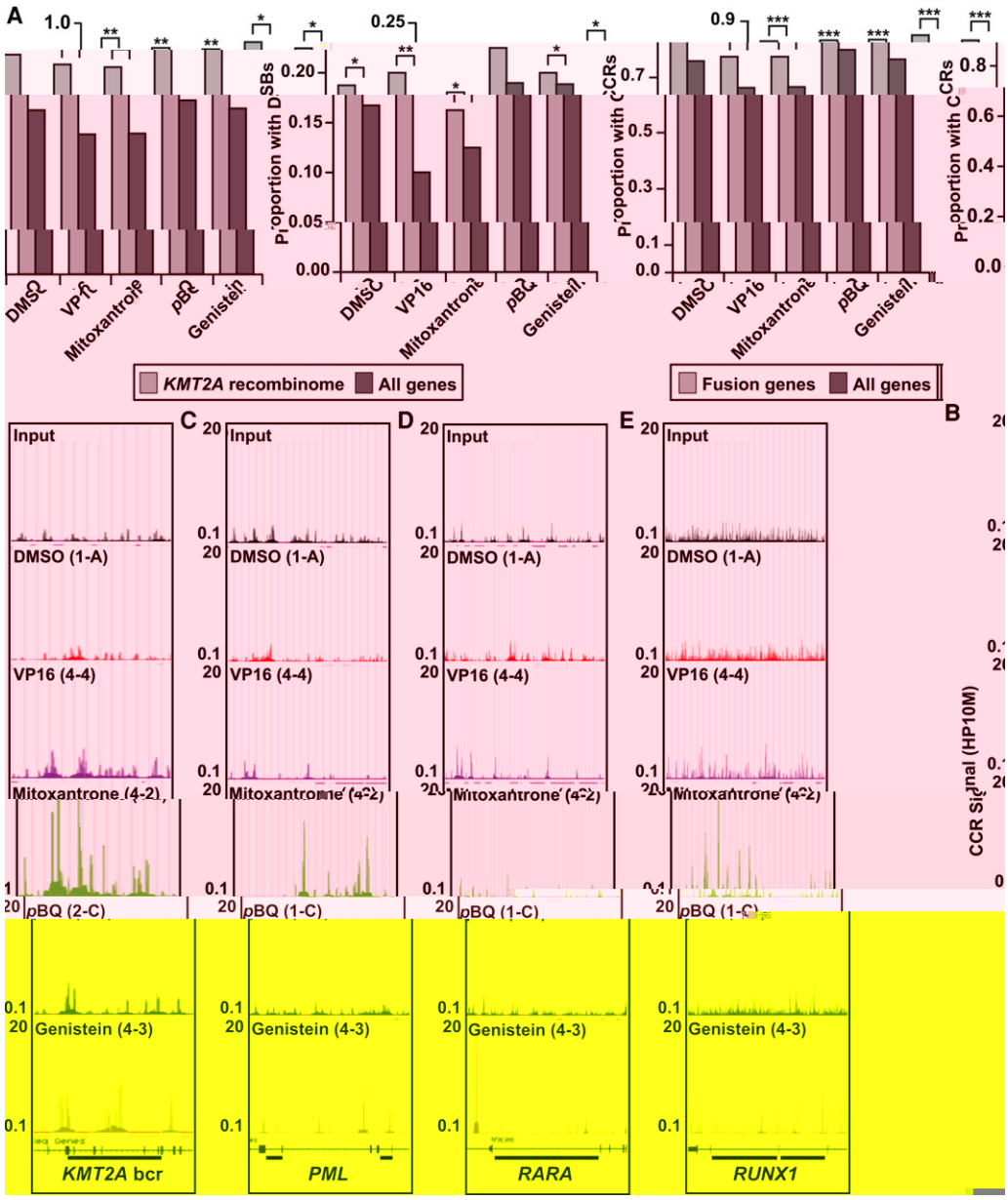
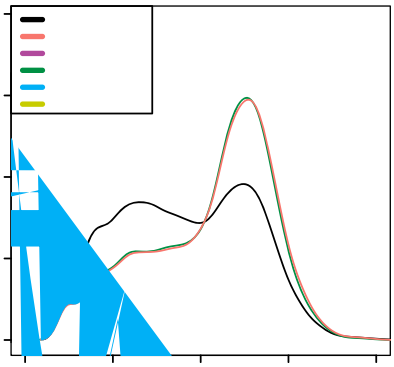
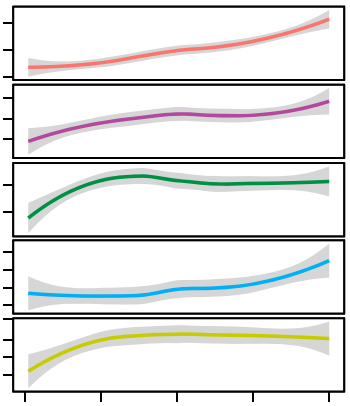


Figure 4. TOP2A cleavage in genes involved in oncogenic translocations. (*A*, *left*) Larger proportions of genes containing CCRs in *KMT2A* recombinome compared with all coding genes. (**) $P < 1 \times 10^{-5}$ for DMSO, VP16, mitoxantrone; (*) $P < 1 \times 10^{-4}$ for *p*BQ, genistein; χ^2 test. (*A*, *middle*) Larger proportions of genes containing TOP2A DSBs in *KMT2A* recombinome compared with all coding genes. (*) $P < 0.05$ for DMSO, mitoxantrone, *p*BQ, genistein; (**) $P = 0.00027375$ for VP16; χ^2 test. (*A*, *right*) Larger proportions of cancer fusion genes (Mitelman et al. 2016) containing CCRs compared with all coding genes. (***) $P < 2.2 \times 10^{-16}$; χ^2 test. Amplified samples; same treatments merged where applicable (Supplemental Table S1). (*B–E*) CCR signals (HP10M) in individual amplified samples along regions of genes involved in leukemia-associated translocations linked to TOP2 poisons (bars). Panels show sonicated input and different treatments (*top* in panels). Gene models from GRCh38/hg38 in the UCSC Genome Browser (*bottom*) (Kent et al. 2002; http://genome.ucsc.edu/) correspond to tracks shown. (*B*) *KMT2A*. Bar, 8.3-kb bcr spanning exon 7 through exon 13 positions 118,481,830–118,490,167; NC_000011.10. (*C*) *PML*. *Left* bar, 1.45-kb intron 3 bcr, positions 74,023,409–74,024,856; *right* bar, 1.06-kb intron 6 bcr, positions 74,033,415–74,034,477; NC_000015.10. (*D*) *RARA*. Bar, 16.9-kb intron 2 bcr, positions 40,332,397–40,348,315; NC_000017.11. (*E*) *RUNX1*. *Left* bar, 25-kb intron 6 bcr, positions





Clustering of chromatin features overlapping with CCRs

TOP2A CCRs are associated with open chromatin and enhancer elements

Discussion

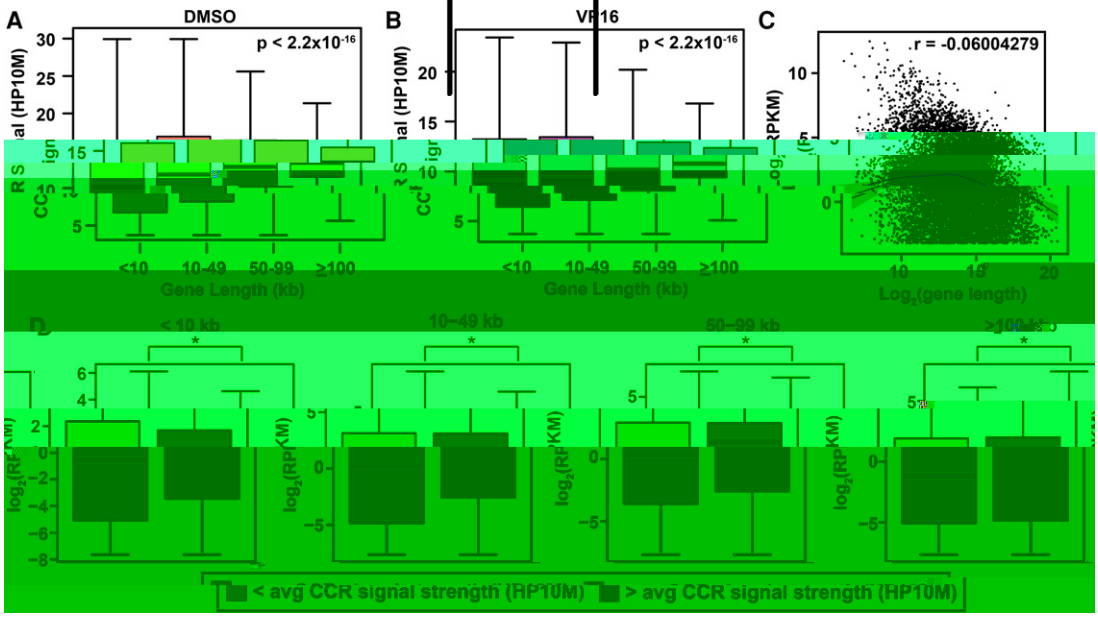


Figure 6. Independent associations of CCR signal strength with coding gene length and transcript abundance. (*A*,*B*) Correlation between length and genic CCR signal strength (HP10M) in DMSO-treated (*A*) and VP16-treated (*B*) samples. Gene length from GRCh38/hg38 by categories on *x*-axis. (Boxes) 25th to 75th percentiles; (whiskers) fifth to 95th percentiles; (horizontal lines) median for each length interval. χ^2 test *P*-values, *top right* in panels. Amplified samples; same treatments merged (Supplemental Table S1). (*C*) Scatter plot of protein-coding transcript abundance versus gene length based on two RNA-seq data sets for untreated K562 cells (GEO accession number GSE46718) (Bansal et al. 2014) and gene length from GRCh38/hg38. Smooth line was predicted by the gam method (Supplemental Methods). (Shading) Confidence interval around smoothed trend line. Genes with RPKM > 0.1 plotted. *r*-value (*top right*) shows slight overall negative correlation. (*D*) Box and whisker plots of genic CCR signal strength versus transcript abundance within indicated length categories subdivided based on </> average genic CCR signal strength. (*) $P < 0.05$; χ^2 test. Note correlation between genic CCR signal strength and transcript abundance across all lengths. Union set of all amplified samples (Supplemental Table S1).

KMT2A

KMT2A PML RARA RUNX1

Saccharomyces cerevisiae

KMT2A

KMT2A

KMT2A- –

Methods

Cell treatment and library preparation

Childhood leukemias
–
DNA Repair (Amst) **5** –

J Biol Chem **28** –

Oncogene **18** –

Front Cell Dev Biol **3**

Am J Epidemiol **168** –

Nucleic Acids Res **34**
–

Nat Struct Mol Biol **17**
–

EMBO J **2** –

Open Biochem J **5** –

Nucleic Acids Res **40**
–

β
Science **312** –

Proc Natl Acad Sci **111**
–

Genome Res **12**
–

Nat Struct Mol Biol **20** –

α
Proc Natl Acad Sci
10 –

J
Biol Chem **277** –

Biochemistry **43** –

Proc Natl
Acad Sci **8** –

β
Mol Cell
Biol **26** –

Leukemia
27 –

N Engl J Med **352**
–

α
Nature **413** –

Nucleic
Acids Res **35** –

PLoS One **4**

Nat Struct Mol Biol **20** –

,
BMC Genomics **8**

Escherichia coli
Biochemistry **40** –

“ ”
Genes Chromosomes Cancer **48** –

J Mol Recognit **7** –

Cell **144** –

Nat Chem
Biol **11** –

Ann N Y Acad Sci **1310** –

DNA Repair (Amst) **5**
–

Proc Natl Acad Sci –

Blood **111** –

Blood **2**
–

β
PLoS One **3**
– –

Cancer Res **58** –

Blood **100** –

,
Cancer Epidemiol Biomarkers Prev **14**
–

Proc Natl Acad Sci **108**
–

Mol Cell Biol **22** –

α
Nat Commun **4**

Nature **48** –

β
Proc Natl Acad Sci **10** –

Cancer Res
5 –

Biochemistry **28** –

Oncogene **22** –

Received July 17, 2016; accepted in revised form April 5, 2017.

Xiang Y , James W. Da enport, Karen A. Urtishak, et al.

Genome Res. p blished online April 6, 2017
Access the most recent ersion at doi:10.1101/gr.211615.116