



HAL
open science

Chromatin-associated MRN complex protects highly transcribing genes from genomic instability

Kader Salifou, Callum Burnard, Poornima Basavarajaiah, Giuseppa Grasso, Marion Helmoortel, Victor Mac, David Depierre, Céline Franckhauser, Emmanuelle Beyne, Xavier Contreras, et al.

► To cite this version:

Kader Salifou, Callum Burnard, Poornima Basavarajaiah, Giuseppa Grasso, Marion Helmoortel, et al.. Chromatin-associated MRN complex protects highly transcribing genes from genomic instability. *Science Advances* , 2021, 7 (21), 10.1126/sciadv.abb2947 . hal-03875724

HAL Id: hal-03875724

<https://hal.science/hal-03875724>

Submitted on 2 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

MOLECULAR BIOLOGY

Chromatin-associated MRN complex protects highly transcribing genes from genomic instability

Kader Salifou^{1†}, Callum Burnard^{1,2†}, Poornima Basavarajaiah^{1†}, Giuseppa Grasso¹, Marion Helmsmoortel¹, Victor Mac¹, David Depierre², Céline Franckhauser¹, Emmanuelle Beyne¹, Xavier Contreras¹, Jérôme Dejardin³, Sylvie Rouquier¹, Olivier Cuvier², Rosemary Kiernan^{1*}

MRN-MDC1 plays a central role in the DNA damage response (DDR) and repair. Using proteomics of isolated chromatin fragments, we identified DDR factors, such as MDC1, among those highly associating with a genomic locus upon transcriptional activation. Purification of MDC1 in the absence of exogenous DNA damage revealed interactions with factors involved in gene expression and RNA processing, in addition to DDR factors. ChIP-seq showed that MRN subunits, MRE11 and NBS1, colocalized throughout the genome, notably at TSSs and bodies of actively transcribing genes, which was dependent on the RNAPII transcriptional complex rather than transcription per se. Depletion of MRN increased RNAPII abundance at MRE11/NBS1-bound genes. Prolonged MRE11 or NBS1 depletion induced single-nucleotide polymorphisms across actively transcribing MRN target genes. These data suggest that association of MRN with the transcriptional machinery constitutively scans active genes for transcription-induced DNA damage to preserve the integrity of the coding genome.

INTRODUCTION

Execution of the appropriate transcriptional program is fundamental during growth, development, and response to environmental stimuli. Although an essential DNA-dependent process, transcription comes at a cost. It can induce double-strand breaks (DSBs) that are a deleterious form of DNA damage and can compromise the genome if erroneously repaired. DSBs are predominantly repaired by two canonical pathways: nonhomologous end joining (NHEJ) or homologous recombination (HR). HR involves extensive DNA end resection and uses an intact copy of the damaged locus to produce error-free repair. NHEJ ligates broken DSB ends with no or limited processing and frequently results in genomic mutations. Both pathways compete to repair a DSB (1), and mechanisms underlying pathway choice are not entirely clear. The nuclease activities of MRE11, a subunit of the MRE11-Rad50-NBS1 complex (MRN), are a critical determinant of pathway choice (2). MRE11 endonuclease activity initiates resection, thereby licensing HR. The exonuclease activities of MRE11 and EXO1/BLM (exonuclease 1/BLM RecQ like helicase) bidirectionally resect toward and away from the DNA end, which commits to HR and disfavors NHEJ.

While the complexity of the DNA break is an important factor in DNA damage response (DDR) pathway choice, recent studies have also identified the local nuclear environment and epigenetic landscape as determinants of repair pathway usage, predisposing to either NHEJ or HR pathways (3, 4). At actively transcribed regions, H3K36me3 is associated with a preference for DNA repair through the HR pathway (5–7). Nuclear organization has emerged as a key parameter in the formation of chromosomal translocations, implying that DSBs cluster into higher-order structures. Microscopy-based techniques and more recent chromosome conformation capture assays have shown that DSBs are mobile and form higher-order

structures detected as clusters (8–11). Initially thought to be factories for DNA repair, recent data have suggested that clustering may protect DSBs from spurious pairing to suppress translocations. Thus, DSBs in actively transcribed chromatin are not repaired in G₁ but remain in clusters until HR can proceed in G₂ (9). DSB clustering depends on several DNA repair factors including the MRN complex (8, 9, 11). Thus, a model emerges in which DSBs in active chromatin are sequestered in higher-order structures that permit the HR repair pathway to proceed under conditions that minimize the formation of deleterious translocations. However, the molecular mechanism by which HR is chosen as a predominant pathway for DSB repair in transcriptionally active regions remains unclear. Furthermore, most studies have been performed by inducing DNA damage using irradiation or chemicals or artificially cutting the genome. Mechanisms involved in endogenous transcription-associated DNA repair remain poorly understood.

To address the recruitment of DNA damage repair (DDR) factors at transcriptionally active regions, we used an unbiased proteomic approach, proteomics of isolated chromatin fragments (PICh) (12), of the inducibly transcribed HIV-1 promoter. Upon transcriptional activation, DNA repair factors were among the most recruited interactants. In particular, mediator of DNA damage checkpoint (MDC1) that associates with the MRN complex was recruited specifically following activation of transcription. Identification of the interactome of endogenous MDC1 by tandem affinity purification followed by mass spectrometry similarly identified numerous factors implicated in transcription and cotranscriptional RNA processing. Coimmunoprecipitation analysis showed that MDC1 interacts with RNA polymerase II (RNAPII) and coactivators, such as P-TEFb. ChIP-seq (chromatin immunoprecipitation followed by high-throughput sequencing) analysis of MRN subunits, MRE11 and NBS1, revealed an association with genes, specifically those that were transcriptionally active. MRE11 and NBS1 demonstrated covariation with RNAPII upon heat shock-induced transcriptional activation. Blockade of transcriptional elongation altered the profile of NBS1 to mirror that of RNAPII, suggesting that MRN is recruited through its association with the transcriptional

Copyright © 2021
The Authors, some
rights reserved;
exclusive licensee
American Association
for the Advancement
of Science. No claim to
original U.S. Government
Works. Distributed
under a Creative
Commons Attribution
NonCommercial
License 4.0 (CC BY-NC).

¹CNRS-UMR 9002, Institute of Human Genetics (IGH)/University of Montpellier, Gene Regulation Lab, 34396, France. ²Center of Integrative Biology (CBI) CNRS/Université de Toulouse (UMR 5077), 31000, France. ³CNRS-UMR 9002, IGH/University of Montpellier, Biology of Repeated Sequences Lab, 34396, France.

*Corresponding author. Email: rosemary.kiernan@igh.cnrs.fr

†These authors contributed equally to this work.

machinery rather than the level of transcription per se. Following MRE11 or NBS1 depletion, the occurrence of single-nucleotide polymorphisms (SNPs) was highly correlated to the binding of MRE11, NBS1, or RNAPII. These data suggest that the MRN complex, through its interaction with the transcription machinery, scans active genes for transcription-induced DNA damage.

RESULTS

DDR factors are associated with active transcription

We investigated the recruitment of DDR factors during transcription using an unbiased proteomic method, PICh (12) of the HIV-1 promoter region, which becomes highly transcribed upon addition of its cognate transactivator, Tat. PICh was performed using chromatin extracts of cells containing tandem integrated copies of the HIV-1 long terminal repeat (LTR) linked to MS2 binding sites (13), in the presence or absence of the transactivator protein, Tat (fig. S1, A and B). Mass spectrometry identified 438 proteins associated with the locus in the control condition and 622 proteins associated with the locus upon transactivation by Tat. Upon transcriptional activation, 58 proteins were highly gained [fold change (FC), >7], 294 were moderately gained (FC, >2 and <7), 29 were moderately lost (FC, <-2 and >-7), and 3 were highly lost (FC, <-7) (Fig. 1A and table S1). Gene ontology revealed a number of pathways that were enriched in cells expressing Tat, compared to control untreated cells (fig. S1C). As expected, factors involved in gene expression were among the most highly gained. A number of factors involved in DNA repair were associated with HIV-1 chromatin specifically upon transactivation of transcription with Tat. Factors implicated in both the NHEJ and HR pathways were identified (table S1). However, NHEJ factors, such as DNA-dependent protein kinase (DNA-PK) and Ku70/80 are known to be implicated in transcription (14, 15), including Tat-mediated transcription of the HIV-1 LTR (16–18). Furthermore, DNA-PK has been shown to interact with HIV-1 Tat, while Ku70/80 has been shown to bind HIV-1 TAR RNA (16, 18, 19). Therefore, it is not clear whether recruitment of these factors following Tat transactivation was due to DNA damage or increased Tat recruitment and transcription. A key factor in the HR pathway, MDC1 was highly represented (Fig. 1A). To validate the interaction of DDR factors with actively transcribed HIV-1, we performed ChIP in cells containing an integrated HIV-1 LTR linked to a luciferase reporter, as described previously (20). Tat-induced activation of HIV-1 transcription increased recruitment of RNAPII as expected (fig. S1D). Similarly, association of MDC1, MRE11, and Ku80 were also significantly enhanced in the presence of Tat (fig. S1D), in agreement with the recruitment of these factors upon transcriptional activation identified by PICh. We furthermore determined that γ H2AX, a marker of double-stranded DNA breaks, was also increased following Tat transactivation (fig. S1E), suggesting that DDR factors may be recruited following DNA damage induced by Tat-mediated transcription.

These data indicated that DDR factors such as MDC1 and MRN may be closely linked with active transcription. To further explore this, we identified the interactome of MDC1 in the absence of exogenously induced DNA damage or Tat transactivation. Using CRISPR-Cas9 genome editing technique, a Flag-hemagglutinin (HA) tag was appended to the N terminus of endogenous MDC1 protein in human embryonic kidney (HEK) 293T cells. Tandem affinity purification of Flag-HA MDC1 was performed, as described previously

(Fig. 1B) (20–22). Gene ontology analysis of the interactants identified association of MDC1 with factors in the DNA damage repair pathway and the cell cycle, as expected (Fig. 1C). MDC1 interactants were also implicated in gene expression pathways (Fig. 1C and table S2). Association of endogenous MDC1 with RNA processing factors, such as SNRNP200, PRP8, HNRNPM, EFTUD2, and SF3B1, was validated by coimmunoprecipitation analysis (Fig. 1D, left) in cells that do not express HIV-1 Tat. Interaction between MDC1 and RNAPII (total and C-terminal repeat domain phosphorylated forms), as well as subunits of P-TEFb, cyclin-dependent kinase 9 (CDK9), and cyclin T1, was also detected by coimmunoprecipitation analysis (Fig. 1D, right), confirming a previous report (23). Moreover, immunoprecipitates of cyclin T1 contained MDC1 and MRN subunits, MRE11 and NBS1 (fig. S1F). Furthermore, MDC1 was specifically associated with the small, active P-TEFb complex, as shown by glycerol gradient analysis (fig. S1G). Together, these data show that the highly transcribing HIV-1 locus recruits DDR factors such as MDC1 and MRN and, conversely, that DDR factor MDC1 is physically associated with factors implicated in transcription and RNA processing, independently of the presence of HIV-1 Tat. These data reveal that the transcription and DDR pathways are intimately linked at the biochemical level through physical interactions.

MRN complex is associated with chromatin

Given its interaction with factors implicated in RNAPII transcription, we next sought to determine the profile of chromatin-associated MDC1/MRN complex across the coding genome. Anti-MDC1 antibodies proved unsuitable for ChIP-seq. We therefore performed ChIP-seq analysis of MRN subunits, NBS1 and MRE11, as well as RNAPII in cells in the absence of exogenous DNA damage. The average profile of signal for both MRE11 and NBS1 at active genes resembled that of RNAPII, although the amount of signal detected was lower (Fig. 2A). While the highest accumulation of MRE11 and NBS1 ChIP-seq reads occurred near the transcription start site (TSS), significant levels of binding were also detected in the gene body and downstream of the transcription end site (TES; Fig. 2A). The binding profile of MRE11 and NBS1 on representative genes, RBM17 and GNG12, is shown (Fig. 2B). Association of MDC1 with several genes, including RBM17 (RNA Binding Motif Protein 17) and GNG12 (G Protein Subunit Gamma 12), was also confirmed by ChIP-quantitative polymerase chain reaction (qPCR) (fig. S2A). More than 50,000 peaks were detected throughout the genome for either MRE11 or NBS1, and 110,000 peaks were detected for RNAPII. Although the highest signal for both MRE11 and NBS1 was found around TSSs (Fig. 2, A and B), the number of peaks overlapping TSSs represented 2.19 and 1.44%, respectively (Fig. 2C). MRE11 and NBS1 were also detected on TESs/transcription termination sites (<2%) and enhancers (3.16 and 1.74%, respectively) (Fig. 2C). The majority of MRE11 and NBS1 peaks were found on gene bodies (approximately 52%), albeit the average size of genes (>26 kb) (24) could account for this result because of the higher chance of overlap with the gene body compared to TSSs/TESSs. However, the signal associated with peaks in gene bodies was significantly lower than that detected at TSSs (Fig. 2, B and C). Furthermore, the distribution of ChIP-seq signal for MRE11 and NBS1 along gene bodies was similar to that of RNAPII (fig. S2B). Note that peaks of MRE11 and NBS1 mapped approximately 20 base pairs (bp) upstream of RNAPII, on average (Fig. 2B and fig. S2C), which may be due to an artefact of cross-linking a large multisubunit complex during the ChIP experiment.

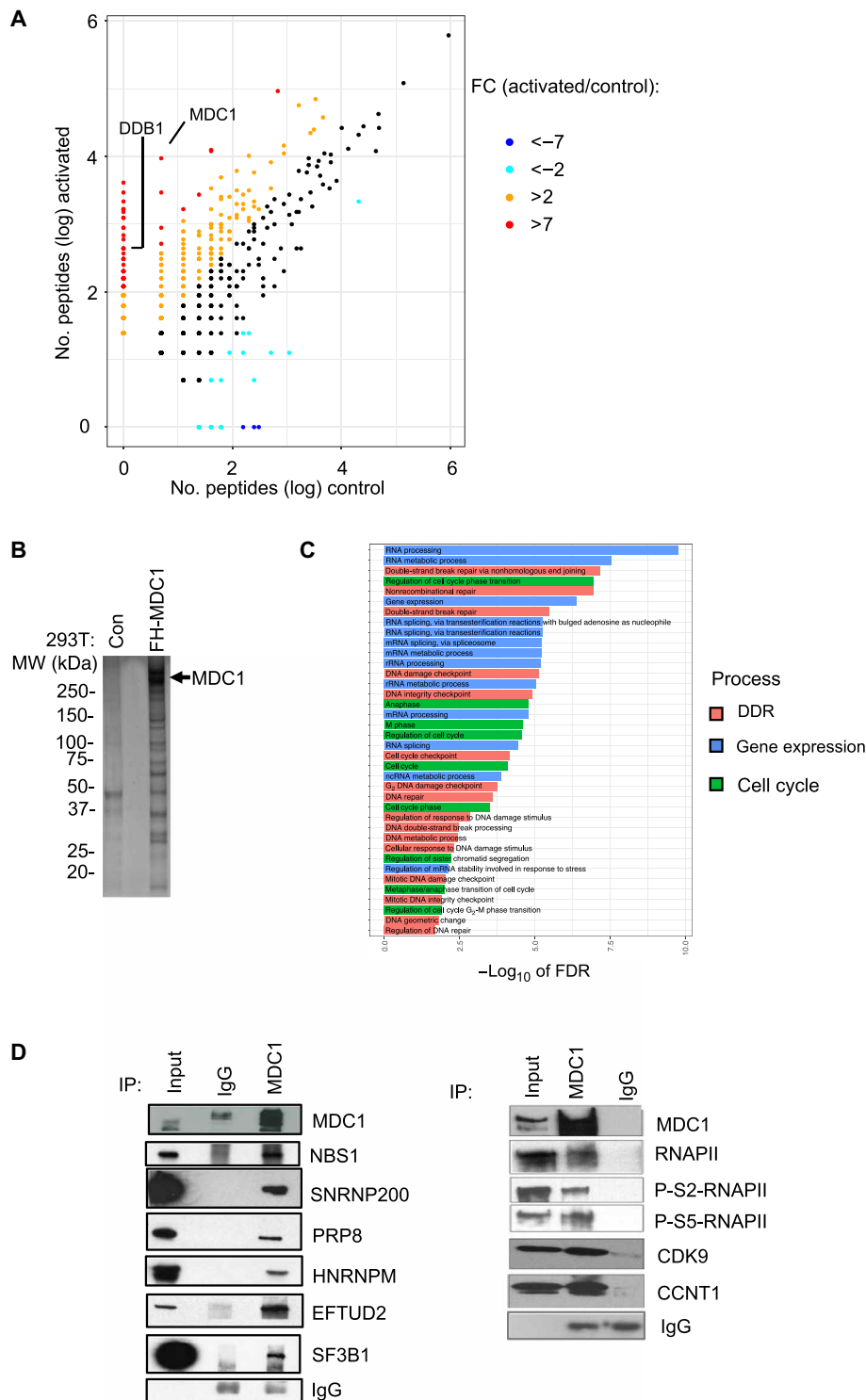


Fig. 1. DNA repair factors are associated with active transcription. (A) Scatter plot representation of proteins differentially associated with an integrated HIV-1 LTR-minigene upon HIV-1 promoter transactivation by Tat, from highly gained (FC, >7) to highly lost (FC, <7), as identified by PICh coupled to mass spectrometry. (B) SDS–polyacrylamide gel electrophoresis (SDS–PAGE) analysis followed by silver staining of eluates of tandem affinity purified nuclear extracts from HEK-293T expressing Flag-HA fused to endogenous MDC1 (FH-MDC1) or control cells (con), in the absence of HIV-1 Tat. (C) Gene Ontology analysis carried out submitting the interactants of FH-MDC1 identified by tandem affinity purification followed by mass spectrometry. GO terms having the highest significance are shown. FDR, false discovery rate. (D) Coimmunoprecipitation analysis of MDC1 performed using nuclear extracts of HEK-293T in the absence of HIV-1 Tat and the indicated antibodies. Normal immunoglobulin G (IgG) was used as a negative control. Nuclear extracts (5%) and immunoprecipitates were separated by SDS–PAGE and immunoblotted using the antibodies indicated in the figure.

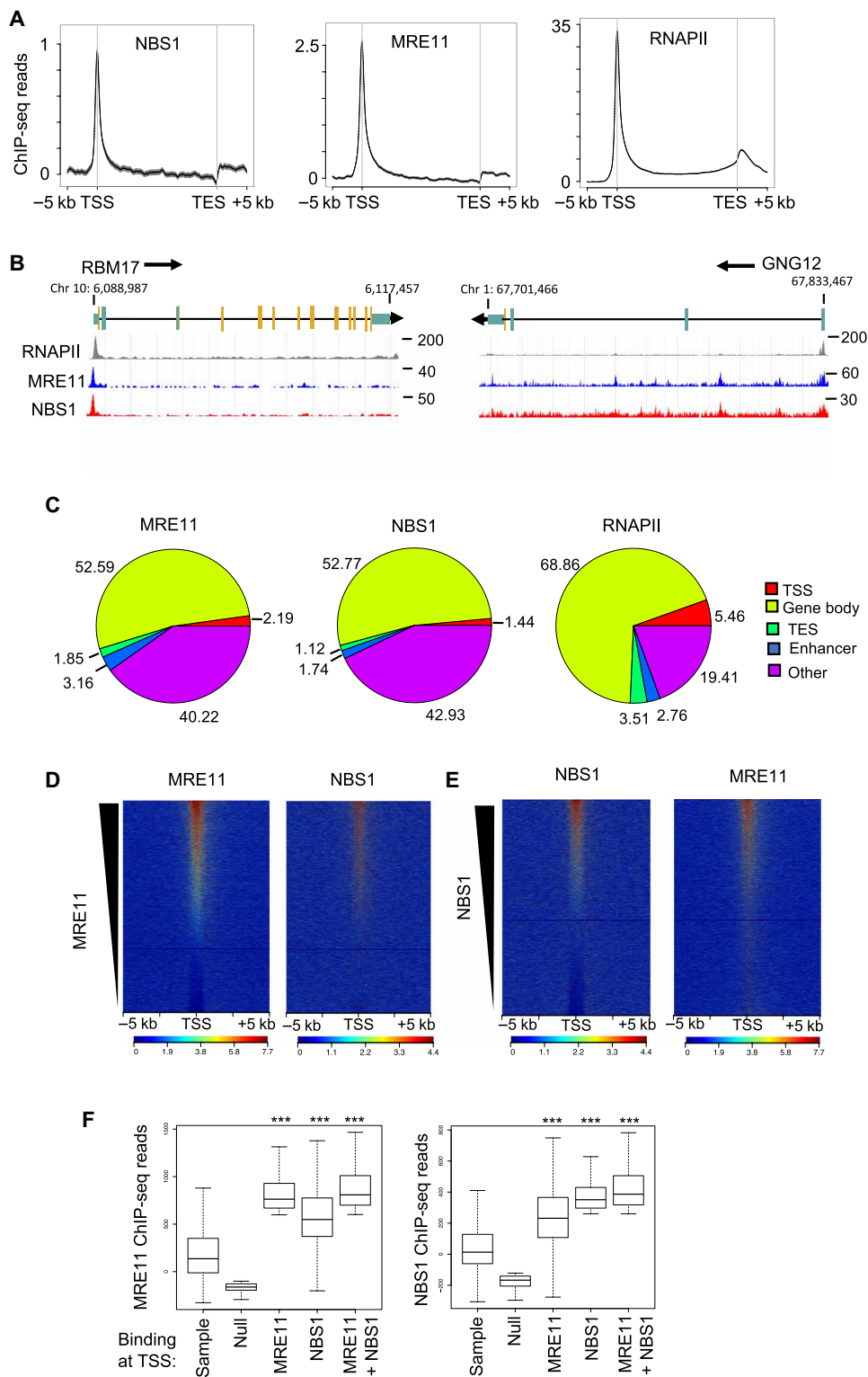


Fig. 2. MRN complex is associated with chromatin. (A) Scaled average density profiles of MRE11, NBS1, and RNAPII ChIP-seq reads across genes, ± 5 kb. (B) Browser shots of RNAPII, MRE11, or NBS1 ChIP-seq signal over a set of representative genes, as indicated. A schematic representation of the gene is shown above. (C) Pie charts showing the genomic distribution of ChIP-seq peaks of MRE11, NBS1, and RNAPII, as indicated. (D and E) ChIP-seq heatmaps centered and rank-ordered on MRE11 (D) or NBS1 (E). ChIP-seq reads of NBS1 (D) or MRE11 (E) were plotted respecting the same ranking. (F) Box plots of MRE11 or NBS1 ChIP-seq reads at genes for which the TSS was highly bound by MRE11, NBS1, both, or neither, as indicated, compared to a random sample of genes (*** $P < 0.001$, Wilcoxon test; $n = 708, 479, 1631, 1631, \text{ and } 708$, respectively).

Further genomic analysis of the localization of MRN on chromatin was performed by ranking genes according to the binding of MRE11 at TSSs (number of ChIP-seq reads per TSS, highest to lowest), as shown in Fig. 2D (left). Ranking genes by MRE11 reads also efficiently sorted the NBS1 signal (Fig. 2D). Similarly, gene ranking based on NBS1 read intensity at the TSS also efficiently ranked the MRE11 signal (Fig. 2E), highlighting a good correlation in their loading onto genes. Next, genes were grouped on the basis of high TSS binding of either MRE11 or NBS1, both factors together, or neither factor. As expected, genes highly bound by MRE11 had significantly more ChIP-seq reads for this factor compared to the random sample group (Fig. 2F, compare box 3 with 1; $P < 1 \times 10^{-20}$). The group of genes that were highly bound by NBS1 also had a significantly high number of MRE11 reads at the TSS compared to the sample group ($P < 1 \times 10^{-20}$; left box plot, compare box 4 with 1). Moreover, the highest number of MRE11 reads were found at cobound genes when compared to the sample group (left box plot, compare box 5 with 1; $P < 1 \times 10^{-20}$) or to TSSs highly bound by MRE11 only ($P < 1 \times 10^{-7}$; left box plot, compare box 5 with 3). The same trend was observed for NBS1 binding (Fig. 2F, right box plot). TSSs highly bound by NBS1 showed a high number of NBS1 reads as expected (compare box 4 with 1; $P < 1 \times 10^{-20}$), while TSSs highly bound by MRE11 also showed significantly high binding of NBS1 compared to the random sample group of genes (compare box 3 with 1; $P < 1 \times 10^{-20}$). As observed for MRE11, the highest number of NBS1 reads were found at genes that were cobound by MRE11 (compare box 5 to 4; $P < 1 \times 10^{-13}$). Correlation of MRE11 and NBS1 ChIP-seq reads furthermore confirmed that the two factors co-associate with the same regions (fig. S2D). These data indicate that MRE11 and NBS1 co-associate with a subset of human genes, likely in the context of the MRN DNA repair complex. They furthermore support proteomic data (Fig. 1) that uncovered previously unidentified biochemical interactions between MDC1/MRN complex and factors implicated in transcription and cotranscriptional mRNA processing.

MRN complex is associated with actively transcribed regions

Since MRE11 and NBS1 were found to be highly associated with TSSs and gene bodies, we wished to further investigate their association with transcription. First, genes were ranked according to the number of RNAPII ChIP-seq reads surrounding the TSS (± 1 kb), from highest to lowest, which likely reflects the transcriptional activity of each gene. Next, respecting the same ranking, ChIP-seq reads for MRE11 and NBS1 were plotted. The resulting heatmaps (Fig. 3A) show that ranking genes by RNAPII also efficiently ranked both MRE11 and NBS1. Similarly, the averaged density profiles showed that MRE11 and NBS1 were more associated with genes having high RNAPII at the TSS than those having low RNAPII (Fig. 3A, top).

We next analyzed the amount of RNAPII at the TSS of genes highly bound by MRE11, NBS1, both MRE11 and NBS1, or neither factor (null) compared to a random sample group of genes (sample). Genes highly bound by either NBS1 or MRE11, alone or together, had significantly higher levels of RNAPII at the TSS compared to the random sample group of genes (Fig. 3B; boxes 3, 4, and 5, compared to box 1; $P > 1 \times 10^{-20}$). Pairwise correlation of MRE11, NBS1, and RNAPII ChIP-seq reads (figs. S2D and S3A) suggested that these factors co-associate with the same regions. To further demonstrate co-association of MRE11, NBS1, and RNAPII at cellular genes and HIV-1 LTR, Re-ChIP was performed (fig. S3B). This analysis demonstrated that chromatin fragments immunoprecipitated using NBS1 or MRE11

antibody could be reprecipitated by antibody against RNAPII, indicating that both proteins were colocalized on chromatin.

The association of MRE11 and NBS1 at the TSS of genes harboring varying amounts of RNAPII at the TSS was validated by ChIP-qPCR. The abundance of MRE11 or NBS1 largely mirrored that of RNAPII (fig. S3C). Furthermore, intersection analyses of ChIP-seq data showed a preferential colocalization of MRE11/NBS1 ChIP-seq peaks with active genes, defined as the top 50% of expressed genes when ranked by normalized RNA sequencing (RNA-seq) reads (25). Considering the peaks of each factor found at TSSs, 84% of MRE11 peaks ($n = 1597$), 79% of NBS1 peaks ($n = 687$), and 88% of MRE11 + NBS1 peaks ($n = 672$) were found at TSSs of active genes. Since RNAPII is also found at active enhancers, we asked whether MRE11 and NBS1 might associate with these regions. Both MRE11 and NBS1 were localized to enhancers, as compared to TSSs and random intergenic regions (Fig. 3C). Similar to TSSs, MRE11 and NBS1 were detected at enhancers with high levels of RNAPII, which are likely active enhancers, and were poorly associated with enhancers having low levels of RNAPII (Fig. 3D). Together, these data indicate the MRE11 and NBS1 subunits of MRN complex are associated with actively transcribing regions in a manner that appears to be tightly correlated to that of RNAPII.

We next tested whether the MRN complex co-associates with RNAPII following induction of a specific transcriptional program. To this end, cells were heat-shocked to induce the expression of a subset of genes. ChIP-seq of RNAPII, MRE11, and NBS1 was performed in heat-shocked and untreated cells. MRE11 and NBS1 binding was analyzed at genes induced upon heat shock, that is, those showing an increase in RNAPII association. As shown in Fig. 3E, both MRE11 and NBS1 became significantly associated with the TSS (left) and gene body (right) of up-regulated genes compared to a sample group of genes. Increased association of RNAPII and NBS1 was confirmed by ChIP-qPCR at several heat shock genes (Fig. 3F). The DSB marker, γ H2AX was also increased following heat shock (Fig. 3G), suggesting that higher levels of transcription were associated with increased DNA damage at these regions. On the other hand, α -amanitin treatment was used to diminish RNAPII association with genes. As shown in fig. S3D, α -amanitin treatment induced loss of RNAPII as well as both MRE11 and NBS1 at target genes. Together, these data confirm that the binding of MRN tightly correlates with that of actively transcribing RNAPII.

MRN association with chromatin depends on RNAPII rather than transcription levels

Association of the MRN complex with actively transcribing regions is highly correlated to that of RNAPII (Fig. 3). However, it is not clear whether MRN is recruited through its interaction with the transcriptional machinery (Fig. 1) (23) or during the process of transcription. To distinguish between these possibilities, we sought to analyze MRN association under conditions where the abundance of RNAPII is not strictly correlated with the amount of transcription. To do so, RNAPII transcriptional elongation was blocked using the adenosine analog 5,6-dichloro-1- β -D-ribofuranosylbenzimidazole (DRB). While DRB abolishes transcription in the coding region, it does not greatly affect the amount of transcriptional initiation. Consequently, DRB treatment causes RNAPII to accumulate at the 5' end of genes disproportionately to the amount of transcription at the site. After 1 hour of DRB treatment, RNAPII accumulated near the TSS and diminished in the coding region and particularly at the

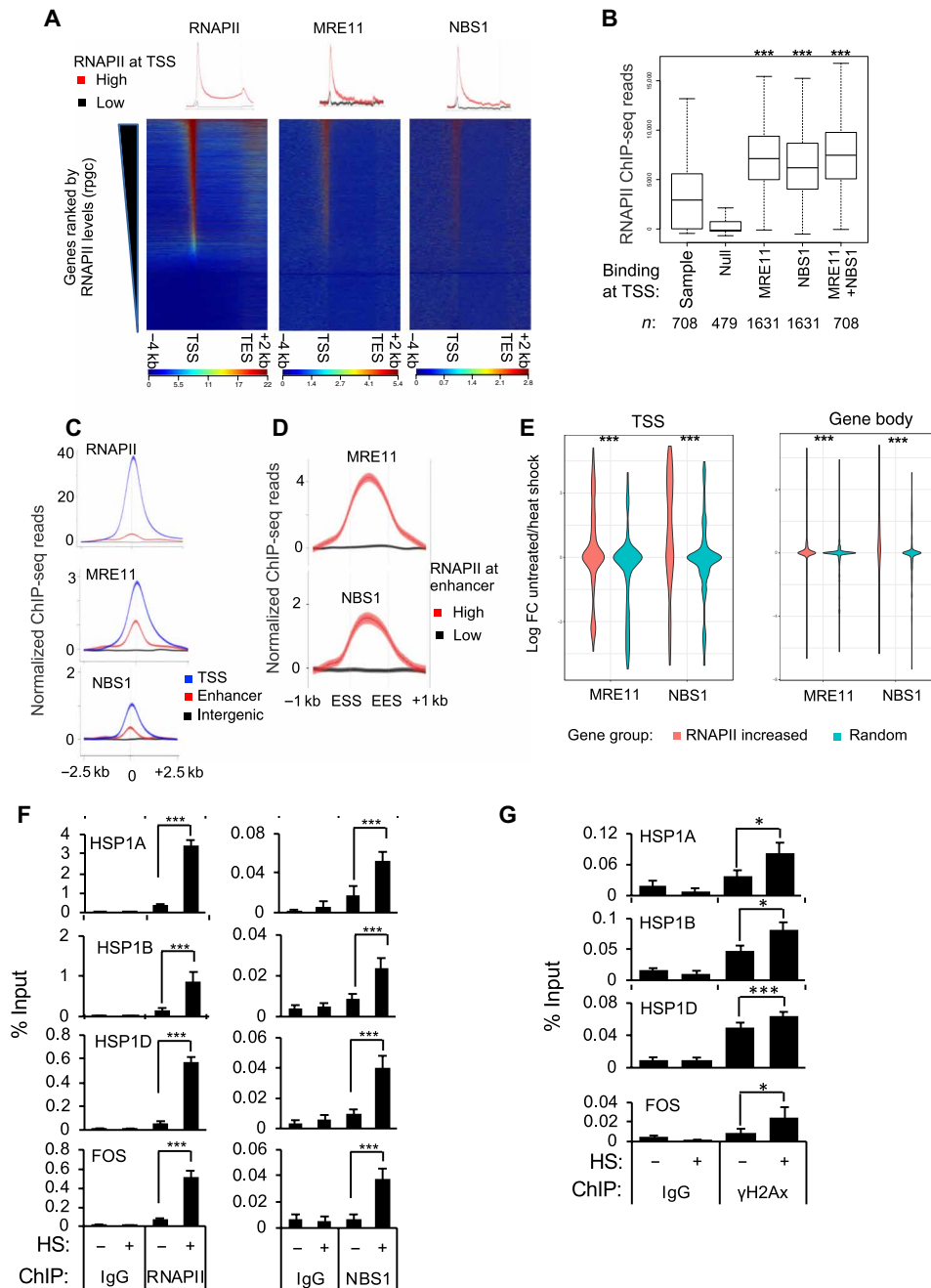


Fig. 3. MRN complex is associated with actively transcribed regions. (A) Heatmaps rank-ordered on normalized RNAPII ChIP-seq reads at the TSS and centered on genes. ChIP-seq reads of MRE11 or NBS1 were plotted respecting the same ranking. Averaged profiles of RNAPII, MRE11, or NBS1 ChIP-seq reads at the highest- and lowest-ranked RNAPII genes are shown above. (B) Box plots of normalized RNAPII ChIP-seq reads at genes highly bound at the TSS by MRE11, NBS1, both MRE11 and NBS1, or neither (null), compared to a random sample of genes ($***P < 0.001$, Wilcoxon test). (C) Averaged profiles of RNAPII, MRE11, or NBS1 ChIP-seq reads centered on TSS, enhancers, or intergenic regions. (D) Averaged profiles of MRE11 or NBS1 ChIP-seq reads at enhancers having the highest- or lowest-ranked RNAPII signal (ESS, enhancer start site; EES, enhancer end site). (E) Violin plots showing changes in MRE11 or NBS1 ChIP-seq reads at the TSS and body of genes at which RNAPII signal was increased in heat shocked samples relative to untreated controls, compared to a random group of genes in the same conditions ($***P < 0.001$, Wilcoxon test). (F) Binding of NBS1 or RNAPII was analyzed by ChIP-qPCR at the indicated genes from cells exposed or not to heat shock (HS). Data represent means \pm SEM ($***P < 0.001$, Student's *t* test; $n = 3$). (G) Association of γ H2AX was analyzed by ChIP-qPCR at the indicated genes from cells exposed or not to heat shock. Data represent means \pm SEM ($***P < 0.001$ and $*P < 0.05$, Student's *t* test; $n = 3$).

3' end of genes in the termination window, indicating that RNAPII elongation was efficiently inhibited (Fig. 4A). A longer blockade with DRB (3-hour treatment) caused RNAPII to accumulate at TSS boundaries, causing apparent accumulation near the start of the gene body as well as at upstream regions. By 1 hour after release, the profile of RNAPII had returned to that in untreated cells at the TSS and in the gene body, with a small reduction still evident at the TES.

Analysis of the profile of NBS1 under conditions of DRB treatment and release showed that it largely mirrored that of RNAPII, although the effects were more modest. The profiles showed the same qualitative effects, such as loss of signal at the TES, and leaking into TSS adjacent regions observed after 3 hours of treatment. The change in profiles of NBS1 and RNAPII were also quantitatively significant at both the TSS and TES regions (Fig. 4B). RNAPII and NBS1 association with chromatin during transcriptional blockade

and release were confirmed by ChIP-qPCR at representative genes (Fig. 4C). Last, we analyzed the correlation between RNAPII and NBS1 association at individual genes for each condition. Association of the two factors with genes was positively correlated at each condition, with the highest correlation at 1 and 3-hour DRB treatment (Fig. 4D). Thus, overall, although the effects measured for NBS1 are more modest than those for RNAPII, the similarities in their binding profiles and dynamics suggest that MRN is likely recruited through its association with the RNAPII transcription complex rather than the amount of transcription per se.

Binding of MRN affects the transcriptional output of target genes

We next wondered whether the presence of MRN affects the level of transcription at its target genes. To test this, MRE11 and NBS1 were

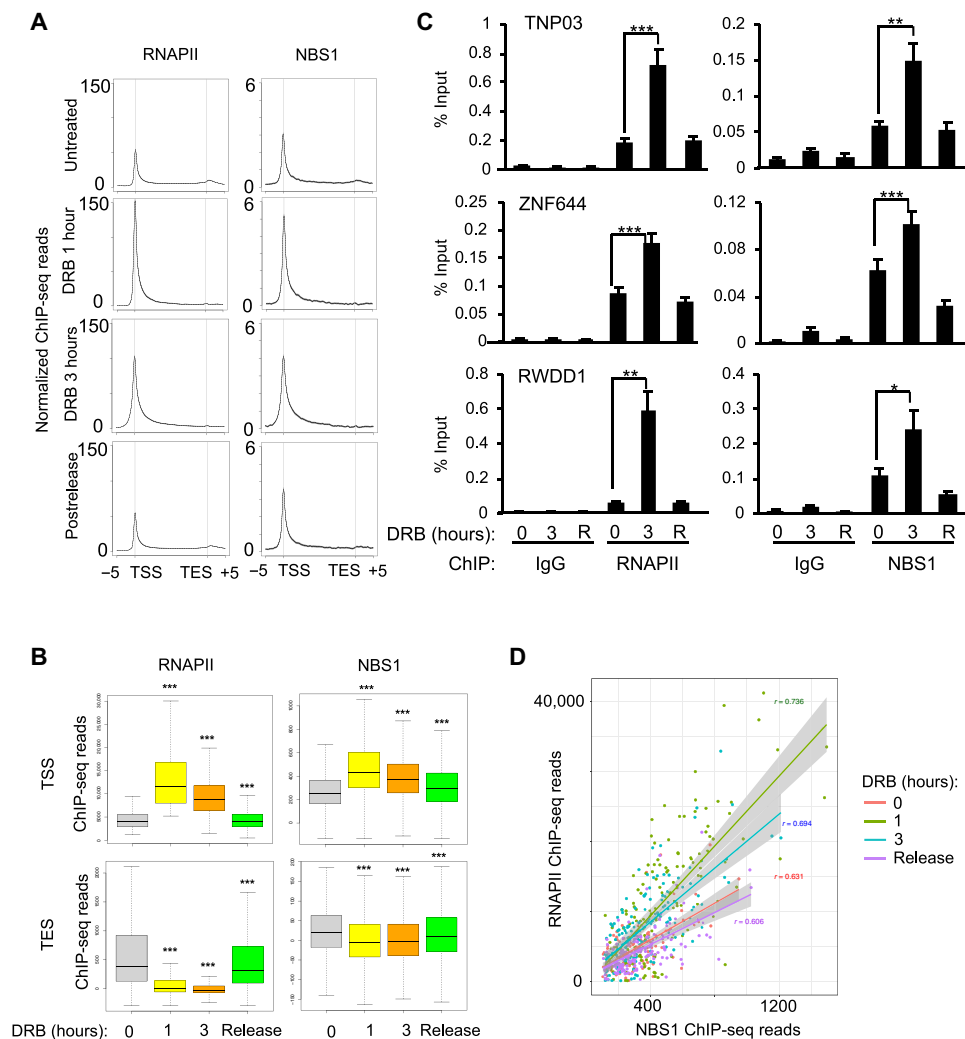


Fig. 4. MRN association with chromatin is correlated with the amount of RNAPII rather than transcription. (A) Average density profiles of ChIP-seq reads of RNAPII and NBS1 across genes showing the highest increase in RNAPII reads at 1-hour DRB treatment compared to a mock-treated control. ChIP-seq reads for RNAPII and NBS1 are shown at the same genes ± 5 kb following treatment with DRB for 1, 3, or 1 hour after release, as indicated. (B) Box plots of RNAPII and NBS1 ChIP-seq reads at the TSS and TES of regions shown in (A) in DRB-treated samples compared to nontreated controls (0) ($***P < 0.001$, Wilcoxon test; $n = 4924$). (C) Association of RNAPII and NBS1 with selected target genes was measured by ChIP-qPCR following mock treatment or DRB treatment for 3 or 1 hour after release, as indicated. Data represent means \pm SEM obtained from three independent experiments ($***P < 0.001$, $**P < 0.01$, and $*P < 0.05$, independent Student's *t* test). (D) Scatter plot showing ChIP-seq reads of RNAPII and NBS1 at genes shown in (A) in untreated samples or samples treated with DRB, as indicated. The coefficient of correlation (*r*) was calculated for each condition as indicated on the graph.

depleted by RNA interference (RNAi). Consistent with previous findings showing that MRE11 is required for stability of the MRN complex (26), knockdown of MRE11 induced loss of MRE11 and destabilized NBS1 in cell extracts (fig. S4A). In contrast, RNAi-mediated depletion of NBS1 did not have a significant impact on MRE11 (fig. S4A), as reported previously (26). Next, RNAi-depleted cells were analyzed by ChIP-seq for the recruitment of each factor and RNAPII. expected, depletion of MRE11 led to a significant loss of MRE11 signal at genes highly bound at the TSS by MRE11, NBS1, or MRE11 + NBS1 (fig. S4B, left; compare blue boxes to gray boxes). Consistent with the reduction of NBS1 in extracts following knockdown of MRE11, NBS1 binding to chromatin was also significantly reduced under the same conditions (fig. S4B, right; compare blue boxes to gray boxes). Similarly, depletion of NBS1 reduced NBS1 binding at TSSs highly bound by NBS1, alone or together with MRE11 (right; compare red boxes to gray boxes). Depletion of NBS1 also diminished MRE11 signal at genes highly bound by MRE11, alone or together with NBS1 (left; red boxes 9 and 15), supporting the idea of cooperative binding to chromatin.

We next analyzed RNAPII association with genes in MRN-depleted and control cells. RNAPII accumulated over genes highly bound at the TSS by MRE11, NBS1, or both, following depletion of either MRE11 or NBS1 compared to a control knockdown (Fig. 5A). RNAPII binding was significantly increased at the TSS of genes highly bound by MRE11, NBS1, or both following loss of MRE11 or NBS1 [Fig. 5B (left and right, respectively) and fig. S5A] and across gene bodies (fig. S5B), which was shown by ChIP-qPCR at several representative genes (fig. S5C). Depletion of either NBS1 or MRE11 altered RNAPII association at a common set of genes, not only at genes where RNAPII levels were increased (fig. S5D, right matrix; compare deciles 1 on X and Y axes) but also at genes where RNAPII association was decreased (fig. S5D, right matrix; compare deciles 10 on X and Y axes; $P < 1 \times 10^{-20}$ by hypergeometric test). We next intersected the changes in RNAPII association with occupancy of MRE11 or NBS1. As shown in Fig. 5C, genes showing the highest increase in RNAPII ChIP-seq reads (deciles 1 on X and Y axes) were also highly associated with MRE11 (left) or NBS1 (right). Although depletion of NBS1 or MRE11 led to loss of RNAPII at a common set of genes (deciles 10 on X and Y axes), these, in contrast, were not significantly bound by either factor.

To determine whether the increase in RNAPII association with genes that occurred upon depletion of MRN reflected an increase in processive transcription, we determined the association of the elongating, serine-2 phosphorylated form of RNAPII (Ser²) at genes highly bound by MRE11 and/or NBS1. Ser² RNAPII accumulated over genes bound by MRE11, NBS1, or both, following depletion of either MRE11 or NBS1 compared to controls (Fig. 5D) in a significant manner (Fig. 5E). MRN was also associated with active enhancers (Fig. 3, C and D). Analysis of RNAPII ChIP-seq showed that, like at genes, loss of MRE11 or NBS1 led to an increase in RNAPII association at enhancer regions (fig. S5E). However, in contrast to genes, Ser² RNAPII ChIP-seq reads were not significantly increased upon depletion of the factors (P value < 0.05). Therefore, MRE11 and NBS1 have a global impact on RNAPII levels genome-wide, frequently at the same regions. Up-regulation of RNAPII occurred at MRN-bound regions, whereas down-regulation of RNAPII may be independent of MRN association. These data suggest that MRN modulates the transcriptional output of target genes.

MRN protects actively transcribed regions from genomic instability

Given the key role of MRN in genomic stability, we next assessed SNPs by whole-genome sequencing in cells exposed to depletion of MRE11 or NBS1 or a control depletion. We discarded common SNPs found in all samples since these are likely to be false positives because of the background genomic environment. Approximately 30,000 SNPs unique to shCon cells were identified, while nearly 100,000 SNPs were detected in MRE11- or NBS1-depleted cells (Fig. 6A). Within genes, the majority of SNPs localized to gene bodies (Fig. 6A), which may reflect the genomic distributions of MRE11 and NBS1 (Fig. 2C). A ranking test by gene set enrichment analysis (GSEA) showed that SNPs found in MRE11- or NBS1-depleted cells could be readily predicted by the binding of MRE11 or NBS1 at TSSs (Fig. 6B). Furthermore, in keeping with the strict association between RNAPII and MRN occupancy at genes, SNP frequency was also well predicted by binding of RNAPII at the TSS.

Association of the DSB marker, γ H2AX, with MRN-bound genes following loss of either MRE11 or NBS1 compared to control samples was also analyzed (Fig. 6C). γ H2AX was detected at higher levels following loss of either MRE11 or NBS1 compared to a control knockdown at genes bound by MRE11, NBS1, or both factors (Fig. 6C and fig. S6). To determine whether the SNPs detected corresponded to regions showing increased levels of γ H2AX, a GSEA was performed. As shown in Fig. 6D, SNPs found in MRE11- or NBS1-depleted cells could be predicted by increased detection of γ H2AX. Together, these data suggest that binding of MRN over gene bodies preserves genes from DNA damage-induced mutations detected as SNPs. This highlights the key function of the MRN complex in maintaining genomic stability at actively transcribed regions.

DISCUSSION

Recent data strongly point to a central role for the MRN complex in the resolution of transcription-associated DNA repair. However, mechanistic details about the recruitment of MRN to sites of transcription-associated DNA damage are not clear. Using the unbiased proteomic technique, PICH, we found that DNA repair factors were among the most recruited proteins that specifically interacted with the chromatin region during transcription of the inducible HIV-1 LTR. In particular, MDC1, which interacts with MRN complex, became highly associated with the actively transcribed locus and was largely absent before transcriptional stimulation. Furthermore, we identified the interactome of MDC1 in cells in the absence of exogenous DNA damage. Consistent with results obtained by PICH at the highly active HIV-1 minigene, MDC1 interacted with many factors implicated in transcription and cotranscriptional RNA processing, indicating an association between MDC1 and active transcription of cellular genes. A robust interaction between MDC1 and RNAPII, as well as P-TEFb, was detected in the absence of exogenous DNA damage, suggesting a constitutive association with transcription. ChIP-seq revealed widespread localization of endogenous MRE11 and NBS1 across genes. Association of both MRE11 and NBS1 with genes was strongly correlated with transcriptional activity. Our data further suggest that MRN association with active genes is dependent on the presence of the RNAPII transcriptional complex rather than the level of transcription per se.

While it is unclear precisely how MRN associates with the RNAPII complex, note that MDC1 interacts with topoisomerase II

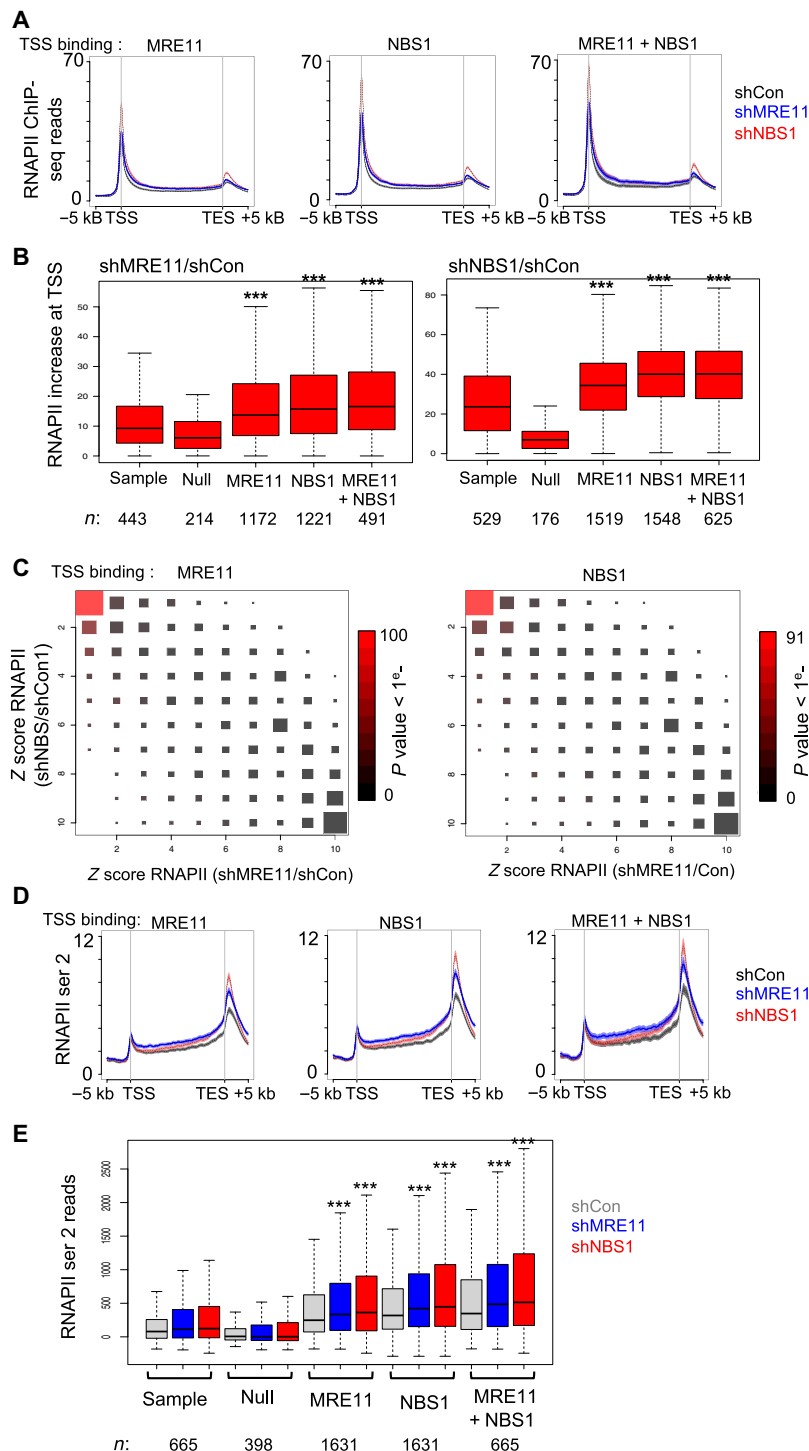


Fig. 5. Binding of MRN affects the transcriptional output of target genes. (A) Averaged profiles of RNAPII ChIP-seq reads across genes (± 5 kb) highly bound at the TSS by MRE11, NBS1, or both in samples treated with short hairpin RNA (shRNA) targeting MRE11, NBS1, or a nontargeting control. (B) Box plots of RNAPII increase at the TSS of genes highly bound at the TSS by MRE11, NBS1, both MRE11 and NBS1, or neither (null), compared to a sample group of genes, following knockdown of MRE11 or NBS1, relative to a control knockdown. ($***P < 0.001$, Wilcoxon test). (C) Hinton diagram showing changes in RNAPII ChIP-seq reads at genes following knockdown of MRE11 or NBS1, relative to a control knockdown, ranked by deciles from the most up-regulated (1) to the most down-regulated (10), at genes highly bound at the TSS by either MRE11 (left) or NBS1 (right). The scale represents P value for the intersection of RNAPII increase with TSS binding of MRE11 or NBS1 (10^{-0} to 10^{-100} and 10^{-0} to 10^{-91} , respectively, hypergeometric test). The size of the box represents the number of genes, as shown in fig. S5D. (D) Averaged profiles of phospho-Ser² RNAPII ChIP-seq reads across genes highly bound at the TSS by MRE11, NBS1, or both following knockdown of MRE11 or NBS1 or a control. (E) Box plots representing phospho-Ser² RNAPII ChIP-seq reads at the TES of genes for which the TSS was highly bound by MRE11, NBS1, both MRE11 and NBS1, or neither (null), compared to a sample group of genes, following knockdown of MRE11 or NBS1 or a control ($***P < 0.001$, Wilcoxon test).

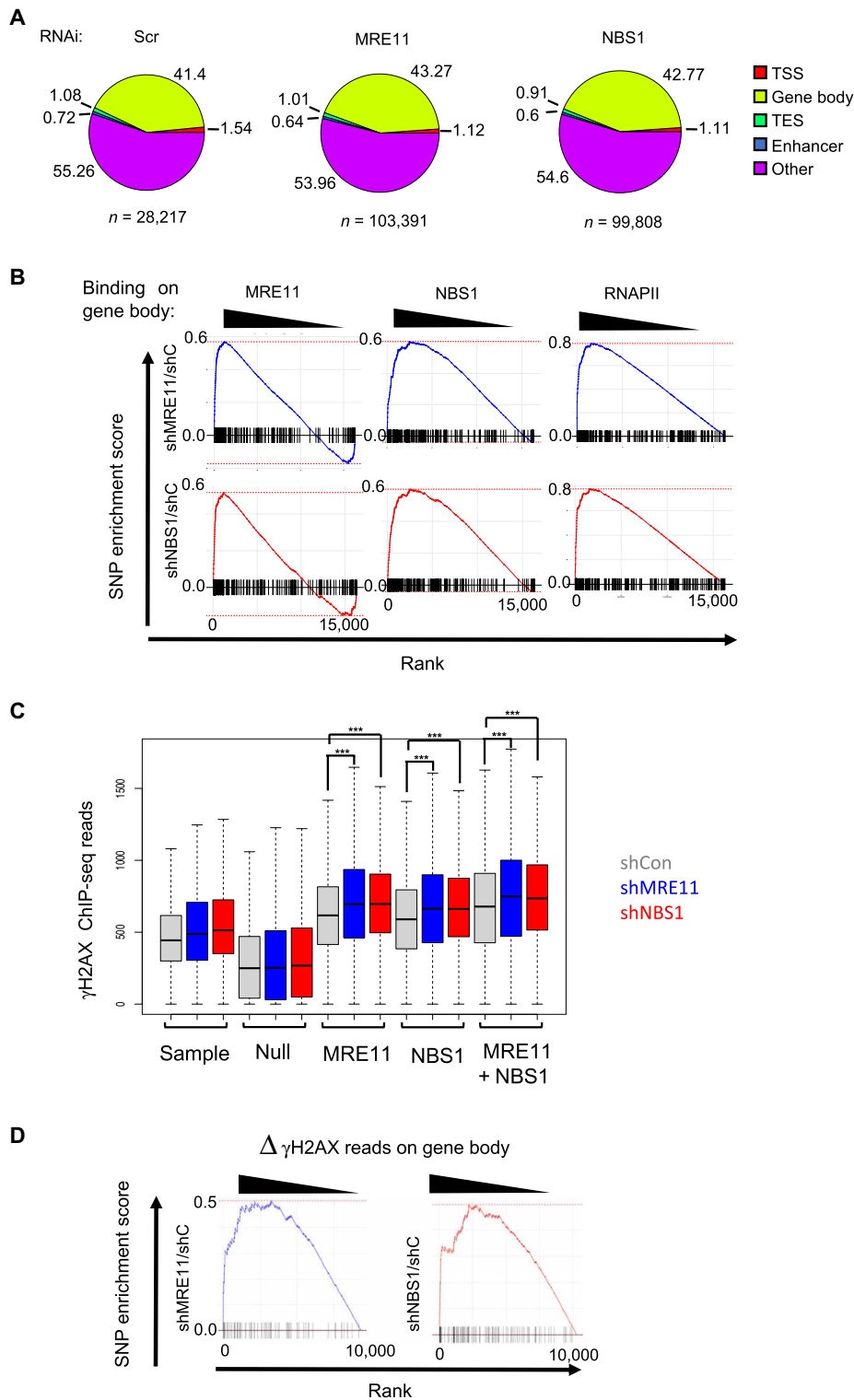


Fig. 6. MRN protects actively transcribed regions from genomic instability. (A) Pie charts showing the genomic distribution of SNPs detected by analyzing genomic DNA extracted from cells following depletion of MRE11, NBS1, or a control, as indicated. (B) GSEA of SNPs detected by analyzing genomic DNA extracted from cells following depletion of MRE11 and NBS1 compared to shCon sample, as indicated, ranked by ChIP-seq reads of MRE11, NBS1, or RNAPII in gene bodies, as indicated (all tests were significant; $P < 10^{-3}$, Wilcoxon test). (C) Box plots representing γ H2AX ChIP-seq read depth across the body of genes for which the TSS was highly bound by MRE11 or NBS1, both, or neither (null), compared to a sample group of genes, following knockdown of MRE11 or NBS1 or a control knockdown, as indicated ($***P < 0.001$, Wilcoxon test). (D) GSEA of SNPs detected by analyzing genomic DNA extracted from cells following depletion of MRE11 or NBS1 compared to a control depletion (shCon), ranked by the increase in γ H2AX ChIP-seq reads across the gene body under the same conditions (all tests were significant; $P < 10^{-3}$, Wilcoxon test).

(TOP2) through its C-terminal BRCT domains (26). TOP2 is a component of the RNAPII transcriptional complex and is required for transcription through chromatin (27). Whether MDC1 associates with the RNAPII complex through TOP2 or another transcription-associated factor, MRN, is found at the most highly transcribed genes, and its association with genes is highly correlated to that of RNAPII. Notably, upon depletion of either MRE11 or NBS1, MRN target genes accumulated DNA damage in an MRN- and RNAPII-dependent manner. Together, these data suggest that MRN is associated with the RNAPII transcription complex probably to scan for transcription-associated DNA damage and initiate repair.

NHEJ and HR pathways compete for the repair of DSBs (1). The degree of chromatin compaction is thought to influence repair pathway choice. DSBs occurring in open chromatin undergo end resection and are predominantly repaired by HR, which results in faithful repair and suppresses dangerous mutations from arising in coding regions of the genome (7, 9, 28). The preference for using HR to repair DSBs in transcribed regions is not due to cell cycle-dependent availability of factors. Increasing evidence indicates that DNA damage occurring on active genes is repaired through a specialized mechanism involving HR. The current model suggests that upon DNA damage, the damaged site is targeted by MRN complex and undergoes the initial steps in HR, which disfavors repair of the site by NHEJ. These MRN-marked sites subsequently pair and cluster into higher-order structures to be repaired flawlessly during G₂ phase of the cell cycle by the HR pathway. This mechanism preserves the fidelity of the genome in coding regions. While the preference for HR at actively transcribed is likely due to a combination of factors including nuclear position and the epigenetic landscape, we speculate that the physical link between MRN, which is required for commitment to HR, and RNAPII transcription complex might contribute to the preference for HR at transcribed regions. Being a component of the transcription complex, MRN may have immediate access to transcription-induced DSBs, facilitating end resection and commitment to HR.

Several lines of evidence indicate that the presence of DSBs in coding regions leads to transcriptional arrest. For example, transcriptional silencing occurs on chromatin in the vicinity of DSBs in an ATM (ataxia telangiectasia mutated)-dependent manner (29). Paradoxically, we observed that transcription was modestly increased in cells depleted of MRE11 or NBS1. There are several possible explanations for this apparently conflicting result. It has recently been shown that in response to ultraviolet irradiation, RNAPII is released from promoter-proximal regions into gene bodies to promote detection of DNA damage within genes (30). Thus, the increased RNAPII occupancy observed upon loss of MRN could also be a scanning mechanism to rapidly detect DNA damage within transcribed genes. Alternatively, the measured increase in transcription could be due to failure to establish HR at damaged sites in the absence of MRN. The use of HR to repair damage at active genes privileges genome fidelity at a slight cost to gene expression. Transcription of genes that have sustained DNA damage will be arrested until repair is completed in G₂. Therefore, at the global level, expression of the gene will be somewhat diminished, but DNA repair will be flawless. In cells depleted of MRN, which is required to commit to HR, DNA repair most likely occurs by NHEJ. In this case, the cost to transcription will be minimal, as NHEJ operates throughout the cell cycle. However, repair will likely occur at the expense of genome fidelity, as NHEJ repair is characterized by the appearance of SNPs. This is the outcome in cells depleted of MRN. RNAPII occupancy at both the TSS and gene body was higher

than in control cells. However, whole-genome sequencing revealed that DNA repair was highly error prone, with the appearance of thousands of SNPs in the gene bodies of target genes. Furthermore, the number of SNPs detected was highly correlated to the abundance of MRN at the gene under control conditions and the presence of γ H2AX in MRN-depleted conditions. This scenario implies that MRN does not directly influence transcription but that the absence of MRN predisposes to an alternative DNA repair pathway that is more favorable for gene expression but at the cost of genome fidelity.

MATERIALS AND METHODS

Cell culture, treatment, and lentiviral infection

HeLa cells were grown in Dulbecco's modified Eagle's minimal essential medium (DMEM) (Sigma-Aldrich, D6429), supplemented with 10% fetal calf serum (FCS; Eurobio Scientific, CVFSVF00-01) and containing 1% penicillin-streptomycin (Sigma-Aldrich, P4333). HEK-293T were grown in Hepes-modified DMEM (Sigma-Aldrich, D6171), supplemented with 10% FCS (Eurobio Scientific, CVFSVF00-01) and containing 1% penicillin-streptomycin (Sigma-Aldrich, P4333). All cells were grown in a humidified incubator at 37°C with 5% CO₂. Where indicated, cells were subjected to heat shock at 42°C for 1 hour, and the medium was replaced before fixation. Cells were treated with α -amanitin (Sigma-Aldrich, A2263) for 16 hours at 20 μ g/ml or with 150 μ M DRB (Sigma-Aldrich, D1916) or mock-treated with dimethyl sulfoxide where indicated for 0, 1, or 3 hours. After 3 hours of DRB treatment, cells were washed and incubated for 1 hour in fresh cell culture medium.

Production of short hairpin RNA (shRNA)-expressing lentiviral particles was performed as described previously (20) using plasmids expressing shRNAs targeting MRE11 (Sigma-Aldrich MISSION shRNA, TRCN000338391), NBS1 (Sigma-Aldrich MISSION shRNA, TRCN000288622), or a nontargeting control (obtained through Addgene, plasmid 1864), as shown in table S5. For knockdown experiments, HeLa cells were transduced with lentiviral particles and harvested 4 days later, as described previously (20).

Transactivation of the luciferase reporter gene was achieved by transfecting HeLa cells in 15-cm dishes with 8 μ g of pcDNA3 vector or pcDNA3-Tat.Flag using TransIT-X2 transfection reagent (Mirus, MIR 6000), according to the manufacturer's instructions. After 24 hours, cells were used for ChIP or Re-ChIP assays.

Antibodies

Antibodies used in this study are shown in table S3.

CRISPR-Cas9-mediated editing of endogenous MDC1 gene

A single guide RNA (sgRNA) targeting the MDC1 gene around the ATG translation start site was cloned in pSpCas9 (BB)-2A-green fluorescent protein plasmid (Addgene, no. 48138). The plasmid was then transfected into HEK-293T cells along with a single-stranded oligodeoxynucleotide (table S4) harboring the Flag-HA sequence flanked by homology sequences to MDC1 around the cleavage site. Single cells were isolated and amplified. HEK-293T clones expressing Flag-HA MDC1 were identified by PCR and confirmed by Western blot using anti-HA and anti-Flag antibodies.

Coimmunoprecipitation analysis

Coimmunoprecipitation was performed using nuclear extracts of HEK-293T cells. Cells were lysed in ice-cold hypotonic buffer [20 mM

tris (pH 7.6), 10 mM KCl, and 1.5 mM MgCl₂] supplemented with EDTA-free complete protease inhibitor mixture (Roche) for 15 min on ice. NP-40 was added at 0.5% final, and extracts were centrifuged 1 min at 14,000g/4°C. The pellet (nuclei) was resuspended in nucleosome buffer [20 mM tris (pH 7.6), 150 mM NaCl, 1.5 mM MgCl₂, 2.5 mM CaCl₂, and 0.5 μl of 100 mM phenylmethylsulfonyl fluoride] and incubated with micrococcal nuclease (2 × 10³ U/ml; New England Biolabs) for 2 hours at 4°C. Lysates were cleared by centrifugation at 14,000g/4°C for 10 min and diluted in immunoprecipitation buffer [50 mM tris (pH 7.6), 150 mM NaCl, and 1% NP-40] supplemented with protease inhibitors. Protein concentration was determined using the Bradford reagent (Bio-Rad). Immunoprecipitations were performed on 400 μg of protein extracts with the indicated antibodies and rotated overnight at 4°C. Protein G Dynabeads were washed three times in immunoprecipitation buffer, added to protein extracts/antibody solution, and incubated for 2 hours at 4°C. Immunoprecipitates were washed extensively with the immunoprecipitation buffer, resuspended in protein sample loading buffer, boiled for 5 min, and analyzed by Western blotting using the antibodies shown in table S3.

Proteomics of isolated chromatin fragments

Proteomics of isolated chromatin segments was performed as described previously (12) using U2OS cells expressing HIV-LTR-MS2 (13). Sequences of probes used are shown in table S4. Silver staining was performed according to the manufacturer's instruction (SilverQuest, Invitrogen). Mass spectrometry was performed at the Taplin Facility, Harvard University, Boston, MA.

MDC1 protein complex purification

MDC1 complexes were purified from nuclear extracts of HEK-293T cells stably expressing Flag-HA-MDC1 by two-step affinity chromatography, as described previously (20). Sequential Flag and HA immunoprecipitations were performed on equal amounts of proteins. Silver staining was performed according to the manufacturer's instructions (SilverQuest, Invitrogen). Mass spectrometry was performed at the Taplin Facility, Harvard University, Boston, MA.

Glycerol gradient sedimentation analysis

Separation of active and inactive P-TEFb complexes was performed as described previously (31). Briefly, glycerol gradients (10 to 30%) were formed by pipetting 2 ml of each of the glycerol fraction (10, 15, 20, 25, and 30% v/v) in buffer A [20 mM Hepes (pH 7.9), 0.3 M KCl, 0.2 mM EDTA, and 0.1% NP-40] into centrifugation tubes (Beckman, 331372). Gradients were formed by standing for 6 hours at 4°C. Cells were lysed in 0.5 ml of buffer A [10 mM tris-HCl (pH 7.4), 150 mM NaCl, 2 mM EDTA, 1% NP-40, and 0.1% protease inhibitor] for 30 min at 4°C. The lysates were centrifuged at 10,000g for 10 min, and the supernatants were loaded into tubes with the preformed glycerol gradients. Protein complexes were then fractionated by centrifugation in an SW 41 Ti rotor (Beckman) at 38,000 rpm for 21 hours. Fractions (0.5 ml) were collected, precipitated with trichloroacetic acid, and lastly analyzed by immunoblotting with the appropriate antibodies.

ChIP, library preparation, and sequencing

ChIP-seq (32) was performed from HeLa cells using the ChIP-IT High Sensitivity Kit from Active Motif (reference 53040) according

to the manufacturer's instructions. Sonication was performed using the Qsonica Q700 Sonicator with microtip of 1/8 inches (reference 4418) at 11% amplitude and 13 min of processing time (30-s "ON" and 30-s "OFF"). Each ChIP used 30 μg of chromatin along with 4 μg of antibody detecting MRE11, NBS1, RNAPII, or phospho-Ser² RNAPII (table S3). ChIP-seq libraries were constructed using the Next Gen DNA Library Kit (Active Motif, 53216 and 53264). Library quality was assessed using Agilent 2100 Bioanalyzer and Agilent High Sensitivity DNA assay. High-throughput sequencing was performed by sequence-by-synthesis technique using a NextSeq 500 (Illumina) at the Genom'ic Facility, Institut Cochin, Paris.

ChIP and Re-ChIP-qPCR

RNAPII, NBS1, MRE11, and MDC1 ChIP were performed using the iDeal ChIP-qPCR Kit (Diagenode, catalog no. C01010180) following the manufacturer's instructions. HeLa cells were sonicated using the Bioruptor Pico (Diagenode, catalog no. B01060001) for 8 cycles of 30-s ON and 30-s OFF at high-power setting. For RNAPII, MRE11, and MDC1 ChIP, 3 μg of antibody was used, while 1.5 μg of antibody was used for NBS1 ChIP. γH2AX ChIP was performed using the ChIP-IT Express Enzymatic Kit (Active Motif, catalog no. 53009), following the manufacturer's instructions. Chromatin was digested for 7 min, and 50 μg of chromatin and 3 μg of antibody were used for each γH2AX ChIP.

Re-ChIP was performed using the Re-ChIP-IT Kit (Active Motif, catalog no. 53016) according to the manufacturer's instruction. Chromatin for the Re-ChIP was prepared using the ChIP-IT High Sensitivity Kit (Active Motif, catalog no. 53040), as described above. For each Re-ChIP, 50 μg of chromatin and 3 μg of antibody were used. Antibody used and sequences of primers used for real-time qPCR analysis are shown in table S4.

Whole-genome sequencing

Genomic DNA was extracted from cells treated with lentiviral particles expressing shCon, shMRE11, or shNBS1 using the QIAGEN DNeasy Blood and Tissue Kit (reference 69504), according to the manufacturer's instructions. To avoid RNA contamination, extracts were treated with ribonuclease A (QIAGEN, 19101) according to the manufacturer's instructions. Equal amounts of DNA were used for library preparation. Whole-genome sequencing was performed by Novogene, Cambridge, UK.

Bioinformatic analysis

For PICh and mass spectrometry proteomics data analysis, uniquely mapping peptides were counted for each protein in each condition. Proteins whose abundance was greater than sevenfold than that in the control condition were analyzed using Gene Ontology enrichment analysis (33, 34), searching for enriched biological processes.

For analysis of ChIP-seq data, sequencing reads were first filtered, using `fastq_illumina_filter`, and quality control of filtered reads was performed using `FastQC`. Filtered reads were then aligned onto the HG38 genome (24) using the `Burrows-Wheeler Aligner` (35) with default parameters. The sorted BAM files generated by `SAMtools` (36) keeping only reads with a mapping quality at least 30 were then normalized by `DeepTools`' (37) `bamCoverage` function, with a bin size of 10 bp. RPGC normalization was applied, with an effective genome size of 2,913,022,398 bp according to `DeepTools`' user manual instructions. Files were then further normalized by subtracting an RPGC (reads per genomic content) normalized input

data file, using bigwigCompare. The input data used for normalization were generated by averaging four inputs from separate assays, so as to minimize variability and biases that can be introduced during input normalization.

From these normalized data files, peak calling was performed using NormR's enrichR function, searching for enrichment of each BAM file of ChIP-seq reads against the input BAM file using a false discovery rate correction. Genomic Ranges (38) was then used to determine overlap between the peak range and genomic features of interest, such as genes with a TSS and TES from GRCh38 and enhancers in HeLa S3 from ENCODE. Profile matrices were extracted from the normalized data files using DeepTools' computeMatrix using a bin size of 10 bp. For each gene, matrices for both TSS and TES were used, with 5-kb flanking on each side of the feature, as well as gene bodies, which were scaled to 4 kb in length and with 4-kb flanking before the TSS and 2-kb flanking after the TES. For enhancers, the enhancer body was scaled to 500 bp, with 1-kb flanking on either side. Using these profile matrices, quantification of normalized reads was calculated by summing the score of each appropriate bin for the feature. Unless indicated otherwise, this was from start to end for gene body and for enhancers and 500 bp before and after TSS and TES.

RNAPII binding variation between conditions was calculated using z scores, which were calculated as follows $Z \text{ score} = \frac{(KD - WT)}{(KD - WT)^2}$ (i.e., difference weighted by mean signal), which transforms the distribution of variations into a normal distribution, allowing for better statistical interpretation of the variations.

Scatter plots, box plots, violin plots, pie charts, and bar plots were created using either basic R plotting functions or ggplot2 functions (39). For box plots, the gene groups analyzed were "sample" (randomly chosen genes within the sample group), "null" (genes within the lowest decile of ChIP-seq reads for a given protein), MRE11 or NBS1 (genes within the highest decile of ChIP-seq reads for individual factor binding), and "MRE11 + NBS1" (genes within the highest decile of ChIP-seq reads for both factors). Average binding profiles of proteins across genomic features of interest were generated using the plotAverage function "seqplots" (40). Heatmaps on genomic features were created using genomation's (41) gridHeat function, using the profile matrices generated by DeepTools. GSEA was performed using fgsea as described previously (42), and all enrichment plots were created using the plotEnrichment function of the plotrix package. Decile matrices were created using color2D matplot from the plotrix package. Genes ranked by z score of RNAPII variation were split into deciles, for both shMRE11 versus shCon and shNBS1 versus shCon. Then, for each pair of deciles, the number of genes in the intersection of those two deciles was saved into a matrix, creating a 10 by 10 matrix of integers. The number of genes in each intersection was tested for significance by hypergeometric test, creating a 10 by 10 matrix of P values. The genes at each intersection were also tested by hypergeometric test for significant enrichment in either top MRE11 or NBS1 binding levels (decile 1 of genes ranked by MRE11 or NBS1 binding). The group of active genes was determined by first ranking all genes from highest to lowest expression, using HTSeq count (Bioconductor) run on RNA-seq data in HeLa cells described in (25). The top 50% of ranked genes were selected as "active genes." For analysis of whole-genome sequencing data, following processing by ANNOVAR (43) on the sequencing platform, SNP location was then overlapped with gene bodies using Genomic Ranges functions, which yields a quantification per gene.

Statistical analysis

Data presented as histograms are shown as means \pm SD. Comparison between two groups was analyzed by two-tailed Student's t test, and asterisks represented significance defined as $*P < 0.05$, $**P < 0.01$, or $***P < 0.001$. Other statistical methods are described above in the "Bioinformatic analysis" section.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/7/21/eabb2947/DC1>

[View/request a protocol for this paper from Bio-protocol.](#)

REFERENCES AND NOTES

1. D. Setiawati, D. Durocher, Shieldin—The protector of DNA ends. *EMBO Rep.* **20**, e47560 (2019).
2. A. Shibata, D. Moiani, A. S. Arvai, J. Perry, S. M. Harding, M. M. Genois, R. Maity, S. van Rossum-Fikkert, A. Kertokallio, F. Romoli, A. Ismail, E. Ismalaj, E. Petricci, M. J. Neale, R. G. Bristow, J. Y. Masson, C. Wyman, P. A. Jeggo, J. A. Tainer, DNA double-strand break repair pathway choice is directed by distinct MRE11 nuclease activities. *Mol. Cell* **53**, 7–18 (2014).
3. A. Janssen, G. A. Breuer, E. K. Brinkman, A. I. van der Meulen, S. V. Borden, B. van Steensel, R. S. Bindra, J. R. LaRocque, G. H. Karpen, A single double-strand break system reveals repair dynamics and mechanisms in heterochromatin and euchromatin. *Genes Dev.* **30**, 1645–1657 (2016).
4. C. Lemaître, A. Grabarz, K. Tsouroula, L. Andronov, A. Furst, T. Pankotai, V. Heyer, M. Rogier, K. M. Attwood, P. Kessler, G. Dellaire, B. Klaholz, B. Reina-San-Martin, E. Soutoglou, Nuclear position dictates DNA repair pathway choice. *Genes Dev.* **28**, 2450–2463 (2014).
5. S. X. Pfister, S. Ahrabi, L. P. Zalmas, S. Sarkar, F. Aymard, C. Z. Bachrati, T. Helleday, G. Legube, N. B. la Thangue, A. C. G. Porter, T. C. Humphrey, SETD2-dependent histone H3K36 trimethylation is required for homologous recombination repair and genome stability. *Cell Rep.* **7**, 2006–2018 (2014).
6. F. Aymard, B. Bugler, C. K. Schmidt, E. Guillou, P. Caron, S. Briois, J. S. Iacovoni, V. Daburon, K. M. Miller, S. P. Jackson, G. Legube, Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. *Nat. Struct. Mol. Biol.* **21**, 366–374 (2014).
7. T. Yasuhara, R. Kato, Y. Hagiwara, B. Shiotani, M. Yamauchi, S. Nakada, A. Shibata, K. Miyagawa, Human Rad52 promotes XPG-mediated R-loop processing to initiate transcription-associated homologous recombination repair. *Cell* **175**, 558–570.e11 (2018).
8. J. A. Aten, J. Stap, P. M. Krawczyk, C. van Oven, R. A. Hoebe, J. Essers, R. Kanaar, Dynamics of DNA double-strand breaks revealed by clustering of damaged chromosome domains. *Science* **303**, 92–95 (2004).
9. F. Aymard, M. Aguirrebengoa, E. Guillou, B. M. Javierre, B. Bugler, C. Arnould, V. Rocher, J. S. Iacovoni, A. Biernacka, M. Skrzypczak, K. Ginalska, M. Rowicka, P. Fraser, G. Legube, Genome-wide mapping of long-range contacts unveils clustering of DNA double-strand breaks at damaged active genes. *Nat. Struct. Mol. Biol.* **24**, 353–361 (2017).
10. P. M. Krawczyk, T. Borowski, J. Stap, T. Cijssouw, R. ten Cate, J. P. Medema, R. Kanaar, N. A. P. Franken, J. A. Aten, Chromatin mobility is increased at sites of DNA double-strand breaks. *J. Cell Sci.* **125**, 2127–2133 (2012).
11. V. Roukos, T. C. Voss, C. K. Schmidt, S. Lee, D. Wangsa, T. Misteli, Spatial dynamics of chromosome translocations in living cells. *Science* **341**, 660–664 (2013).
12. J. DeJardin, R. E. Kingston, Purification of proteins associated with specific genomic loci. *Cell* **136**, 175–186 (2009).
13. S. Boireau, P. Maiuri, E. Basyuk, M. de la Mata, A. Knezevich, B. Pradet-Balade, V. Bäcker, A. Kornblihtt, A. Marcello, E. Bertrand, The transcriptional cycle of HIV-1 in real-time and live cells. *J. Cell Biol.* **179**, 291–304 (2007).
14. J. F. Goodwin, K. E. Knudsen, Beyond DNA repair: DNA-PK function in cancer. *Cancer Discov.* **4**, 1126–1139 (2014).
15. C. A. Sartorius, G. S. Takimoto, J. K. Richer, L. Tung, K. B. Horwitz, Association of the Ku autoantigen/DNA-dependent protein kinase holoenzyme and poly(ADP-ribose) polymerase with the DNA binding domain of progesterone receptors. *J. Mol. Endocrinol.* **24**, 165–182 (2000).
16. R. F. Chun, O. J. Semmes, C. Neuveut, K. T. Jeang, Modulation of Sp1 phosphorylation by human immunodeficiency virus type 1 Tat. *J. Virol.* **72**, 2615–2629 (1998).
17. S. Tyagi, A. Ochem, M. Tyagi, DNA-dependent protein kinase interacts functionally with the RNA polymerase II complex recruited at the human immunodeficiency virus (HIV) long terminal repeat and plays an important role in HIV gene expression. *J. Gen. Virol.* **92**, 1710–1720 (2011).
18. S. M. Zhang, H. Zhang, T.-Y. Yang, T.-Y. Ying, P.-X. Yang, X.-D. Liu, S.-J. Tang, P.-K. Zhou, Interaction between HIV-1 Tat and DNA-PKcs modulates HIV transcription and class switch recombination. *Int. J. Biol. Sci.* **10**, 1138–1149 (2014).

19. W. Kaczmarek, S. A. Khan, Lupus autoantigen Ku protein binds HIV-1 TAR RNA in vitro. *Biochem. Biophys. Res. Commun.* **196**, 935–942 (1993).
20. X. Contreras, K. Salifou, G. Sanchez, M. Helmsmoortel, E. Beyne, L. Bluy, S. Pelletier, E. Roussel, S. Rouquier, R. Kiernan, Nuclear RNA surveillance complexes silence HIV-1 transcription. *PLoS Pathog.* **14**, e1006950 (2018).
21. D. Latreille, L. Bluy, M. Benkirane, R. E. Kiernan, Identification of histone 3 variant 2 interacting factors. *Nucleic Acids Res.* **42**, 3542–3550 (2014).
22. B. Sobhian, N. Laguet, A. Yatim, M. Nakamura, Y. Levy, R. Kiernan, M. Benkirane, HIV-1 Tat assembles a multifunctional transcription elongation complex and stably associates with the 75K snRNP. *Mol. Cell* **38**, 439–451 (2010).
23. F. Michellini, S. Pitschiaya, V. Vitelli, S. Sharma, U. Gioia, F. Pessina, M. Cabrini, Y. Wang, I. Capozzo, F. Iannelli, V. Matti, S. Francia, G. V. Shivashankar, N. G. Walter, F. d'Adda di Fagagna, Damage-induced lncRNAs control the DNA damage response through interaction with DDRNAs at individual double-strand breaks. *Nat. Cell Biol.* **19**, 1400–1411 (2017).
24. E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczy, R. Levine, P. McEwan, K. McKernan, J. Meldrum, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, Y. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Solinger, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, M. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Showlken, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J.-F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H.-M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, C. Raymond, N. Shimizu, K. Kawasaki, S. Minooshima, G. A. Evans, M. Athanasiou, R. Schultz, B. A. Roe, F. Fuen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. McCombie, M. de la Bastide, N. Dedhia, H. Blöcker, K. Hornischer, G. Nordisiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H.-C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kasprzyk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. Smit, E. Stupka, J. Szustakowski, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A. Patrino, M. J. Morgan, P. de Jong, J. J. Catanese, K. Osoegawa, H. Shizuya, S. Choi, Y. J. Chen, J. Szustakowski; International Human Genome Sequencing Consortium, Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
25. B. Stadelmayer, G. Micas, A. Gamot, P. Martin, N. Malirat, S. Koval, R. Raffel, B. Sobhian, D. Severac, S. Rialle, H. Parrinello, O. Cuvier, M. Benkirane, Integrator complex regulates NELF-mediated RNA polymerase II pause/release and processivity at coding genes. *Nat. Commun.* **5**, 5531 (2014).
26. Z. H. Zhong, W. Q. Jiang, A. J. Cesare, A. A. Neumann, R. Wadhwa, R. R. Reddel, Disruption of telomere maintenance by depletion of the MRE11/RAD50/NBS1 complex in cells that use alternative lengthening of telomeres. *J. Biol. Chem.* **282**, 29314–29322 (2007).
27. N. Mondal, J. D. Parvin, DNA topoisomerase IIalpha is required for RNA polymerase II transcription on chromatin templates. *Nature* **413**, 435–438 (2001).
28. W. T. Lu, B. R. Hawley, G. L. Skalka, R. A. Baldock, E. M. Smith, A. S. Bader, M. Malewicz, F. Z. Watts, A. Wilczynska, M. Bushell, Drosha drives the formation of DNA:RNA hybrids around DNA break sites to facilitate DNA repair. *Nat. Commun.* **9**, 532 (2018).
29. N. M. Shanbhag, I. U. Rafalska-Metcalf, C. Balane-Bolivar, S. M. Janicki, R. A. Greenberg, ATM-dependent chromatin changes silence transcription in cis to DNA double-strand breaks. *Cell* **141**, 970–981 (2010).
30. M. D. Lavigne, D. Konstantopoulos, K. Z. Ntakou-Zamplara, A. Liakos, M. Fousteri, Global unleashing of transcription elongation waves in response to genotoxic stress restricts somatic mutation rate. *Nat. Commun.* **8**, 2076 (2017).
31. X. Contreras, M. Barboric, T. Lenasi, B. M. Peterlin, HMBA releases P-TEFb from HEXIM1 and 75K snRNA via PI3K/Akt and activates HIV transcription. *PLoS Pathog.* **3**, 1459–1469 (2007).
32. A. Barski, S. Cuddapah, K. Cui, T. Y. Roh, D. E. Schones, Z. Wang, G. Wei, I. Chepelev, K. Zhao, High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).
33. M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, G. Sherlock, Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
34. The Gene Ontology Consortium, The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47**, D330–D338 (2019).
35. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
36. J. K. Bonfield, J. Marshall, P. Danecek, H. Li, V. Ohan, A. Whitwham, T. Keane, R. M. Davies, HTSlib: C library for reading/writing high-throughput sequencing data. *Gigascience* **10**, giab007 (2021).
37. F. Ramirez, D. P. Ryan, B. Grünig, V. Bhardwaj, F. Kilpert, A. S. Richter, S. Heyne, F. Dündar, T. Manke, deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
38. M. Lawrence, W. Huber, H. Pagès, P. Aboyoun, M. Carlson, R. Gentleman, M. T. Morgan, V. J. Carey, Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **9**, e1003118 (2013).
39. H. Wickham, *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2016).
40. P. Stempor, J. Ahringer, SeqPlots—Interactive software for exploratory data analyses, pattern discovery and visualization in genomics. *Wellcome Open Res.* **1**, 14 (2016).
41. A. Akalin, V. Franke, K. Vlahovicek, C. E. Mason, D. Schubele, Genomation: A toolkit to summarize, annotate and visualize genomic intervals. *Bioinformatics* **31**, 1127–1129 (2015).
42. A. Heurteau, C. Perrois, D. Depierre, O. Fosseprez, J. Humbert, S. Schaak, O. Cuvier, Insulator-based loops mediate the spreading of H3K27me3 over distant micro-domains repressing euchromatin genes. *Genome Biol.* **21**, 193 (2020).
43. K. Wang, M. Li, H. Hakonarson, ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).

Acknowledgments: We thank C. Dargemont and members of the Cuvier and Kiernan laboratories for helpful criticism, J. Hamroune and the Genom'ic platform for sequencing, A. Heurteau and Gen2i for help with bioinformatics, and L. Bluy for technical assistance. **Funding:** This work was supported by ARC, ANRS, MSD Avenir (Hide, Inflamm&Seq), and European Research Council (CoG RNAmedTGS) to R.K.; ARC (4027) to S.R.; and Fondation pour la Recherche Médicale (FRM DEQ20160334940) to O.C. **Author contributions:** K.S., P.B., G.G., M.H., V.M., C.F., X.C., and S.R. performed experiments and analyzed data. C.B. and D.D. analyzed data. R.K. and S.R. designed the experiments. O.C. designed data analysis. R.K. wrote the manuscript with input from all authors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. ChIP-seq data have been deposited at GEO (GSE 143591). Additional data related to this paper may be requested from the authors.

Submitted 14 February 2020

Accepted 31 March 2021

Published 21 May 2021

10.1126/sciadv.abb2947

Citation: K. Salifou, C. Burnard, P. Basavarajiah, G. Grasso, M. Helmsmoortel, V. Mac, D. Depierre, C. Franckhauser, E. Beyne, X. Contreras, J. Dejardin, S. Rouquier, O. Cuvier, R. Kiernan, Chromatin-associated MRN complex protects highly transcribing genes from genomic instability. *Sci. Adv.* **7**, eabb2947 (2021).