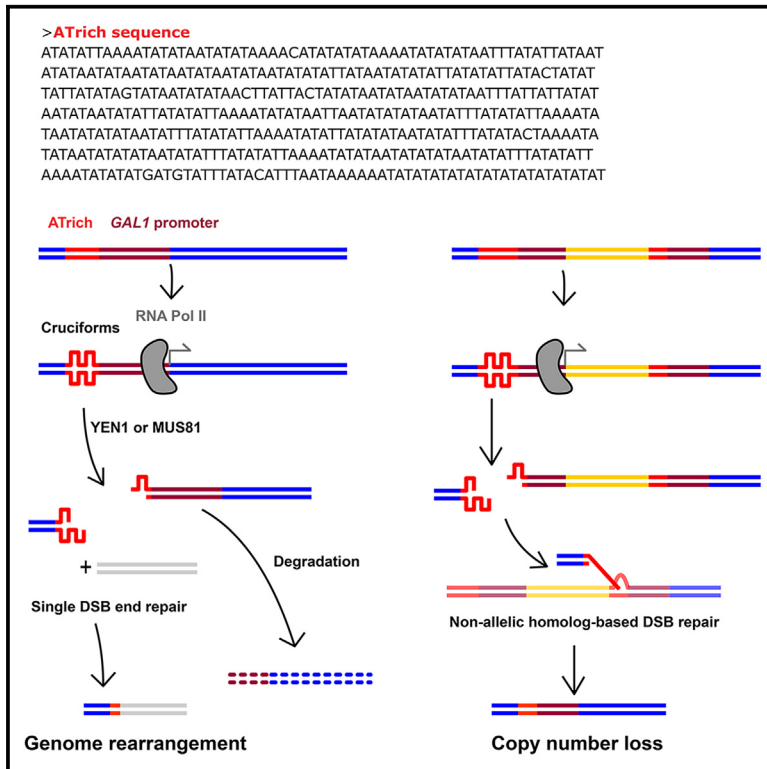


# Downstream transcription promotes human recurrent CNV associated AT-rich sequence mediated genome rearrangements in yeast

## Graphical abstract



## Authors

Feng-Yun Xie, Xiao-Guohui Zhang,  
 Juan Chen, ..., Shen Yin, Xiang-Hong Ou,  
 Jun-Yu Ma

## Correspondence

ouxianhong2003@163.com (X.-H.O.),  
 majy@gd2h.org.cn (J.-Y.M.)

## In brief

Properties of biomolecules; Techniques  
 in genetics; Molecular genetics; Model  
 organism

## Highlights

- Transcription exacerbates the fragility of upstream AT-rich sequence
- AT-rich sequences mediate copy number loss in yeast
- MUS81, YEN1, RAD1/10, RAD51, and Rnase H regulate the fragility of AT-rich
- DNA topoisomerase poisons increase AT-rich instability



## Article

# Downstream transcription promotes human recurrent CNV associated AT-rich sequence mediated genome rearrangements in yeast

Feng-Yun Xie,<sup>1,2</sup> Xiao-Guohui Zhang,<sup>3</sup> Juan Chen,<sup>1,2</sup> Xin Xu,<sup>2</sup> Sen Li,<sup>1,2</sup> Tian-Jin Xia,<sup>4</sup> Lei-Ning Chen,<sup>1,2</sup> Shen Yin,<sup>4</sup> Xiang-Hong Ou,<sup>1,2,\*</sup> and Jun-Yu Ma<sup>1,2,5,\*</sup>

<sup>1</sup>Guangzhou Municipal Key Laboratory of Metabolic Diseases and Reproductive Health, The Affiliated Guangdong Second Provincial General Hospital of Jinan University, Guangzhou, China

<sup>2</sup>Reproductive Medicine Center, The Affiliated Guangdong Second Provincial General Hospital of Jinan University, Guangzhou, China

<sup>3</sup>The Second School of Clinical Medicine, Southern Medical University, Guangzhou, China

<sup>4</sup>College of Life Sciences, Qingdao Agricultural University, Qingdao, China

<sup>5</sup>Lead contact

\*Correspondence: [ouxianghong2003@163.com](mailto:ouxianghong2003@163.com) (X.-H.O.), [majy@gd2h.org.cn](mailto:majy@gd2h.org.cn) (J.-Y.M.)

<https://doi.org/10.1016/j.isci.2024.111508>

## SUMMARY

AT-rich sequence can cause structure variants such as translocations and its instability can be accelerated by replication stresses. When human 16p11.2 or 22q11.2 recurrent copy number variant (reCNV) associated AT-rich sequence was inserted upstream *GAL1* promoter in yeast genome, we found that downstream transcription could promote AT-rich forming cruciform structure and mediate gross genome rearrangements. When genes were flanked with direct repeats containing AT-rich sequence, copy number loss of these genes would be stimulated. Transcription-mediated AT-rich instability can be alleviated by disrupting *MUS81* or *YEN1* and exacerbated by disrupting *RAD1/10*. Deletion of homologous recombination-associated genes can not only increase AT-rich fragility but also alter the breakpoint positions. AT-rich stability was also affected by DNA topoisomerase poisons. Our results reveal that transcription can promote AT-rich-mediated *de novo* genome rearrangement, which might be helpful for understanding the mechanism of reCNV formation in humans.

## INTRODUCTION

In both prokaryotes and eukaryotes, palindromic DNA sequences with inverted repeats can form specific cruciform-like structures. The cruciform extrusion of palindromic DNA is promoted by either DNA replication or transcription. At the early S phase, DNA replication can form approximately  $3 \times 10^5$  cruciforms in a single mammalian cell.<sup>1</sup> During transcription, promoter activation induces negative supercoiling of DNA, which can promote the transcription-mediated cruciform formation.<sup>2–4</sup> In mouse G2 phase-like oocytes, cruciform DNAs can be observed in growing oocytes which are transcriptionally active, but are collapsed in fully grown oocytes which are transcriptionally silent.<sup>5</sup> Evidence shows that palindromic DNA composed of AT base pairs is most vulnerable to cruciform formation, whereas palindromes with centric GC pairs or imperfect palindromes require more driving forces for cruciform extrusion.<sup>4</sup> Newly formed cruciform structures, in turn, may affect both DNA replication and transcription. Purified cruciform DNA is highly enriched in DNA fragments containing replication origins.<sup>6</sup> Depletion of the cruciform binding domain of 14-3-3 proteins would retard replication initiation in yeast.<sup>7</sup> Promoter upstream cruciforms can enhance promoter activity in prokaryotic systems,<sup>8</sup>

suggesting that cruciforms may play a role in transcriptional regulation. Although cruciform DNAs have been discovered for decades and are widespread in somatic and germline cells, their biological functions are still not well understood.

In addition to their beneficial effects, cruciform structures can also have some adverse effects on genome stability. Like other non-B DNA structures, such as G-quadruplexes and Z-DNA,<sup>9</sup> unsolved cruciforms during DNA transactions may induce DNA double-strand breaks (DSBs). Long palindromic AT-rich repeats (PATRR) have been found located at the breakpoint regions of recurrent chromosomal translocations in humans, such as t(8; 22)(q24; q11),<sup>10</sup> t(17; 22)(q11; q11),<sup>11</sup> and t(11; 22)(q23; q11).<sup>12</sup> These PATRR regions are enriched with DSB repair-associated protein markers such as RAD51 and  $\gamma$ H2A.X,<sup>13</sup> indicating that AT-rich indeed induces DSBs and may be repaired by RAD51-mediated homologous recombination.

Palindromic AT-rich sequences have been reported to be cleaved by structure-specific DNA endonucleases, resulting in the generation of DSBs. Study in a plasmid-based system suggested that GEN1 is responsible for the PATRR cruciform cleavage.<sup>14</sup> In microsatellite instability-associated cancer cells, the four-way junction-specific endonuclease MUS81 can cleave long AT-rich sequences when the RecQ DNA helicase gene



*WRN* is depleted.<sup>15</sup> As the formation of cruciforms is promoted by transcription-induced negative supercoiling,<sup>16</sup> cell specific gene expression might induce different cruciform profiles in different cells. AT-rich induced recurrent t(11; 22)(q23; q11) translocations are highly frequently formed in human sperm but couldn't be detected in lymphoblasts or fibroblasts,<sup>17</sup> which might be caused by testicular specific transcription of genes upstream or downstream of the PATRR sequences.<sup>5</sup> However, whether the fragilities of AT-rich sequences are aggravated by transcription hasn't been analyzed.

In addition to translocations, the integrity of the human genome is also disrupted by copy number variations (CNVs). CNVs can form sporadically or recurrently in the genome and are mainly associated with common complex human diseases such as neuropsychiatric disorders.<sup>18,19</sup> The formation of recurrent CNVs (reCNVs) is mediated by non-allelic homologous recombination (NAHR) repairing of DSBs at genomic segmental duplication regions.<sup>20</sup> In these reCNVs, 16p11.2 and 22q11.2 reCNVs are associated with diseases such as autism, epilepsy, intellectual disability, schizophrenia, and other brain-related disorders.<sup>21</sup> According to the UK Biobank,<sup>22</sup> these two reCNVs affect approximately 0.056% and 0.067% of people, respectively. Although the pathogenesis of 16p11.2 and 22q11.2 reCNVs has been comprehensively studied, mechanisms regarding the *de novo* generation of these reCNVs still lack investigation. It has been estimated that 90–95% of 22q11.2 deletions are *de novo* generated during reproduction and affect about 1 in 1000 fetuses.<sup>23</sup> For 16p11.2 reCNVs, *de novo* cases are found in 67.8% of deletion carriers and 25.0% of duplication carriers, respectively.<sup>24</sup> Deep investigation of the formation mechanisms of these reCNVs is critical for preventing birth defect associated disorders.

In this study, we found that AT-rich sequences were associated with CNV formation and located at the breakpoint regions of 16p11.2 and 22q11.2 reCNVs. Using the yeast system, we analyzed the fragility of the 16p11.2/22q11.2 reCNV associated AT-rich sequences and found that they could form cruciform structures, which made them unstable in yeast. The genome fragility of AT-rich sequence can be induced either by transcription-dependent or transcription-independent mechanisms, and is influenced by various structure-specific endonucleases and DSB repair proteins. These results are helpful for understanding the formation mechanism of 16p11.2 and 22q11.2 reCNVs in human germline cells.

## RESULTS

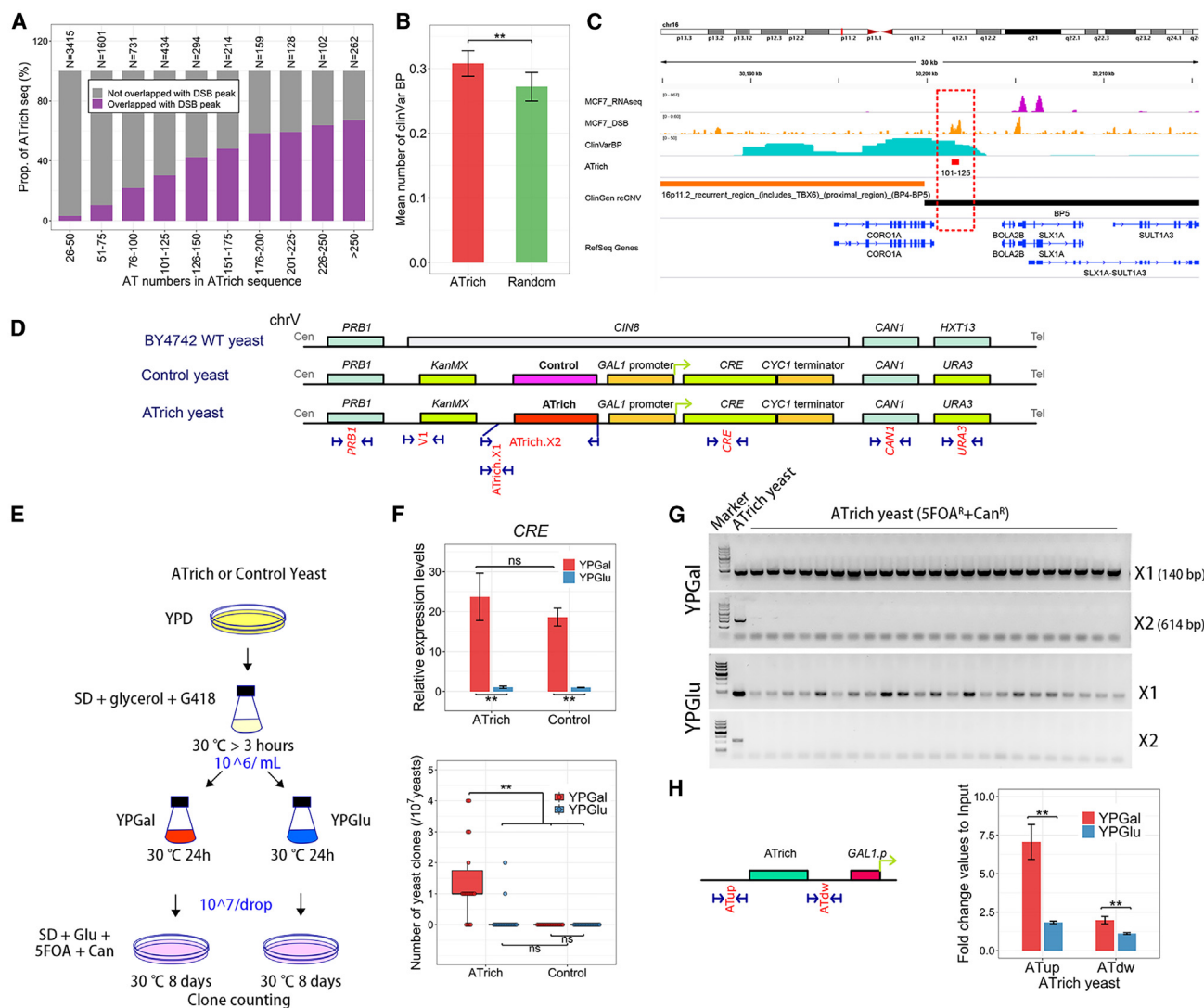
### Promoter upstream AT-rich sequences induce gross genome rearrangements

To analyze the genome stability of AT-rich sequences in the human genome, we extracted AT-rich sequences from the human genome (hg19) that meet three criteria: they contain at least one (AT)<sub>3</sub> sequence, the content of A and T bases is over 90% in a 100 bp DNA region, and they contain more than 25 AT units in the AT-rich sequence. As a result, 7340 AT-rich sequences (Table S3) were identified in human genome (2.37 AT-rich seq per 1 Mbp of DNA). To analyze whether the genome stability of AT-rich regions decreases with an increase in AT units, these

AT-rich sequences were divided into 10 groups based on the number of AT units (26–50, 51–75, 76–100, 101–125, 126–150, 151–175, 176–200, 201–225, 226–250, and >250). Then the number of AT-rich sequences that overlapped with DSB peaks which were detected in the normal MCF7 cell line<sup>25</sup> was counted. These DSB peaks which located within the region of 500 bp before and after AT-rich sequence were marked as AT-rich associated DSB peaks. As a result, the proportion of AT-rich sequences overlapped with DSB peaks increases with their AT unit numbers (Figure 1A), indicating that an increase in AT numbers in AT-rich sequences is detrimental to their stability.

Then, we analyzed whether AT-rich sequences are associated with the formation of CNVs in human genome. CNV breakpoints information was archived by ClinVar.<sup>26</sup> The number of ClinVar CNV breakpoints located within 2.5 Kbp regions flanking the AT-rich or randomly selected control sequences was counted. As a result, we found AT-rich sequences associated with higher number of ClinVar CNV breakpoints than that of their corresponding random sequence groups (mean value 0.31 for AT-rich group vs. 0.27 for random sequence group,  $p < 0.01$ , Wilcoxon test, Figure 1B), indicating that AT-rich sequences are more likely to cause structure variants like CNVs than random control sequences. Then the genes located within the 5 Kbp flanking DNA regions of AT-rich sequences were analyzed. We found that these AT-rich associated genes are enriched in the nervous system development pathway, regardless of whether these AT-rich sequences overlapped with ClinVar CNV breakpoints (Table S4). Then, we analyzed the AT-rich sequences that overlapped with DSB peaks in MCF7 cells and were located at the breakpoints of reCNVs. As a result, 29 AT-rich sequences were found within 8 breakpoint regions associated with reCNVs. These reCNVs included 10q22.3q23.2 (1 AT-rich region), 16p13.11 (3 AT-rich regions), 16p11.2 (1 AT-rich region), 22q11.21 (11 AT-rich regions), and 22q11.2 (24 AT-rich regions) reCNVs (Table S4).

In the breakpoint region of 16p11.2 reCNV, the AT-rich sequence (chr16:30201399–30201815) locates downstream of *CORA1A* and *BOLA2B* genes and upstream of *SLX1A* and *SLX1A-SULT1A3* genes (Figure 1C). To analyze whether the 16p11.2 reCNV associated AT-rich sequence could be a source of structural variant formation, we investigated its genome stability in a budding yeast system. In the *URA3* inactive BY4742 yeast, the *HXT13* gene was replaced with an active *URA3* gene. AT-rich and a control sequence (chr16:30200828–30201319) (Data S1) were then inserted into the *CIN8* gene region, respectively. Both AT-rich and control sequences were positioned upstream of a *GAL1* promoter, which controls the expression of an exogenous gene *CRE* (Figure 1D). Then, the AT-rich and Control yeasts that lost their chrV ends containing the *CAN1* gene, which makes yeast sensitive to Canavanine (Can), and the *URA3* gene, which makes yeast sensitive to 5-fluoro-orotic acid (5-FOA), could be selected using the 5-FOA + Can plates. To analyze the effects of transcription on AT-rich sequence stability, both AT-rich and Control yeasts were cultured in synthetic defined medium (SD) + 2% Glycerol + G418 for more than 3 h and 10<sup>6</sup> yeasts were cultured in 10 mL YPGal or YPGlu media for 24 h. Subsequently, water drops containing 10<sup>7</sup> yeasts were spotted on the 5-FOA + Can selection



**Figure 1. Transcription promotes upstream AT-rich instability in yeast**

(A) In the human MCF7 cell line, the stability of AT-rich decreases as the AT units increases. As the AT unit number increases in AT-rich sequences, the proportion of AT-rich sequences overlapping with DSB peaks also increased.

(B) Within the 2500 bp flanking regions, AT-rich associated with more ClinVar CNV breakpoints (ClinVar BP) than the randomly selected sequences. \*\*,  $p < 0.01$ , Wilcoxon test.

(C) An AT-rich sequence with 101–125 AT units (within the red dashed box) is located at the breakpoint 5 (BP5) region of the 16p11.2 recurrent CNV.

(D) The AT-rich sequence and a control sequence are inserted into the upstream region of the *GAL1* promoter in BY4742 yeast.

(E) Method for evaluating the gross genome rearrangement rate in yeasts. Yeasts that lose their chrV ends (containing *URA3* and *CAN1* genes) can exhibit resistance to 5-FOA and Can, forming double-resistant (RR) clones.

(F) The *CRE* gene expression is upregulated by the YPGal-induced *GAL1* promoter activation, and the *CRE* gene expression levels are comparable in AT-rich and control yeasts (upper). YPGal culturation increases the RR clone number in AT-rich yeast but not in Control yeast. RR clones can also be generated in YPGlu AT-rich yeasts, but their numbers are not significantly different from those in YPGlu Control yeasts. \*\*,  $p < 0.01$ ; ns, not significant; Student's *t* test.

(G) The breakpoints in both YPGal and YPGlu RR AT-rich clones are located at the AT-rich regions.

(H) YPGal culture promotes the cruciform structure formation of AT-rich sequence. \*\*,  $p < 0.01$ ; ns, not significant; Student's *t* test. DNA marker sizes, see STAR Methods.

plates. After 8 days of culturing, the RR yeast clone numbers were counted to evaluate the *de novo* gross genome rearrangement induced by the AT-rich or control sequence (Figure 1E). As a result, the *GAL1* promoter was activated in both the AT-rich and Control yeasts by YPGal (Figure 1F), but no RR clone was

found in the Control yeasts. There were only 3 clones found in 18 drops of YPGlu AT-rich yeasts, but an average of 1.39 clones per  $10^7$  yeasts ( $C/10^7$ ) was found in YPGal AT-rich yeasts, which was significantly higher than that in the other groups ( $p < 0.01$ , Figure 1F).

To analyze the breakpoint position where gross genome rearrangement was generated in AT-rich yeasts, we amplified the DNA fragment covering the AT-rich sequence (X2 in Figure 1D) and an adjacent upstream fragment of AT-rich sequence (X1 in Figure 1D). As a result, we identified the breakpoints of RR AT-rich yeast clones induced by YPGal and YPGlu, which were situated in the AT-rich regions (Figure 1G). Then, we analyzed whether the *GAL1* promoter mediated transcription promotes cruciform DNA formation at AT-rich regions. Using the cruciform DNA antibody-based immunoprecipitation-PCR method, we found that cruciform structures were highly enriched in the AT-rich regions in YPGal-cultured AT-rich yeasts (Figure 1H). This indicates that transcription promotes the extrusion of cruciform structures from upstream AT-rich regions.

With similar methods, we analyzed the genome stability of another human AT-rich sequence which locates at the breakpoint D (BP-D) region of the 22q11.2 reCNV. For this AT-rich sequence (chr22:21681177-21681911, 22q11.2 AT-rich), there is a noncoding gene *FAM230H* (NR\_136559.2) located at its 251 bp downstream region (Figure 2A; Table S4). This 22q11.2 AT-rich and its associated control sequence (chr22:21675551-21677460, 22q11.2 Control) were combined with the *GAL1* promoter and inserted into the yeast genome (Figure 2B; Data S2). When the 22q11.2 Control yeasts were cultured in YPGal or YPGlu for 24 h, we found that both YPGal and YPGlu could induce RR yeast clone formation, but there was no significant difference in RR clone number between these two groups (2.61 C/10<sup>7</sup> when cultured by YPGal and 2.00 C/10<sup>7</sup> when cultured by YPGlu, Figure 2C). However, for the 22q11.2 AT-rich yeast, the YPGal induced RR clone numbers were significantly larger than that induced by YPGlu (6.11 C/10<sup>7</sup> for YPGal cultured yeasts and 0.50 C/10<sup>7</sup> for YPGlu cultured yeasts,  $p < 0.01$ , Figure 2C), indicating YPGal induced *GAL1* promoter activation increased the fragility of 22q11.2 AT-rich sequence. Then the breakpoints of the 22q11.2 AT-rich and Control sequences induced RR yeasts were analyzed. For the 22q11.2 AT-rich yeasts, we found all the breakpoints in YPGlu and YPGal induced RR yeasts located at the AT-rich regions (23/23 RR clones in YPGlu group and 23/23 RR clones in YPGal group). For the 22q11.2 Control yeasts, we found all the breakpoints in YPGlu and YPGal induced RR yeasts located in the downstream regions of the 22q11.2 Control sequence (23/23 RR clones in YPGlu group and 23/23 RR clones in YPGal group, Figure 2D). Similar with 16p11.2 reCNV associated AT-rich sequence, we found *GAL1* promoter activation could also promote the cruciform structure formation in the 22q11.2 AT-rich region (Figure 2E).

### AT-rich sequences mediate the copy number loss in yeast

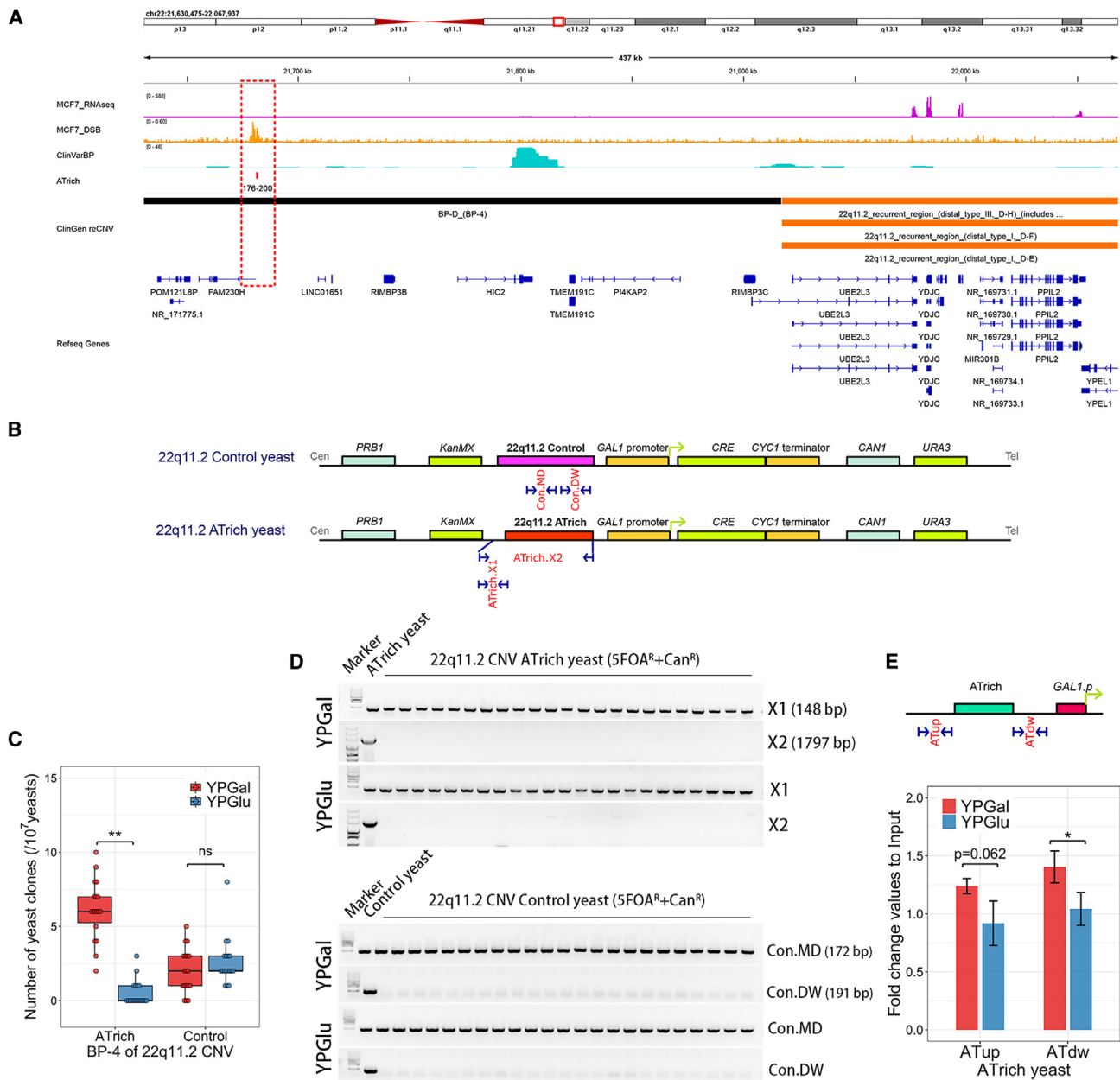
In human genome, reCNV formation are mostly mediated by the NAHR.<sup>27</sup> To analyze whether 16p11.2 reCNV associated AT-rich sequence (hereafter the AT-rich and Control sequences refer in particular to the 16p11.2 reCNV associated AT-rich and Control sequences respectively) could mediated CNV in yeast, AT-rich-*GAL1p-CRE-LEU2* and Control-*GAL1p-CRE-LEU2* cassettes were inserted into the AT-rich yeasts by replacing the *AVT2* gene on chrV (Figure 3A). The new yeast strains

were named as ATrich-ATrich yeast and ATrich-Control yeasts respectively. In the ATrich-ATrich yeasts, *CAN1* gene were surrounded by *KanMX-ATrich-GAL1p-CRE* and *ATrich-GAL1p-CRE-LEU2* cassettes. In the ATrich-Control yeasts, *CAN1* gene were surrounded by *KanMX-ATrich-GAL1p-CRE* and *Control-GAL1p-CRE-LEU2*. If AT-rich induced DSBs could be repaired by NAHR and caused the loss of *CAN1* gene, then the mutated yeasts could be selected by the Can plates. After cultured in SD ARG-/URA-/LEU-/G418 media overnight, 10<sup>6</sup> ATrich-ATrich and ATrich-Control yeasts were cultured in 10 mL YPGal or YPGlu media for 24h respectively, and then the Can resistant yeasts were selected by two kinds of plates: ARG- Can plates, which select the yeast whose *CAN1* is lost or mutated, regardless whether *LEU2* and *URA3* are lost; and the ARG-/URA-/LEU-/Can plates, which select the yeast whose *CAN1* is lost or mutated but *LEU2* and *URA3* are still exist (Figure 3B).

As a result, when ATrich-ATrich yeasts were selected by ARG-Can plates or ARG-/URA-/LEU-/Can plates, YPGal media significantly increased the Can resistant yeast clone number comparing with that caused by YPGlu (when yeasts were selected by ARG-/Can plates, 98.22 C/10<sup>4</sup> for YPGal group vs. 34.83 C/10<sup>4</sup> for YPGlu group,  $p < 0.01$ ; when yeasts were selected by ARG-/URA-/LEU-/Can plates, 101.33 C/10<sup>4</sup> for YPGal group vs. 37.89 C/10<sup>4</sup> for YPGlu group,  $p < 0.01$ ; Figure 3C). when ATrich-Control yeasts were selected by Can plates, YPGal media also significantly increased the Can resistant yeast clone number (when yeasts were selected by ARG-/Can plates, 7.61 C/10<sup>6</sup> for YPGal group vs. 2.11 C/10<sup>6</sup> for YPGlu group,  $p < 0.01$ ; when yeasts were selected by ARG-/URA-/LEU-/Can plates, 7.17 C/10<sup>6</sup> for YPGal group vs. 2.50 C/10<sup>6</sup> for YPGlu group,  $p < 0.01$ ; Figure 3C). For both ATrich-ATrich and ATrich-Control yeasts, when yeasts were selected by different Can plates (Figure 3B), the Can resistant clone numbers had no significant difference (Figure 3C).

Then we analyzed whether *CAN1* gene were lost and *URA3* gene still exist in the Can resistant yeast clones using PCR method. As a result, for all the Can resistant yeasts had been analyzed (23 clones for each group), the *CAN1* gene fragment couldn't be amplified but *URA3* gene fragments had been amplified (Figure 3D). To validate whether *CAN1* genes were lost in these Can resistant yeast clones, we amplified the genome region from *KanMX* to the *LEU2* (Figure 3E) from the YPGal and YPGlu induced Can resistant ATrich-ATrich and ATrich-Control yeasts (8 clones for each group). After single molecule sequencing, we found both YPGal and YPGlu induced Can resistant ATrich-ATrich yeasts had lost their *CAN1* gene by an NAHR manner,<sup>27</sup> indicating AT-rich sequence in yeast could induce copy number loss. For the YPGal induced Can resistant ATrich-Control yeasts, we found all 8 clones lost their *CAN1* genes and the Control sequences. For the YPGlu induced Can resistant ATrich-Control yeasts, we found one clone lost their *CAN1* gene and the Control sequence, but the other 7 clones lost their AT-rich sequences and *CAN1* genes (Figure 3E; Data S3). The exact mechanism for how *CAN1*-Control sequences were deleted in the ATrich-Control yeast is not known. It might be caused by NAHR-repair of DSBs occurring in the *GAL1p-CRE* region.





**Figure 2. Transcription promotes upstream 22q11.2 reCNV associated AT-rich instability**

(A) Location of 22q11.2 reCNV associated AT-rich sequence (within the red dashed box) in human genome.

(B) The 22q11.2 reCNV associated AT-rich sequence and control sequence are inserted into the upstream region of the *GAL1* promoter in BY4742 yeast.

(C) For the 22q11.2 AT-rich yeast, YPGal induced *GAL1* promoter activation increased the 22q11.2 AT-rich fragility. For the 22q11.2 control yeast, no significant difference of the 22q11.2 Control sequence fragility has been found between the YPGlu and YPGal group.

(D) For the mutated 22q11.2 AT-rich yeasts induced by either YPGlu or YPGal, the breakpoints locate in the AT-rich regions. For the mutated 22q11.2 Control yeasts, the breakpoints locate in the downstream region of the Control sequence.

(E) YPGal culturation promotes the cruciform structure formation of 22q11.2 associated AT-rich sequence. \*,  $p < 0.05$ ; \*\*,  $p < 0.01$ ; ns, not significant; Student's *t* test. DNA marker sizes, see STAR Methods.

In the ATrich-Control yeast, there is a short homologous sequence (green in Figure 3F) in the ATrich-*GAL1p*-*CRE* and Control-*GAL1p*-*CRE* cassettes, which might mediate the NAHR-mediated ATrich-*CAN1* loss in the YPGlu induced Can resistant ATrich-Control yeast (Figure 3G).

### Deletion of *RAD1*, *RAD2*, or *RAD10* promotes transcription-mediated AT-rich instability

It has been reported that AT-rich formed DNA structures could be cleaved by MUS81<sup>28</sup> and GEN1 (also referred as YEN1 in yeast).<sup>14</sup> MUS81 has been reported to be associated with



AT-rich genome instability induced by replication stress,<sup>28,29</sup> but it was unknown whether it causes the transcription-mediated AT-rich instability. To comprehensively analyze which DNA structure-specific endonucleases caused the transcription-mediated AT-rich instability, we deleted structure-specific endonuclease genes *MUS81*, *YEN1*, *RAD1* (also referred as *ERCC1* or *XPF*), *RAD2* (also referred as *ERCC5* or *XPG*), and *RAD27* (also referred as *FEN1*)<sup>30</sup> in the AT-rich yeast, and evaluated the stability of AT-rich regions. As a result, compared to normal AT-rich yeast (1.67 and 0.11 RR C/10<sup>7</sup> when cultured by YPGal and YPGlu respectively), the deletion of *RAD1*, *RAD2*, *YEN1*, or *MUS81* had no obvious effects on AT-rich stability when the *GAL1* promoter was suppressed by YPGlu. However, the RR clones were decreased in *yen1Δ* (0.94 C/10<sup>7</sup>,  $p < 0.05$ ) and *mus81Δ* (0.50 C/10<sup>7</sup>,  $p < 0.01$ ) AT-rich yeasts, but were increased in *rad1Δ* (12.33 C/10<sup>7</sup>,  $p < 0.01$ ) and *rad2Δ* (2.67 C/10<sup>7</sup>,  $p < 0.05$ ) AT-rich yeasts (Figure 4A). Using breakpoint analysis by PCR, we found that 22 out of 23 (22/23) RR *rad1Δ* clones were generated by DNA cleavage at the AT-rich region, while 1/23 clone was generated by DNA cleavage between the AT-rich region and the *CRE* gene body (Figures 4B and S1). It has been reported that *RAD1* participates in homologous recombination repair of DSB when resected 3' overhang contains AT-rich induced stem loops and other non-B structures.<sup>31</sup> The increase in RR clones in *rad1Δ* AT-rich yeasts might be caused by interference in the repair of DSBs generated at AT-rich regions.

When *RAD27* was deleted, we observed a significant increase in RR clones in both YPGal (115.89 C/10<sup>7</sup>,  $p < 0.01$ ) and YPGlu (27.72 C/10<sup>7</sup>,  $p < 0.01$ ) cultured yeasts compared to normal AT-rich yeasts (Figure 4A). However, we found that only 3/23 of RR *rad27Δ* clones were generated by AT-rich cleavage in the YPGal group, and no clone was generated by AT-rich cleavage in the YPGlu group (Figure 4B). In the YPGal *rad27Δ* RR clones, 1/23 clone was generated by DNA cleavage between AT-rich sequence and its upstream *KanMX* gene (Figure S1). In the YPGlu RR *rad27Δ* clones, we found that the breakpoints in 10/23 clones were located between the AT-rich sequence and *KanMX* gene (Figure S1). These data indicate that *RAD27* deletion had more unfavorable influences on the overall yeast genome stability, making it difficult to determine whether *rad27Δ* directly affects AT-rich stability.

To validate the effects of *MUS81* and *YEN1* on AT-rich stability, these two genes were also deleted in the *rad1Δ* AT-rich yeasts. After culturing in YPGal, we found that the mean RR clone numbers in *rad1Δ yen1Δ* yeasts (3.11 C/10<sup>7</sup>) and *rad1Δ mus81Δ* yeasts (1.17 C/10<sup>7</sup>) were significantly less than that in *rad1Δ* yeasts (10.72 C/10<sup>7</sup>,  $p < 0.01$ , Figure 4C). The number of clones in *rad1Δ mus81Δ* yeasts was significantly lower than

that in *rad1Δ yen1Δ* yeasts ( $p < 0.01$ , Figure 4C). These data confirm that both *YEN1* and *MUS81* participate in the cleavage of AT-rich formed DNA structures, and indicate that *MUS81* plays a more critical role in cleaving AT-rich structures than *YEN1* in yeasts.

It has been reported that *SLX1/4* and *RAD1/10* complexes participate in the *MUS81* cleavage of replication stress-induced AT-rich instability.<sup>28</sup> *SLX4* complexed with *RAD1/10* participates in the removal of 3' overhangs during single-strand annealing repair of DSB. It can also form a complex with *SLX1* as a co-activator to enhance the 5'-flap endonuclease activity of *SLX1*.<sup>32</sup> To analyze whether *SLX1* and *SLX4* are associated with transcription-mediated AT-rich instability, the *SLX1* and *SLX4* genes were deleted in normal AT-rich yeast and *rad1Δ* AT-rich yeast. After being cultured in YPGal, the RR clone numbers generated in *slx1Δ* (2.05 C/10<sup>7</sup>) and *slx4Δ* (1.94 C/10<sup>7</sup>) yeasts showed no significant difference compared to normal AT-rich yeasts (2.00 C/10<sup>7</sup>). Similarly, the RR clone numbers in *rad1Δ slx1Δ* (11.56 C/10<sup>7</sup>) and *rad1Δ slx4Δ* (10.56 C/10<sup>7</sup>) yeasts also did not significantly differ from *rad1Δ* AT-rich yeasts (12.28 C/10<sup>7</sup>, Figure 4D), indicating that *SLX1* and *SLX4* may not be involved in the transcription-mediated AT-rich instability.

After deleting of *RAD10* in normal AT-rich yeast and *rad1Δ* AT-rich yeast, we assessed the stability of AT-rich. As a result, deletion of *RAD10* also increases the mean RR clones in the YPGal group (6.78 C/10<sup>7</sup> vs. 1.06 C/10<sup>7</sup> in normal AT-rich yeasts,  $p < 0.01$ ). However, the mean clone numbers in *rad10Δ* yeasts were lower than those in *rad1Δ* yeasts (12.44 C/10<sup>7</sup>). There was no obvious effect on RR clone numbers in YPGlu-cultured yeasts by *RAD10* deletion (0.28 C/10<sup>7</sup> in *rad10Δ* AT-rich yeasts vs. 0.11 C/10<sup>7</sup> and 0.28 C/10<sup>7</sup> in normal and *rad1Δ* AT-rich yeasts, Figure 4E). When cultured in YPGal, RR clones generated in the *rad1Δ rad10Δ* AT-rich yeasts (12.61 C/10<sup>7</sup>) showed no significant difference compared to those in *rad1Δ* AT-rich yeasts but were significantly higher than in *rad10Δ* AT-rich yeasts ( $p < 0.01$ ). When cultured in YPGlu, RR clones generated in the *rad1Δ rad10Δ* AT-rich yeast showed a slight increase (1.17 C/10<sup>7</sup>) compared to that in *rad1Δ*, *rad10Δ*, and normal AT-rich yeasts ( $p < 0.01$ , Figure 4E).

#### Homologous recombination participates in transcription-mediated AT-rich DSB repair

As the *RAD1/10* complex might attenuate AT-rich sequence instability by promoting homologous recombination repair of the AT-rich-generated DSB,<sup>31</sup> we analyzed whether the homologous recombination associated factor RNase H could increase AT-rich instability. When *RNH1* and *RNH201* were double deleted in normal AT-rich yeast and *rad1Δ* AT-rich yeast, the methyl methanesulfonate (MMS) resistance capability, which

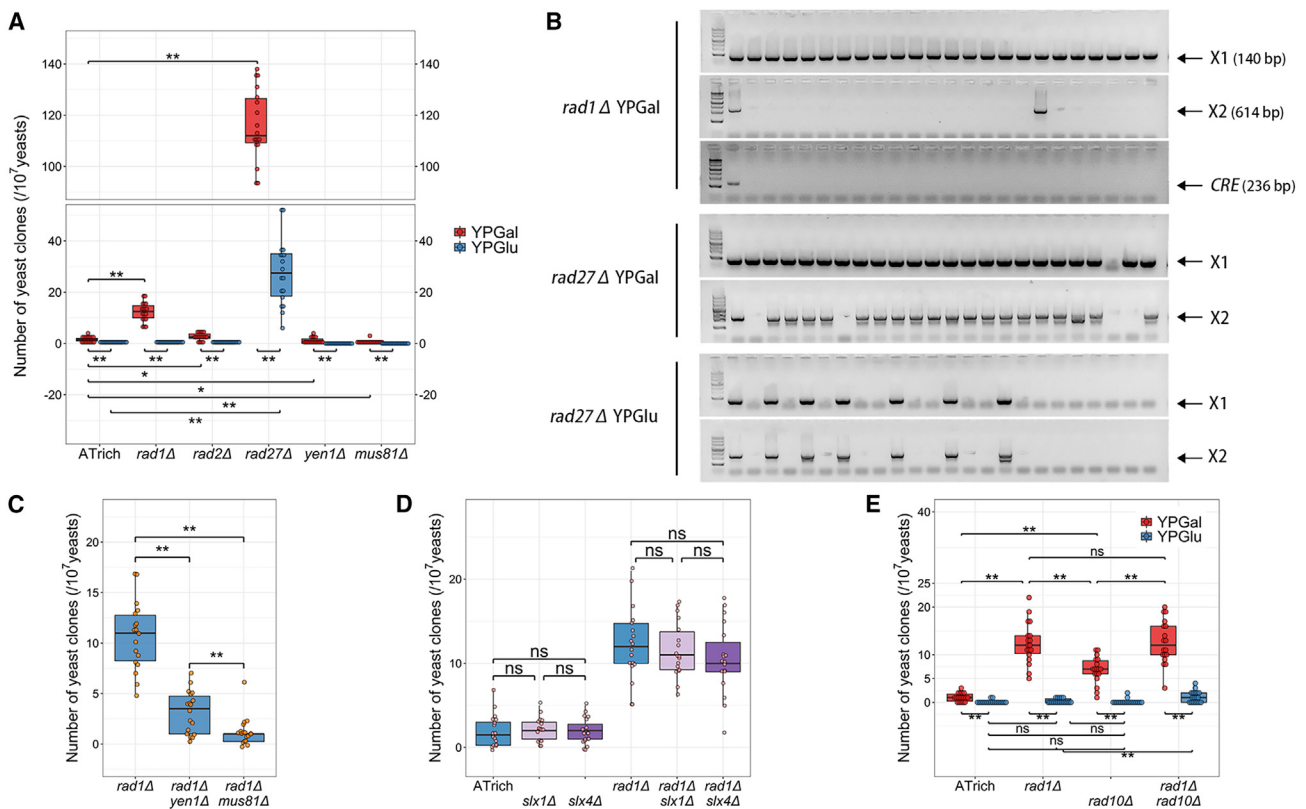
(D) Detection of whether *CAN1* gene is deleted in the Can resistant yeast clones by PCR method.

(E) PacBio HiFi sequencing of the mutated regions of Can resistant yeasts. For each yeast group, 8 clones are chosen for HiFi sequencing. For ATrich-ATrich yeasts, both YPGlu and YPGal induced Can resistant clones (8 in 8 clones) lost their genome regions containing AT-rich sequence, *CRE* and *CAN1* genes. For ATrich-Control yeasts, 8 in 8 YPGal induced Can resistant clones and 1 in 8 YPGlu induced Can resistant clone lost their genome regions containing *CRE*, *CAN1* and Control sequence; 7 in 8 YPGlu induced Can resistant clone lost their genome regions containing AT-rich sequence, *CRE* and *CAN1*.

(F) The genome elements of ATrich-Control yeast.

(G) The HiFi sequencing results of YPGlu induced Can resistant ATrich-Control yeast clones which lost the AT-rich sequence. \*\*,  $p < 0.01$ ; ns, not significant; Student's *t* test. DNA marker sizes, see STAR Methods.





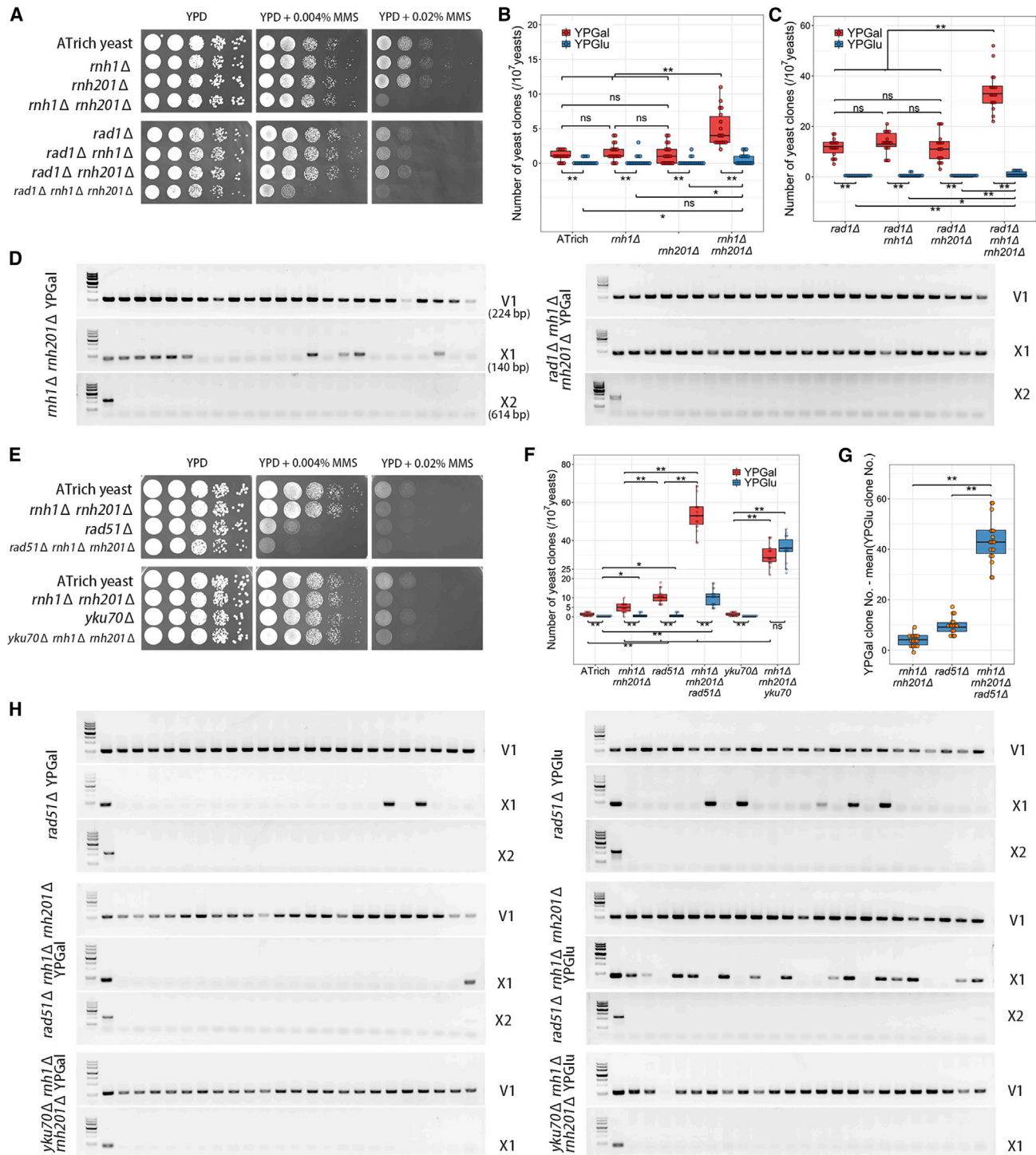
**Figure 4. RAD1/10 complex plays roles in AT-rich stability**

(A) Disruption of *RAD1* or *RAD2* can significantly increase the YPGal-induced RR clone numbers, whereas disruption of *MUS81* or *YEN1* decreased the RR clone numbers. *RAD27* disruption increased RR clone numbers in both YPGal and YPGlu cultured yeasts. (B) The breakpoints in *rad1Δ* RR clones are mostly (22 out of 23 clones) located in AT-rich regions, whereas in *rad27Δ* YPGal and YPGlu RR yeasts, the number of breakpoints located in AT-rich regions is 3 and 0, respectively (Figure S1). First lane: DNA marker; second lane: normal AT-rich yeasts; third and other lanes: RR yeasts. (C) Disruption of *YEN1* or *MUS81* in *rad1Δ* yeast decreases the YPGal-induced RR clone numbers. (D) Disruption of *SLX1* or *SLX4* has no obvious effects on the AT-rich stability in YPGal culturation conditions. (E) YPGal-cultured *rad10Δ* yeasts can generate more RR clones than normal AT-rich yeasts but fewer than *rad1Δ* yeasts. The YPGal-induced RR clone numbers are comparable between *rad1Δ* yeasts and *rad1Δ rad10Δ* yeasts. \*,  $p < 0.05$ ; \*\*,  $p < 0.01$ ; ns, not significant; Student's *t* test. DNA marker sizes, see STAR Methods.

mostly represents the efficiency of homologous recombination repair,<sup>33,34</sup> was decreased in these yeasts (Figure 5A). In addition, *RNH1* and *RNH201* double deletion also slightly increased the sensitivity of AT-rich yeasts to DSB inducer Zeiocin, but it hadn't obviously increase yeast sensitivity to  $H_2O_2$ . No obvious difference of yeast sensitivity to  $H_2O_2$  and Zeiocin had been found between *RAD1* deleted and normal AT-rich yeast. But deletion of *RAD1* in *rnh1Δ rnh201Δ* AT-rich yeasts slightly increased the yeast sensitivity to  $H_2O_2$  (Figure S2). When cultured in YPGal, *rnh1Δ rnh201Δ* AT-rich yeasts generated more RR clones ( $4.94 C/10^7$ ) than *rnh1Δ* ( $1.61 C/10^7$ ), *rnh201Δ* ( $1.33 C/10^7$ ), and normal ( $1.06 C/10^7$ ) AT-rich yeasts ( $p < 0.01$ , Figure 5B). When cultured in YPGlu, *rnh1Δ rnh201Δ* AT-rich yeasts also generated slightly more RR clones ( $0.61 C/10^7$ ,  $p < 0.05$ ) than *rnh201Δ* ( $0.22 C/10^7$ ) and normal ( $0.22 C/10^7$ ) AT-rich yeasts, but not more than that in *rnh1Δ* ( $0.33 C/10^7$ ) AT-rich yeasts (Figure 5B). Similarly, *rad1Δ rnh1Δ rnh201Δ* triple disrupted AT-rich yeasts generated significantly more RR clones

( $34.06 C/10^7$ ) than *rad1Δ rnh1Δ* ( $13.44 C/10^7$ ), *rad1Δ rnh201Δ* ( $11.56 C/10^7$ ), and *rad1Δ* ( $11.61 C/10^7$ ) AT-rich yeasts ( $p < 0.01$ , Figure 5B) when cultured in YPGal. When cultured in YPGlu, *rad1Δ rnh1Δ rnh201Δ* AT-rich yeasts produced more RR clones ( $1.00 C/10^7$ ) than *rad1Δ rnh1Δ* ( $0.39 C/10^7$ ,  $p < 0.05$ ), *rad1Δ rnh201Δ* ( $0.22 C/10^7$ ,  $p < 0.01$ ), and *rad1Δ* ( $0.11 C/10^7$ ,  $p < 0.01$ ) AT-rich yeasts (Figure 5C).

The breakpoints of RR AT-rich yeasts resulting from the double deletion of *RNH1/201* and the triple deletion of *RNH1/RNH201/RAD1* were analyzed using PCR. As a result, we found breakpoints in 9/23 *rnh1Δ rnh201Δ* RR clones located at AT-rich sequences, whereas in 14/23 clones the breakpoints were located between AT-rich sequence and the *KanMX* gene (Figure 5D). However, all breakpoints in YPGal-induced *rad1Δ rnh1Δ rnh201Δ* RR clones were located at the AT-rich region (Figure 5D). Similar to the breakpoint positions in YPGal-induced RR clones, in the YPGlu-induced *rnh1Δ rnh201Δ* RR clones, 11/23 breakpoints were located in AT-rich regions, and 12/23



**Figure 5. Homologous recombination proteins are involved in the repair of transcription-mediated DSB in AT-rich**

(A) The sensitivity of AT-rich yeasts to methyl methanesulfonate (MMS) is increased by double deletion of *RNH1/201* or single deletion of *RAD1*. Triple deletion of *RAD1* and *RNH1/201* further increases the MMS sensitivity of AT-rich yeasts.

(B and C) *RNH1/201* double deletion significantly increases the number of RR clones in normal AT-rich yeasts and *rad1Δ* AT-rich yeasts cultured in YPGal.

(D) Breakpoint identification of YPGal-induced *rh1Δ rh201Δ* RR yeasts and *rh1Δ rh201Δ rad1Δ* RR yeasts.

(E) Deletion of *RAD51*, but not *YKU70*, increases the sensitivity of normal AT-rich yeasts and *RNH1/201* double-deleted AT-rich yeasts to MMS.

(F) The effects of *RAD51* or *YKU70* deletion on the gross genome rearrangement rates of normal and *rh1Δ rh201Δ* AT-rich yeasts.

(legend continued on next page)

breakpoints were situated between the AT-rich sequence and the *KanMX* gene. For YPGlu-induced *rad1Δ rnh1Δ rnh201Δ* RR clones, 20/23 breakpoints were located in AT-rich regions, 1/23 located between AT-rich sequence and *CRE* gene, and 2/23 located downstream of the *CRE* gene (Figure S1).

To further study the repair mechanism of AT-rich-induced DSBs, we deleted the critical homologous recombination repair factor gene *RAD51* and the non-homologous end-joining factor gene *YKU70* in normal and *rnh1Δ rnh201Δ* AT-rich yeast. As a result, deletion of *RAD51* but not *YKU70*, increased the sensitivity of normal and *rnh1Δ rnh201Δ* AT-rich yeasts to MMS (Figure 5E). In addition, deletion of *RAD51* obviously increased the sensitivity of normal AT-rich and *rnh1Δ rnh201Δ* AT-rich yeasts to Zeocin, whereas deletion of *YKU70* had no obvious effect of yeast sensitivity to either H<sub>2</sub>O<sub>2</sub> or Zeocin (Figure S2).

When cultured in YPGal, the RR clone numbers significantly increased progressively in *rnh1Δ rnh201Δ* (4.83 C/10<sup>7</sup>), *rad51Δ* (10.39 C/10<sup>7</sup>), and *rad51Δ rnh1Δ rnh201Δ* (53.39 C/10<sup>7</sup>) AT-rich yeasts, compared to normal AT-rich yeast (1.39 C/10<sup>7</sup>,  $p < 0.01$  for every comparison, Figure 5F). When cultured in YPGlu, *rad51Δ rnh1Δ rnh201Δ* AT-rich yeasts generated significantly more RR clones (10.17 C/10<sup>7</sup>) than normal AT-rich yeasts (0.22 C/10<sup>7</sup>). However, *rad51Δ* (0.83 C/10<sup>7</sup>) and *rnh1Δ rnh201Δ* (0.89 C/10<sup>7</sup>) AT-rich yeasts only produced slightly more RR clones than normal AT-rich yeasts (Figure 5F). Then we compared the numbers of RR clones generated by transcription alone between *rad51Δ* single disrupted, *rnh1Δ rnh201Δ* double disrupted, and *rad51Δ rnh1Δ rnh201Δ* triple disrupted AT-rich yeasts by subtracting the mean value of RR clone numbers caused by YPGlu from the RR clone numbers caused by YPGal. As a result, transcription alone induced RR clones were significantly more abundant in *rad51Δ rnh1Δ rnh201Δ* AT-rich yeasts than in *rad51Δ* or *rnh1Δ rnh201Δ* AT-rich yeasts (Figure 5G).

Unlike *RAD51* deletion, we found that *YKU70* single deletion did not affect the AT-rich stability in both YPGlu (0.06 C/10<sup>7</sup>) and YPGal (1.44 C/10<sup>7</sup>) cultured yeasts compared to normal AT-rich yeasts. However, when *YKU70* was deleted in *rnh1Δ rnh201Δ* AT-rich yeasts, the RR clone numbers significantly increased compared to *yku70Δ* AT-rich yeasts in both YPGlu (35.39 C/10<sup>7</sup>,  $p < 0.01$ ) and YPGal (32.50 C/10<sup>7</sup>,  $p < 0.01$ ) cultured yeasts (Figure 5F). In addition, no significant difference in *yku70Δ rnh1Δ rnh201Δ* RR clone numbers was found between the YPGlu and YPGal groups (Figure 5F).

In YPGal-generated *rad51Δ* RR clones, breakpoints of 21/23 clones were located between AT-rich sequence and *KanMX* gene, while 2/23 were located at the AT-rich region. In YPGlu-generated *rad51Δ* RR clones, 18/23 were located between AT-rich sequence and *KanMX* gene, while 5/23 were located in the AT-rich region, with no significant difference compared to the YPGal clones ( $p > 0.05$ , Fisher's exact test). However, for the YPGal-generated *rad51Δ rnh1Δ rnh201Δ* RR yeasts, the number of clones with breakpoints located between AT-rich sequence and *KanMX*

or located in the AT-rich region was 22/23 and 1/23, respectively. In contrast, the number of clones for YPGlu-generated *rad51Δ rnh1Δ rnh201Δ* RR yeasts changed significantly to 9/22 and 14/22, respectively ( $p < 0.01$ , Fisher's exact test, Figure 5H). For the YPGal and YPGlu generated *yku70Δ rnh1Δ rnh201Δ* RR yeasts, the breakpoints were mostly located between AT-rich sequence and *KanMX* gene (23/23 for the YPGal group and 22/23 for the YPGlu group,  $p > 0.05$ , Fisher's exact test, Figure 5H).

### The effects of topoisomerase poisons on the stability of AT-rich sequences

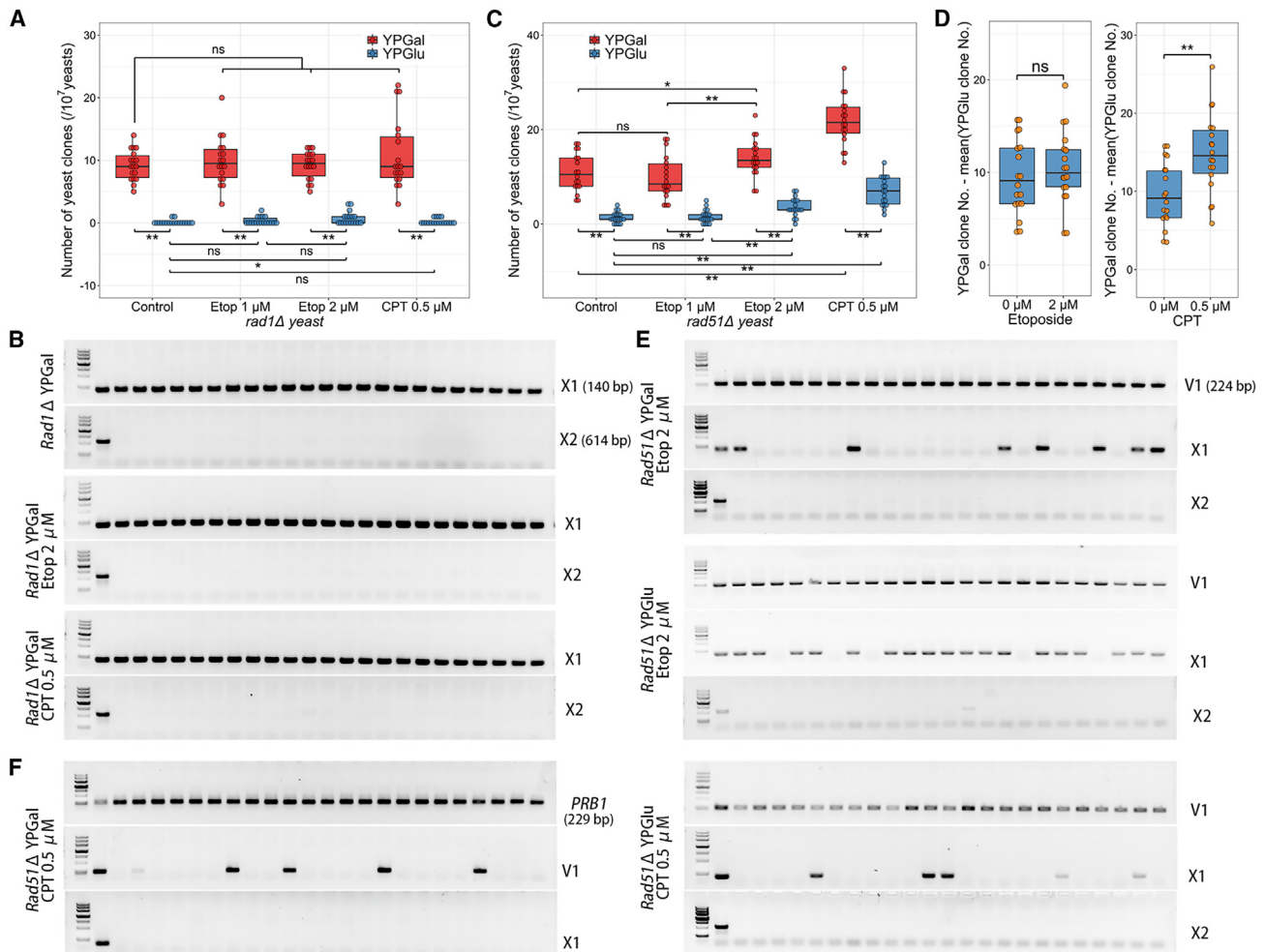
It has been reported that TOP2 plays critical roles in the formation of negative supercoiling at gene boundaries<sup>16</sup> and contributes to the specific DNA structure-mediated DSB formation.<sup>35</sup> Unlike TOP2, although TOP1 also regulates DNA transcription, it mainly function in removing positive supercoils generated ahead of RNA polymerase II.<sup>36</sup> To analyze whether RR clone numbers generated in AT-rich yeasts were associated with DNA topoisomerases, we treated the *rad1Δ* and *rad51Δ* AT-rich yeasts with a low dose of the TOP2 inhibitor Etoposide (Etop, 1 μM or 2 μM) or the TOP1 inhibitor Camptothecin (CPT, 0.5 μM).

Comparing with control *rad1Δ* AT-rich yeasts, when treated with a low dose of Etop or CPT for 24 h in YPGal, there was no significant change in the RR clone number in each group. When treated with Etop or CPT in YPGlu, only the RR clone number in 2 μM Etop treatment group was slightly higher (0.72 C/10<sup>7</sup>) than that in the control group (0.11 C/10<sup>7</sup>,  $p < 0.05$ , Figure 6A). All the breakpoints (23/23) of *rad1Δ* RR clones generated in the control group, 2 μM Etop treatment, and CPT treatment groups in YPGal were found to be located at the AT-rich regions (Figure 6B), indicating that Etop and CPT hadn't affected the mutation features in *rad1Δ* RR yeasts.

Unlike the results in *rad1Δ* AT-rich yeasts, for the *rad51Δ* AT-rich yeasts, we found that 2 μM Etop treatment could significantly increase the RR clone numbers in both YPGal (14.00 C/10<sup>7</sup>) and YPGlu (3.56 C/10<sup>7</sup>) groups when compared to that in the control yeasts (10.94 C/10<sup>7</sup> in YPGal,  $p < 0.05$ ; 1.39 C/10<sup>7</sup> in YPGlu,  $p < 0.01$ ) or 1 μM Etop-treated yeasts (9.89 C/10<sup>7</sup> in YPGal,  $p < 0.01$ ; 1.61 C/10<sup>7</sup> in YPGlu,  $p < 0.01$ , Figure 6C). However, after deducing the mean value of RR clone numbers induced by YPGlu, we found no significant difference in the transcription-mediated RR clone numbers between 2 μM Etop-treated and control *rad51Δ* AT-rich yeasts (Figure 6D). When compared to the control *rad51Δ* AT-rich yeasts, we observed an increase in RR clone numbers when treated with CPT in both YPGlu (6.94 C/10<sup>7</sup>,  $p < 0.01$ ) and YPGal (21.83 C/10<sup>7</sup>,  $p < 0.01$ ) groups (Figure 6C). After deducing the mean value of RR clone number caused by YPGlu, the RR clone number in CPT-treated group was still larger than that in the control group (Figure 6D), indicating that CPT might have increased the transcription-mediated AT-rich fragility.

(G) Compared to *RAD51* single deletion or *RNH1/201* double deletion, *RAD51* and *RNH1/201* triple deletion significantly increases the transcription induced RR clone numbers in AT-rich yeasts.

(H) Breakpoint identification of YPGal- or YPGlu-induced RR clones generated from *rad51Δ*, *rad51Δ rnh1Δ rnh201Δ*, and *yku70Δ rnh1Δ rnh201Δ* AT-rich yeasts. First lane: DNA marker; second lane: normal AT-rich yeasts; third and other lanes: RR yeasts. \*,  $p < 0.05$ ; \*\*,  $p < 0.01$ ; ns, not significant; Student's *t* test. DNA marker sizes, see STAR Methods.



**Figure 6. The effects of topoisomerase inhibitors on the AT-rich stability**

(A) A low dose of the topoisomerase II inhibitor Etoposide (Etop) or the topoisomerase I inhibitor camptothecin (CPT) has no significant effect on the YPGal-induced RR clone numbers in *rad1Δ* AT-rich yeasts.  
 (B) Breakpoint identification of *rad1Δ* RR clones induced by YPGal and treated with inhibitors.  
 (C) Both 2  $\mu$ M of Etop and 0.5  $\mu$ M of CPT can increase the YPGal-induced RR clone numbers in *rad51Δ* AT-rich yeasts.  
 (D) But only CPT treatment increased the transcription-mediated RR clone numbers in *rad51Δ* AT-rich yeasts.  
 (E) Breakpoint identification of *rad51Δ* RR clones induced by YPGal or YPGlu and treated with Etop.  
 (F) Breakpoint identification of *rad51Δ* RR clones induced by YPGal or YPGlu and treated with CPT. For all DNA gel electrophoresis results, first lane: DNA marker; second lane: normal AT-rich yeasts; third and other lanes: RR yeasts. \*,  $p < 0.05$ ; \*\*,  $p < 0.01$ ; ns, not significant; Student's *t* test. DNA marker sizes, see STAR Methods.

We then detected the breakpoints of *rad51Δ* RR clones in the yeast treated with 2  $\mu$ M Etop. In the YPGal group, we found the breakpoints in 16/23 RR clones situated between AT-rich sequence and *KanMX* and 7/23 located in AT-rich regions, with no significant difference when compared to normal *rad51Δ* RR clones ( $p > 0.05$ , Fisher's exact test, Figures 5H and 6E). In the YPGlu group, 5/23 breakpoints were found located between AT-rich and *KanMX*, and 18/23 located at AT-rich regions. This distribution was significantly different from that in normal *rad51Δ* RR clones ( $p < 0.01$ , Fisher's exact test, Figures 5H and 6E).

For the breakpoints of *rad51Δ* RR clones in the CPT-treated YPGal-cultured yeasts, 5/23 located between AT-rich and

*KanMX* gene, while 18/23 were distributed even farther, between the *KanMX* gene and *PRB1* gene (Figure 1D). This distribution was significantly different from that in normal YPGal-cultured *rad51Δ* RR clones ( $p < 0.01$ , Chi-squared Test, Figures 5H and 6F). In the YPGlu group, 18/23 breakpoints were located between AT-rich and *KanMX* gene and 5/23 located at AT-rich regions, which was consistent with the findings in normal YPGlu-cultured *rad51Δ* RR clones (Figures 5H and 6F).

## DISCUSSION

In this study, when *rad51Δ* yeasts were treated with Etop or CPT, no breakpoints outside of the genome region from *PRB1* to



AT-rich were identified in the RR clones. Since the *PRB1* gene locates adjacent to the *KanMX-ATrich-GAL1p* cassette, and no additional RR clones were induced by Etop or CPT in *rad1Δ* yeasts, the increase in RR clone numbers in Etop or CPT treated *rad51Δ* yeasts may be attributed to the combined effects of AT-rich and topoisomerase poisons, rather than the individual effects of topoisomerase poison or AT-rich. In *Arabidopsis thaliana*, TOP2 plays a role in DSB homologous recombination repair.<sup>37</sup> It has been proposed that topoisomerases might also participate in relaxing the DNA donor state to promote DNA synthesis in the D-loop.<sup>38</sup> However, the impact of topoisomerases on the AT-rich associated DSB repair still needs to be explored. As CPT can move the breakpoint of *rad51Δ* RR yeasts forward, we propose that TOP1 might participate in the AT-rich DSB repair in *rad51Δ* yeasts. TOP1 mainly functions on DNA torsion removal and regulating DNA replication and transcription by nicking and resealing the DNA.<sup>39,40</sup> Deletion of *TOP1* would limit the transcription of long genes<sup>41</sup> and decrease the DNA replication speed on long chromosomes.<sup>42</sup> Unlike *TOP1* deletion, TOP1 inhibitor CPT suppresses DNA resealing but not nicking activity, and generates covalent TOP1-DNA crosslink.<sup>43</sup> In this study, the effects of CPT on breakpoint positions might be caused by TOP1 poisons but not the accumulation of DNA torsion.

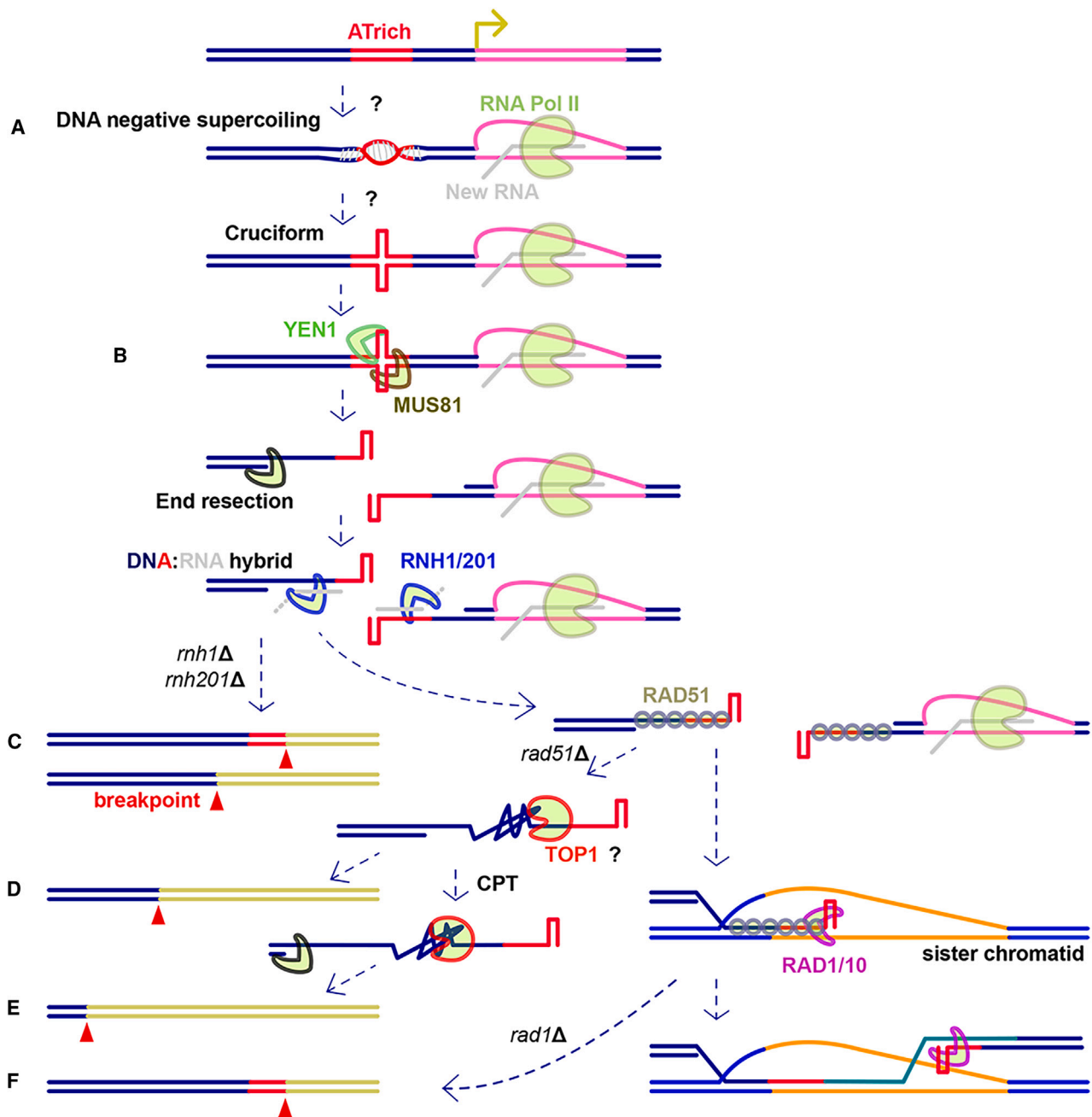
It has been reported that replication stress is a critical driver of AT-rich-mediated genome instability.<sup>44</sup> In this study, we found transcription-mediated AT-rich instability in yeast and investigated how transcription-mediated AT-rich DSBs were repaired (Figure 7). Based on our results, we hypothesized that transcription promotes the accumulation of negative supercoiling at the gene boundaries.<sup>16</sup> Negative supercoiling then stimulate the transformation of AT-rich regions into cruciform structures (Figure 7A).<sup>2,4</sup> Subsequently, these cruciform DNA structures are cleaved by MUS81 or YEN1, generating DSBs with stem structures at their ends (Figure 7B). Afterward, the DSB end is resected, and the 3' overhangs are protected by DNA:RNA hybrid formation. If *RNH1* and *RNH201* are disrupted, the homologous recombination repair will be suppressed, leading to the formation of gross genome rearrangements due to various mechanisms (Figure 7C). Else, the RNA in DNA:RNA hybrids is degraded by RNase H, and the single strand 3' overhangs invade the sister chromatid with the assistance of RAD51. If *RAD51* is disrupted, gross genome rearrangements will form and the breakpoints will be located at the 3' overhangs (Figure 7D). As there is no RAD51-mediated 3' overhang invading, prolonged exposure may lead to the coiling of the 3' overhang. If TOP1 is poisoned, the tangled single-strand DNA might cause DSB end resection to be even longer, leaving a longer 3' overhang and shifting the breakpoint further forward from the AT-rich region (Figure 7E). If the 3' overhang has been exchanged with the template DNA strand by RAD51, in order to synthesize new DNA, the stem structure at its end should be removed by the RAD1/10 complex. If *RAD1* is disrupted, homologous recombination will still be suppressed, leading to gross genome rearrangement with breakpoints in the AT-rich region (Figure 7F).

In addition to transcription-mediated AT-rich instability, we also found that the triple deletion of *RNH1*, *RNH201*, and *YKU70* could significantly exacerbate the instability of AT-rich in a non-transcription-mediated manner. However, no obvious

effect of *YKU70* single disruption has been found on the AT-rich stability in either transcription or non-transcription conditions. Further investigations are needed to determine if there are complementary functions of RNase H and non-homologous end-joining in maintaining the stability of AT-rich regions. In our study, YPGlu-cultured normal 16p11.2 AT-rich yeasts and 22q11.2 AT-rich yeasts could also generate gross genome rearrangements with breakpoints located at AT-rich regions. Double deletion of *RAD1/10*, *RNH1/RNH201*, or single deletion of *RAD51* could promote YPGlu-mediated RR clone formation. These gross genome rearrangements might also be caused by non-transcription-mediated AT-rich instability.

As the BY4742 yeast is haploid and has a different genome background compared to the human genome, the direct analysis of whether AT-rich mediates CNV formation in diploid cells were not conducted in this study. However, as 16p11.2 associated AT-rich region can induce copy number loss in ATrich-ATrich and ATrich-Control yeasts, we hypothesized that this AT-rich region can also generate DSBs and mediate the formation of reCNVs through NAHR in human cells. In this study, the gross genome rearrangements caused by AT-rich in yeasts occur at a rate of approximately 0–100/10<sup>7</sup>/24 h, which is much lower than the rates of reCNV formation (about 1/10000 to 1/1000 per generation) observed at 16p11.2 or 22q11.2 CNVs in humans.<sup>45</sup> However, for the ATrich-ATrich yeasts, the copy number loss rate dramatically increased to ~100/10<sup>4</sup>/24 h, which is much faster than the rate of reCNV formation in humans. Why the copy number loss rate is so high in ATrich-ATrich yeasts is not well known in this study. This high formation rate of reCNVs in human might be caused by the persistent DNA replication in male germ stem cells or DNA transcription in female oocytes for ten years to several decades, which increases the risk of transcription-mediated or non-dependent ATrich-caused genome mutations. All these results revealed a new potential source for the formation of 16p11.2/22q11.2 reCNVs, as well as other AT-rich associated reCNVs, structure variants, and DSBs. This finding might be helpful for preventing AT-rich associated birth defect in humans in the future.

In summary, in this study we found the 16p11.2 and 22q11.2 reCNVs associated AT-rich sequences are unstable in yeast genome. The AT-rich sequence can induce gross genome rearrangement, including copy number loss in yeast genome. We found the stability of AT-rich in yeast is accelerated by downstream transcription. Downstream transcription promotes the formation of cruciform structure of AT-rich sequence. Transcription-mediated genome instability of AT-rich can be attenuated by *MUS81* or *YEN1* disruption and be further accelerated by disruption of homologous recombination associated genes such as RNase H, *RAD1*, and *RAD51*. Triple deletion of *RNH1/201* and *YKU70* can make AT-rich unstable in a transcription-independent manner. In the human genome, 16p11.2 CNV associated AT-rich locates at the downstream region of *BOLA2B* gene, and its homolog located at the intron region of *BOLA2-SMG1P6* fusion gene (chr16:29462040-29462615). The 22q11.2 CNV associated AT-rich locates at the upstream region of *FAM230H* gene, and it has homologs located at the upstream of *FAM230* family genes such as *FAM230E*, *FAM230B*, *FAM230A*, and *FAM230D*. The NAHR repair of



**Figure 7. Predicted model of transcription-mediated DSB formation at AT-rich and the possible DSB repair mechanism**

(A) Downstream transcription may promote the accumulation of negative supercoiling at AT-rich region, which then induces the cruciform formation.

(B) Cruciform structure is cleaved by the YEN1 or MUS81 endonuclease.

(C) After DSB ends are resected, 3' overhangs form DNA:RNA hybrids with newly synthesized RNA. If the RNA in DNA:RNA hybrids cannot be degraded by RNase H, then the DSB will lead to gross genome rearrangements with breakpoints located in AT-rich regions or upstream regions of AT-rich.

(D) If RNA is degraded from DNA:RNA hybrids, but there is no RAD51 to mediate 3' overhang strand exchange, then DSBs will generate gross genome rearrangements with breakpoints located in upstream region of AT-rich.

(E) If TOP1 is poisoned by CPT, the exposed 3' overhangs of AT-rich DSBs in *rad51Δ* yeasts might become entangled, potentially impeding homologous recombination. Then the DSB ends might be resected longer and induce gross genome rearrangements with breakpoints located in upstream regions far away from AT-rich.

(F) If the 3' overhang can form a D loop with template DNA but the stem residues cannot be removed by RAD1/10, normal homologous recombination will be interrupted, leading to gross genome rearrangements with breakpoints located at AT-rich regions. Unverified speculations are marked by question marks.

DSB at the 16p11.2/22q11.2 reCNV associated AT-rich sequences might be one of the causes of 16p11.2/22q11.2 reCNV formation. However, whether the 16p11.2/22q11.2 reCNV associated AT-rich sequences could form cruciform structures and be affected by related gene transcription in human cells still requires further research. Nonetheless, the AT-rich-mediated copy number change might be a possible mechanism of 16p11.2/22q11.2 reCNV formation and suppression of DSB formation in AT-rich regions would be a possible path to prevent reCNV *de novo* formation in humans.

### Limitations of the study

In this study, there are some limitations to be considered. First, beside *GAL1* promoter activation and gene disruption, the different metabolic microenvironments and gene expression profiles between YPGlu and YPGal cultured yeasts might also affect the AT-rich stability or DSB repair, and interfere with the results in this study. Second, in the yeast system, *CIN8*, *HXT13*, and/or *AVT2* genes were substituted, which might have effects on the stability or topological structure of the yeast genome. Third, double deletion of *RNH1/RNH201* will induce the accumulation of R-loops in the yeast genome and make the genome unstable. All these limitations may affect the yeast model used in this study.

### RESOURCE AVAILABILITY

#### Lead contact

Further information will be fulfilled by the lead contact (Jun-Yu Ma, [majy@gd2h.org.cn](mailto:majy@gd2h.org.cn)).

#### Materials availability

The yeast strains are available from the [lead contact](#) upon request. Reagents and materials in this study are commercially available.

#### Data and code availability

- Data: The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive<sup>46</sup> in National Genomics Data Center,<sup>47</sup> China National Center for Bioinformatics/Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRA018885) that are publicly accessible at <https://ngdc.cncb.ac.cn/gsa>. The MCF-7 RNA-seq data was downloaded from SRA (<https://www.ncbi.nlm.nih.gov/sra>, accession number: SRR19737218) and MCF-7 END\_seq data was downloaded from GEO (<https://www.ncbi.nlm.nih.gov/geo>, accession number: GSE99194). The clinvarCNV data was downloaded from UCSC genome browser website (<https://hgdownload.soe.ucsc.edu/goldenPath/hg19/database>) and the recurrent CNV data was downloaded from ClinGen website (<https://search.clinicalgenome.org/kb/downloads>).
- Code: Codes for AT-rich sequence analysis are listed in [Data S4](#) and any other codes and data underlying this article are available from the [lead contact](#) upon request.
- Other items: All other information required to reanalyze the data reported are available from the [lead contact](#) upon request.

### ACKNOWLEDGMENTS

We thank all the members at Clinical Lab and Fertility Preservation Lab of the Reproductive Medicine Center at Guangdong Second Provincial General Hospital for their supports to this study. This work is supported by the National Key R&D Program of China (2022YFC2703204) and the National Natural Science Foundation of China (82271683).

### AUTHOR CONTRIBUTIONS

J.-Y.M., S.Y., L.-N.C., and X.-H.O. designed the research; F.-Y.X., T.-J.X., and J.-Y.M. performed the analysis; F.-Y.X., J.C., X.X., T.-J.X., and S.L. analyzed the results; J.-Y.M. made figures; X.G.Z. and J.Y.M. wrote the article.

### DECLARATION OF INTERESTS

The authors declare no competing interests.

### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Amplification of AT-rich sequences
  - Yeast strains
- METHOD DETAILS
  - Human AT-rich sequence stability analysis
  - Yeast genome rearrangement rate analysis
  - Chemical reagent treatment and yeast survival analysis
  - Cruciform DNA antibody-based chromatin immunoprecipitation
  - Breakpoint and copy number loss identification
  - Quantitative PCR
  - Pacbio HiFi sequencing
- QUANTIFICATION AND STATISTICAL ANALYSIS

### SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2024.111508>.

Received: April 25, 2024

Revised: September 29, 2024

Accepted: November 27, 2024

Published: November 30, 2024

### REFERENCES

1. Ward, G.K., McKenzie, R., Zannis-Hadjopoulos, M., and Price, G.B. (1990). The dynamic distribution and quantification of DNA cruciforms in eukaryotic nuclei. *Exp. Cell Res.* 188, 235–246. [https://doi.org/10.1016/0014-4827\(90\)90165-7](https://doi.org/10.1016/0014-4827(90)90165-7).
2. Mizuuchi, K., Mizuuchi, M., and Gellert, M. (1982). Cruciform structures in palindromic DNA are favored by DNA supercoiling. *J. Mol. Biol.* 156, 229–243. [https://doi.org/10.1016/0022-2836\(82\)90325-4](https://doi.org/10.1016/0022-2836(82)90325-4).
3. Krasilnikov, A.S., Podtelezchnikov, A., Vologodskii, A., and Mirkin, S.M. (1999). Large-scale effects of transcriptional DNA supercoiling in vivo. *J. Mol. Biol.* 292, 1149–1160. <https://doi.org/10.1006/jmbi.1999.3117>.
4. Courey, A.J., and Wang, J.C. (1983). Cruciform formation in a negatively supercoiled DNA may be kinetically forbidden under physiological conditions. *Cell* 33, 817–829. [https://doi.org/10.1016/0092-8674\(83\)90024-7](https://doi.org/10.1016/0092-8674(83)90024-7).
5. Feng, X., Xie, F.Y., Ou, X.H., and Ma, J.Y. (2020). Cruciform DNA in mouse growing oocytes: Its dynamics and its relationship with DNA transcription. *PLoS One* 15, e0240844. <https://doi.org/10.1371/journal.pone.0240844>.
6. Bell, D., Sabloff, M., Zannis-Hadjopoulos, M., and Price, G. (1991). Anti-cruciform DNA affinity purification of active mammalian origins of replication. *Biochim. Biophys. Acta* 1089, 299–308. [https://doi.org/10.1016/0167-4781\(91\)90169-m](https://doi.org/10.1016/0167-4781(91)90169-m).
7. Yahyaoui, W., Callejo, M., Price, G.B., and Zannis-Hadjopoulos, M. (2007). Deletion of the cruciform binding domain in CBP/14-3-3 displays reduced origin binding and initiation of DNA replication in budding yeast. *BMC Mol. Biol.* 8, 27. <https://doi.org/10.1186/1471-2199-8-27>.

8. Dayn, A., Malkhosyan, S., and Mirkin, S.M. (1992). Transcriptionally driven cruciform formation in vivo. *Nucleic Acids Res.* *20*, 5991–5997. <https://doi.org/10.1093/nar/20.22.5991>.
9. Duardo, R.C., Guerra, F., Pepe, S., and Capranico, G. (2023). Non-B DNA structures as a booster of genome instability. *Biochimie* *274*, 176–192. <https://doi.org/10.1016/j.biochi.2023.07.002>.
10. Gotter, A.L., Nimmakayalu, M.A., Jalali, G.R., Hacker, A.M., Vorstman, J., Conforto Duffy, D., Medne, L., and Emanuel, B.S. (2007). A palindrome-driven complex rearrangement of 22q11.2 and 8q24.1 elucidated using novel technologies. *Genome Res.* *17*, 470–481. <https://doi.org/10.1101/gr.6130907>.
11. Kurahashi, H., Shaikh, T., Takata, M., Toda, T., and Emanuel, B.S. (2003). The constitutional t(11;22): another translocation mediated by palindromic AT-rich repeats. *Am. J. Hum. Genet.* *72*, 733–738. <https://doi.org/10.1086/368062>.
12. Kurahashi, H., and Emanuel, B.S. (2001). Long AT-rich palindromes and the constitutional t(11;22) breakpoint. *Hum. Mol. Genet.* *10*, 2605–2617. <https://doi.org/10.1093/hmg/10.23.2605>.
13. Correll-Tash, S., Lilley, B., Salmons Iv, H., Mlynarski, E., Franconi, C.P., McNamara, M., Woodbury, C., Easley, C.A., and Emanuel, B.S. (2021). Double strand breaks (DSBs) as indicators of genomic instability in PATRR-mediated translocations. *Hum. Mol. Genet.* *29*, 3872–3881. <https://doi.org/10.1093/hmg/ddaa251>.
14. Inagaki, H., Ohye, T., Kogo, H., Tsutsumi, M., Kato, T., Tong, M., Emanuel, B.S., and Kurahashi, H. (2013). Two sequential cleavage reactions on cruciform DNA structures cause palindrome-mediated chromosomal translocations. *Nat. Commun.* *4*, 1592. <https://doi.org/10.1038/ncomms2595>.
15. van Wietmarschen, N., Sridharan, S., Nathan, W.J., Tubbs, A., Chan, E.M., Callen, E., Wu, W., Belinky, F., Tripathi, V., Wong, N., et al. (2020). Repeat expansions confer WRN dependence in microsatellite-unstable cancers. *Nature* *586*, 292–298. <https://doi.org/10.1038/s41586-020-2769-8>.
16. Achar, Y.J., Adhil, M., Choudhary, R., Gilbert, N., and Foiani, M. (2020). Negative supercoil at gene boundaries modulates gene topology. *Nature* *577*, 701–705. <https://doi.org/10.1038/s41586-020-1934-4>.
17. Kurahashi, H., and Emanuel, B.S. (2001). Unexpectedly high rate of de novo constitutional t(11;22) translocations in sperm from normal males. *Nat. Genet.* *29*, 139–140. <https://doi.org/10.1038/ng1001-139>.
18. Shaikh, T.H. (2017). Copy Number Variation Disorders. *Curr. Genet. Med. Rep.* *5*, 183–190. <https://doi.org/10.1007/s40142-017-0129-2>.
19. Zhang, F., Gu, W., Hurler, M.E., and Lupski, J.R. (2009). Copy number variation in human health, disease, and evolution. *Annu. Rev. Genom. Hum. Genet.* *10*, 451–481. <https://doi.org/10.1146/annurev.genom.9.081307.164217>.
20. Kim, P.M., Lam, H.Y.K., Urban, A.E., Korb, J.O., Affourtit, J., Grubert, F., Chen, X., Weissman, S., Snyder, M., and Gerstein, M.B. (2008). Analysis of copy number variants and segmental duplications in the human genome: Evidence for a change in the process of formation in recent evolutionary history. *Genome Res.* *18*, 1865–1874. <https://doi.org/10.1101/gr.081422.108>.
21. Vysotskiy, M., Zhong, X., Miller-Fleming, T.W., Zhou, D., Autism Working Group of the Psychiatric Genomics Consortium; Bipolar Disorder Working Group of the Psychiatric Genomics Consortium; Schizophrenia Working Group of the Psychiatric Genomics Consortium; Cox, N.J., and Weiss, L.A. (2021). Integration of genetic, transcriptomic, and clinical data provides insight into 16p11.2 and 22q11.2 CNV genes. *Genome Med.* *13*, 172. <https://doi.org/10.1186/s13073-021-00972-1>.
22. Kendall, K.M., Rees, E., Escott-Price, V., Eimon, M., Thomas, R., Hewitt, J., O'Donovan, M.C., Owen, M.J., Walters, J.T.R., and Kirov, G. (2017). Cognitive Performance Among Carriers of Pathogenic Copy Number Variants: Analysis of 152,000 UK Biobank Subjects. *Biol. Psychiatr.* *82*, 103–110. <https://doi.org/10.1016/j.biopsych.2016.08.014>.
23. McDonald-McGinn, D.M., Sullivan, K.E., Marino, B., Philip, N., Swillen, A., Vorstman, J.A.S., Zackai, E.H., Emanuel, B.S., Vermeesch, J.R., Morrow, B.E., et al. (2015). 22q11.2 deletion syndrome. *Nat. Rev. Dis. Prim.* *1*, 15071. <https://doi.org/10.1038/nrdp.2015.71>.
24. Duyzend, M.H., Nettle, X., Coe, B.P., Baker, C., Nickerson, D.A., Bernier, R., and Eichler, E.E. (2016). Maternal Modifiers and Parent-of-Origin Bias of the Autism-Associated 16p11.2 CNV. *Am. J. Hum. Genet.* *98*, 45–57. <https://doi.org/10.1016/j.ajhg.2015.11.017>.
25. Canela, A., Maman, Y., Jung, S., Wong, N., Callen, E., Day, A., Kieffer-Kwon, K.R., Pekowska, A., Zhang, H., Rao, S.S.P., et al. (2017). Genome Organization Drives Chromosome Fragility. *Cell* *170*, 507–521.e18. <https://doi.org/10.1016/j.cell.2017.06.034>.
26. Landrum, M.J., Lee, J.M., Riley, G.R., Jang, W., Rubinstein, W.S., Church, D.M., and Maglott, D.R. (2014). ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* *42*, D980–D985. <https://doi.org/10.1093/nar/gkt1113>.
27. Dittwald, P., Gambin, T., Szafranski, P., Li, J., Amato, S., Divon, M.Y., Rodríguez Rojas, L.X., Elton, L.E., Scott, D.A., Schaaf, C.P., et al. (2013). NAHR-mediated copy-number variants in a clinical population: mechanistic insights into both genomic disorders and Mendelizing traits. *Genome Res.* *23*, 1395–1409. <https://doi.org/10.1101/gr.152454.112>.
28. Kaushal, S., Wollmuth, C.E., Das, K., Hile, S.E., Regan, S.B., Barnes, R.P., Haouzi, A., Lee, S.M., House, N.C.M., Guyumdzhyyan, M., et al. (2019). Sequence and Nuclease Requirements for Breakage and Healing of a Structure-Forming (AT)<sub>n</sub> Sequence within Fragile Site FRA16D. *Cell Rep.* *27*, 1151–1164.e5. <https://doi.org/10.1016/j.celrep.2019.03.103>.
29. Zhang, H., and Freudenreich, C.H. (2007). An AT-rich sequence in human common fragile site FRA16D causes fork stalling and chromosome breakage in *S. cerevisiae*. *Mol. Cell* *27*, 367–379. <https://doi.org/10.1016/j.molcel.2007.06.012>.
30. Dehé, P.M., and Gaillard, P.H.L. (2017). Control of structure-specific endonucleases to maintain genome stability. *Nat. Rev. Mol. Cell Biol.* *18*, 315–330. <https://doi.org/10.1038/nrm.2016.177>.
31. Li, S., Lu, H., Wang, Z., Hu, Q., Wang, H., Xiang, R., Chiba, T., and Wu, X. (2019). ERCC1/XPF Is Important for Repair of DNA Double-Strand Breaks Containing Secondary Structures. *iScience* *16*, 63–78. <https://doi.org/10.1016/j.isci.2019.05.017>.
32. Guervilly, J.H., and Gaillard, P.H. (2018). SLX4: multitasking to maintain genome stability. *Crit. Rev. Biochem. Mol. Biol.* *53*, 475–514. <https://doi.org/10.1080/10409238.2018.1488803>.
33. Nikolova, T., Ensminger, M., Löbrich, M., and Kaina, B. (2010). Homologous recombination protects mammalian cells from replication-associated DNA double-strand breaks arising in response to methyl methanesulfonate. *DNA Repair* *9*, 1050–1063. <https://doi.org/10.1016/j.dnarep.2010.07.005>.
34. Lundin, C., North, M., Erixon, K., Walters, K., Jenssen, D., Goldman, A.S.H., and Helleday, T. (2005). Methyl methanesulfonate (MMS) produces heat-labile DNA damage but no detectable in vivo DNA double-strand breaks. *Nucleic Acids Res.* *33*, 3799–3811. <https://doi.org/10.1093/nar/gki681>.
35. Szlachta, K., Manukyan, A., Raimer, H.M., Singh, S., Salamon, A., Guo, W., Lobachev, K.S., and Wang, Y.H. (2020). Topoisomerase II contributes to DNA secondary structure-mediated double-stranded breaks. *Nucleic Acids Res.* *48*, 6654–6671. <https://doi.org/10.1093/nar/gkaa483>.
36. Pommier, Y., Nussenzweig, A., Takeda, S., and Austin, C. (2022). Human topoisomerases and their roles in genome stability and organization. *Nat. Rev. Mol. Cell Biol.* *23*, 407–427. <https://doi.org/10.1038/s41580-022-00452-3>.
37. Martínez-García, M., White, C.I., Franklin, F.C.H., and Sanchez-Moran, E. (2021). The role of topoisomerase II in DNA repair and recombination in *Arabidopsis thaliana*. *Int. J. Mol. Sci.* *22*, 13115. <https://doi.org/10.3390/ijms222313115>.



38. Wright, W.D., Shah, S.S., and Heyer, W.D. (2018). Homologous recombination and the repair of DNA double-strand breaks. *J. Biol. Chem.* 293, 10524–10535. <https://doi.org/10.1074/jbc.TM118.000372>.
39. Pommier, Y., Sun, Y., Huang, S.Y.N., and Nitiss, J.L. (2016). Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. *Nat. Rev. Mol. Cell Biol.* 17, 703–721. <https://doi.org/10.1038/nrm.2016.111>.
40. Cho, J.E., and Jinks-Robertson, S. (2018). Topoisomerase I and Genome Stability: The Good and the Bad. *Methods Mol. Biol.* 1703, 21–45. [https://doi.org/10.1007/978-1-4939-7459-7\\_2](https://doi.org/10.1007/978-1-4939-7459-7_2).
41. Mabb, A.M., Simon, J.M., King, I.F., Lee, H.M., An, L.K., Philpot, B.D., and Zylka, M.J. (2016). Topoisomerase 1 Regulates Gene Expression in Neurons through Cleavage Complex-Dependent and -Independent Mechanisms. *PLoS One* 11, e0156439. <https://doi.org/10.1371/journal.pone.0156439>.
42. Kegel, A., Betts-Lindroos, H., Kanno, T., Jeppsson, K., Ström, L., Katou, Y., Itoh, T., Shirahige, K., and Sjögren, C. (2011). Chromosome length influences replication-induced topological stress. *Nature* 471, 392–396. <https://doi.org/10.1038/nature09791>.
43. Pommier, Y. (2006). Topoisomerase I inhibitors: camptothecins and beyond. *Nat. Rev. Cancer* 6, 789–802. <https://doi.org/10.1038/nrc1977>.
44. Irony-Tur Sinai, M., Salamon, A., Stanleigh, N., Goldberg, T., Weiss, A., Wang, Y.H., and Kerem, B. (2019). AT-dinucleotide rich sequences drive fragile site formation. *Nucleic Acids Res.* 47, 9685–9695. <https://doi.org/10.1093/nar/gkz689>.
45. Ma, J.Y., Xia, T.J., Li, S., Yin, S., Luo, S.M., and Li, G. (2024). Germline cell de novo mutations and potential effects of inflammation on germline cell genome stability. *Semin. Cell Dev. Biol.* 154, 316–327. <https://doi.org/10.1016/j.semcdb.2022.11.003>.
46. Chen, T., Chen, X., Zhang, S., Zhu, J., Tang, B., Wang, A., Dong, L., Zhang, Z., Yu, C., Sun, Y., et al. (2021). The Genome Sequence Archive Family: Toward Explosive Data Growth and Diverse Data Types. *Dev. Reprod. Biol.* 19, 578–583. <https://doi.org/10.1016/j.gpb.2021.08.001>.
47. CNCB-NGDC Members and Partners (2021). Database Resources of the National Genomics Data Center, China National Center for Bioinformatics in 2021. *Nucleic Acids Res.* 49, D18–D28. <https://doi.org/10.1093/nar/gkaa1022>.
48. Amberg, D.C., Burke, D.J., and Strathern, J.N. (2005). *Methods in yeast genetics : a Cold Spring Harbor Laboratory course manual* (Cold Spring Harbor Laboratory Press).
49. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoutte, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137. <https://doi.org/10.1186/gb-2008-9-9-r137>.
50. Rehm, H.L., Berg, J.S., Brooks, L.D., Bustamante, C.D., Evans, J.P., Landrum, M.J., Ledbetter, D.H., Maglott, D.R., Martin, C.L., Nussbaum, R.L., et al. (2015). ClinGen—the Clinical Genome Resource. *N. Engl. J. Med.* 372, 2235–2242. <https://doi.org/10.1056/NEJMs1406261>.
51. Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings Bioinf.* 14, 178–192. <https://doi.org/10.1093/bib/bbs017>.
52. Chen, C., and Kolodner, R.D. (1999). Gross chromosomal rearrangements in *Saccharomyces cerevisiae* replication and recombination defective mutants. *Nat. Genet.* 23, 81–85. <https://doi.org/10.1038/12687>.
53. Lelandais, G., Blugeon, C., and Merhej, J. (2016). ChIPseq in Yeast Species: From Chromatin Immunoprecipitation to High-Throughput Sequencing and Bioinformatics Data Analyses. *Methods Mol. Biol.* 1361, 185–202. [https://doi.org/10.1007/978-1-4939-3079-1\\_11](https://doi.org/10.1007/978-1-4939-3079-1_11).
54. Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P.A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* 37, 540–546. <https://doi.org/10.1038/s41587-019-0072-8>.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
cruciform DNA antibody	GeneTex	Cat#GTX54648
<b>Chemicals, peptides, and recombinant proteins</b>		
Mouse IgG (H + L)	Sangon	Cat#D110503
Camptothecin	Acmecc	Cat#7689-03-4
Etoposide	Beyotime	Cat#SC0173
Methyl methanesulfonate	Acmecc	Cat#66-27-3
Protein A/G beads	MCE	Cat#HY-K0202
G418	Sangon	Cat#A600958
Canavanine sulfate salt	Sangon	Cat#A606173
5-fluoroorotic acid	Macklin	Cat#F832427
Zeocin Selection Antibiotic, Sterile	MCE	Cat#HY-K1053
Hydrogen peroxide solution	Macklin	Cat#H792073
DNA marker III	Tiangen	Cat#MD103
Zirconia beads	Youlisheng	Cat#111178579
Snailase	Solarbio	#S8280
<b>Critical commercial assays</b>		
TIANamp Genomic DNA Kit	Tiangen	Cat#DP304
Es Taq Master Mix	CWbio	Cat#CW0690
Universal DNA Purification Kit	Tiangen	Cat#DP214
Hieff Clone Zero TOPO-TA Simple Cloning Kit	Yeasen	Cat#10908ES20
DNA Purification Kit with Magnetic Beads	Beyotime	Cat#D0041M
Rapid Taq Master Mix	Vazyme	Cat#P222-01
EZ-10 DNAaway RNA Mini-Preps Kit	Sangon	Cat#B618133
HiScript III 1st Strand cDNA Synthesis Kit	Vazyme	Cat#R312-01
AceQ qPCR SYBR Green Master Mix	Vazyme	Cat#Q111-02
FastPure Cell/Tissue DNA Isolation Mini Kit	Vazyme	Cat#DC102-01
<b>Deposited data</b>		
Pacbio HiFi Seq data deposited on Genome Sequence Archive ( <a href="https://ngdc.cncb.ac.cn/gsa">https://ngdc.cncb.ac.cn/gsa</a> )	This paper	CRA018885
MCF-7 RNA-seq data	SRA	SRR19737218
MCF-7 END_seq	GEO	GSE99194
clinvarCNV	UCSC	<a href="https://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/">https://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/</a>
Recurrent CNV data	ClinGen	<a href="https://search.clinicalgenome.org/kb/downloads">https://search.clinicalgenome.org/kb/downloads</a>
<b>Experimental models: Organisms/strains</b>		
Yeast strains	This paper	Table S1
<b>Oligonucleotides</b>		
Primers	This paper	Table S2
<b>Recombinant DNA</b>		
16p11.2_AT-rich and 16p11.2_Control cassettes in yeast	This paper	Data S1
22q11.2_AT-rich and 22q11.2_Control cassettes in yeast	This paper	Data S2
<b>Software and algorithms</b>		
MACS3	Zhang et al.	<a href="https://github.com/mac3-project/MACS">https://github.com/mac3-project/MACS</a>
Integrative Genomics Viewer	IGV team	<a href="https://igv.org/">https://igv.org/</a>

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Jupyter notebook	Jupyter	<a href="https://jupyter.org/">https://jupyter.org/</a>
Flye-2.9.5	Pavel Pevzner's lab	<a href="https://github.com/mikolmogorov/Flye">https://github.com/mikolmogorov/Flye</a>
AT-rich sequence extraction code	This paper	Data S4

**EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**

**Amplification of AT-rich sequences**

The 16p11.2 reCNV associated AT-rich sequence at chr16:30201379-30201957 (hg19) and the its corresponding control sequence at chr16:30200828-30201319, and the 22q11.2 reCNV associated AT-rich sequence (chr22:21681065-21682882) and its corresponding control sequence (chr22:21675551-21677460) were amplified from HeLa cell DNA extracted using the TIANamp Genomic DNA Kit (Tiangen, DP304). To amplify the AT-rich sequence from HeLa cell DNA, the PCR amplification temperature was set at 60°C. To amplify the control sequence, the amplification temperature was set at 72°C. Es Taq Master Mix (CW0690, CWbio) was used for the PCR amplification, and the PCR products were purified by Universal DNA Purification Kit (DP214, Tiangen). Then AT-rich and control sequences were inserted into the pESI-T vector using the Hieff Clone Zero TOPO-TA Simple Cloning Kit (10908ES20, Yeasen) for Sanger sequencing. After that, we amplified the 16p11.2\_AT-rich, 16p11.2\_Control, 22q11.2\_AT-rich and 22q11.2\_Control sequences using the M13 F/R primers on the pESI-T vector for further analysis.

**Yeast strains**

In this study, the BY4742 strain of yeast, in which the *HXT13* gene was replaced by the *URA3* gene, was used. Firstly, we amplified the DNA segment containing the *GAL1* promoter (*GAL1p*), *CRE* coding sequence, and *CYC1* terminator (*CYC1t*) from plasmid pSH63. Then the *GAL1p-CRE-CYC1t* was ligated with either an AT-rich or its corresponding control sequence at its upstream. The *ATrich-GAL1p-CRE-CYC1t* and *Control-GAL1p-CRE-CYC1t* cassettes were further ligated with the *KanMX* cassette from plasmid *pFA6a-kanMX6* at its downstream. Then the *KanMX-ATrich-GAL1p-CRE-CYC1t* and *KanMX-Control-GAL1p-CRE-CYC1t* fragments were inserted into the *hxt13::URA3* yeast by replacing the *CIN8* gene. Finally, we obtained four yeast strains, which were labeled as AT-rich yeast (BY4742 *cin8::KanMX-16p11.2\_ATrich-GAL1p-CRE-CYC1t hxt13::URA3*), Control yeast (BY4742 *cin8::KanMX-16p11.2\_Control-GAL1p-CRE-CYC1t hxt13::URA3*), 22q11.2\_AT-rich yeast (BY4742 *cin8::KanMX-22q11.2\_ATrich-GAL1p-CRE-CYC1t hxt13::URA3*) and 22q11.2\_Control yeast (BY4742 *cin8::KanMX-22q11.2\_Control-GAL1p-CRE-CYC1t hxt13::URA3*), see Figures 1D and 2B; Datas S1 and S2.

Thereafter, the 3' ends of 16p11.2\_AT-rich and 16p11.2\_Control sequences were linked with a *GAL1p-CRE-CYC1t-LEU2* cassette, and the 16p11.2\_ATrich-*GAL1p-CRE-CYC1t-LEU2* and 16p11.2\_Control-*GAL1p-CRE-CYC1t-LEU2* cassettes were integrated into the AT-rich yeast by substituting the *AVT2* gene, which located between the *CAN1* and *hxt13::URA3* of yeast chrV. These two new yeast strains were termed as *ATrich-ATrich yeast (AT-rich yeast, avt2:16p11.2\_ATrich-GAL1p-CRE-CYC1t-LEU2)* and *ATrich-Control yeast (AT-rich yeast, avt2:16p11.2\_Control-GAL1p-CRE-CYC1t-LEU2)* respectively, see Figure 3A.

The DNA damage associated genes were disrupted in the AT-rich yeasts. PCR-based methods were used for yeast gene disruption, and the LiAc transformation method was employed to transfer PCR products into yeasts.<sup>48</sup> All yeast strains used in this study are listed in Table S1.

**METHOD DETAILS**

**Human AT-rich sequence stability analysis**

Human AT-rich sequences were extracted from the human genome (hg19) using three criteria as described in the results section. The DSB peaks in the MCF7 cell line were analyzed using the bigwig data of the END\_seq\_MCF7\_NT sample from GSE99194,<sup>25</sup> and called by the MACS3 program (parameters: -cutoff 0.5 -L 30 -g 100).<sup>49</sup> The DSB peaks were associated with AT-rich sequences if their positions (from start site to end site) located within the DNA range from AT-rich start position - 500 bp to AT-rich end position +500 bp. The CNV breakpoint information was extracted from the clinvarCNV data on the UCSC website. Position information of reCNVs was downloaded from the clinGen website.<sup>50</sup> The random sequences corresponding to each AT-rich sequences were randomly extracted from the same chromosome (non-centromere region). The CNV breakpoints located within 2500 bp upstream or downstream of the AT-rich and random sequences were labeled as AT-rich or Random sequence associated breakpoints. Genes located 5 Kbp flanking of AT-rich sequence were extracted as the AT-rich associated genes. Human genome data were visualized using IGV.<sup>51</sup> All scripts used for human genome AT-rich analysis were listed in Data S4.

**Yeast genome rearrangement rate analysis**

The methods for selecting yeasts with gross genome rearrangements have been described in previous literature<sup>52</sup> and are summarized in Figure 1E. Specifically, yeasts were firstly maintained on the synthetic defined (SD) URA-plate with 200 µg/mL G418

(A600958, Sangon). For mutation rate analysis, yeasts were cultured in SD URA-media with 2% Glycerol and G418, and shaking at 200 rpm for more than 3 h. Then,  $10^6$  yeasts in 100  $\mu$ L of water were added to 10 mL of YPGlu (YP + 2% glucose) or YPGal (YP + 2% galactose) medium and shaken for 24 h at 30°C. Then the yeasts were counted and resuspended in water to achieve a concentration of  $10^7$  per 250  $\mu$ L of water. Then 250  $\mu$ L of yeasts (18 repeats) were spotted onto the ARG-selection plates containing 60 mg/L canavanine sulfate salt (Can, A606173, Sangon) and 0.1% 5-fluoroorotic acid (5-FOA, F832427, Macklin). After drying on a sterilized clean bench for 3–5 h at room temperature, the plates were incubated at 30°C for 8 days. Then the selected Can and 5-FOA double-resistant (RR) yeast clones were counted to evaluate the *de novo* gross genome rearrangement rate in yeast.

To analyze the frequency of copy number loss in ATrich-ATrich and ATrich-Control yeasts, these two strain yeasts were firstly cultured in SD LEU-/URA-/G418 medium with 2% glucose at 30°C overnight. Then  $10^6$  yeasts in 100  $\mu$ L of water were added to 10 mL of YPGlu or YPGal medium and cultured at 30°C for 24 h. Then, for the ATrich-ATrich yeasts,  $10^4$  yeasts in 100  $\mu$ L water were dropped on the SD ARG-plates or SD ARG-/LEU-/URA-plates with 60 mg/L canavanine; for the ATrich-Control yeasts,  $10^6$  yeasts in 100  $\mu$ L water (18 repeats) were dropped on the same selection plates as ATrich-ATrich yeasts. After drying on a sterilized clean bench for 3–5 h at room temperature, the selection plates were incubated at 30°C for 8 days and the yeast clones were counted to evaluate the copy number loss rates in yeast.

### Chemical reagent treatment and yeast survival analysis

To test the homologous recombination repair efficiency of yeasts, various quantities of yeasts (approximately  $10^5$ ,  $10^4$ ,  $10^3$ ,  $10^2$ , and 10) in 10  $\mu$ L of water were spotted onto the YPD plate containing 0%, 0.004%, or 0.02% (v/v) methyl methanesulfonate (MMS, 66-27-3, AcmeC). To test the yeast sensitivity to the oxidative damage caused by hydrogen peroxide ( $H_2O_2$ ), the yeasts were firstly treated with 0 mM or 4 mM of  $H_2O_2$  (H792073, Macklin) at 30°C for 30 min, and spotted onto YPD plates. The yeasts treated with 0 mM of  $H_2O_2$  were also spotted on the plates containing 20 or 30  $\mu$ g/mL of Zeiocin (HY-K1053, MCE) to test the sensitivity of yeast to the Zeiocin induced DSBs. After water drying, the yeasts were cultured at 30°C for 2–3 days for the survival analysis. In this study, Camptothecin (CPT, 7689-03-4, AcmeC) and Etoposide (Etop, SC0173, Beyotime) were utilized to inhibit DNA topoisomerase I (TOP1) and topoisomerase II (TOP2), respectively.

### Cruciform DNA antibody-based chromatin immunoprecipitation

The chromatin immunoprecipitation (ChIP) method used in this study was based on the literature<sup>53</sup> with some modifications. In detail, AT-rich yeasts were first cultured in SD + 2% Glycerol + G418 medium at 30°C overnight. Subsequently,  $10^6$  yeasts were transferred into 20 mL YPGlu medium and 20 mL YPGal medium. After 24 h of culturing, yeasts were cross-linked with formaldehyde, and then the cross-linking process was stopped by glycine. Then, yeasts were lysed by beating them three times with zirconia beads (diameter 0.25mm) in lysis buffer at 4°C for 3 min. Then the yeast lysates were sonicated twice using Qsonica (parameters: 30 s on, 30 s off, 80% amplitude, for a total of 4 min). After centrifugation, 10  $\mu$ L suspensions were kept as Input and the remaining samples were incubated with Protein A/G beads (HY-K0202, MCE) that had bonded with cruciform DNA antibodies (clone ID: 2D3, GTX54648, GeneTex). After overnight incubation at 4°C, the beads were washed and resuspended in 100  $\mu$ L of TE-SDS as the immunoprecipitation (IP) samples. 90  $\mu$ L of TE-SDS was added to the 10  $\mu$ L Input samples. Then both IP and Input samples were treated with thermo mixer (1200 rpm at 65°C, TS100, Hangzhou Ruicheng) for 1 h. Afterward, the samples were treated with proteinase K, and the DNA was purified using the Universal DNA Purification Kit (DP214, Tiangen).

### Breakpoint and copy number loss identification

The breakpoints of RR yeast clones were roughly identified using PCR. Clones with a negative PCR bond indicated the loss of the corresponding DNA fragment in yeast. For yeast PCR, the AT-rich, ATrich-ATrich and ATrich-Control yeasts were first treated with 100  $\mu$ L of 0.1M NaOH at 100°C for 1 h, and yeast DNA was isolated using the DNA Purification Kit with Magnetic Beads (D0041M, Beyotime). For the 22q11.2 AT-rich yeasts, the yeast walls were firstly removed by Snailase (S8280, Solarbio), and the genome DNA was isolated by FastPure Cell/Tissue DNA Isolation Mini Kit (DC102-01, Vazyme). Rapid Taq Master Mix (P222-01, Vazyme) was used for PCR amplification. DNA marker III (MD103, Tiangen) was used for the gel electrophoresis of the PCR products (the marker sizes are 200, 500, 800, 1200, 2000, 3000, and 4500 bp).

### Quantitative PCR

To extract RNAs, yeasts were first suspended in an isotonic solution, and the cell wall was removed using Snailase. The RNAs were isolated using the EZ-10 DNAaway RNA Mini-Preps Kit (B618133, Sangon) and then reverse transcribed into cDNAs with the HiScript III 1st Strand cDNA Synthesis Kit (R312-01, Vazyme). The real-time quantitative PCR was used to analyze the relative *CRE* gene expression (*URA3* gene is used as internal reference), and the ChIP-PCR was employed to assess the relative level of ATrich-formed cruciform structures. All quantitative PCR experiments were conducted on the LC480 platform using the AceQ qPCR SYBR Green Master Mix (Q111-02, Vazyme). All the primer sequences used in this study are listed in [Table S2](#).

### Pacbio HiFi sequencing

To analyze the copy number loss in ATrich-ATrich and ATrich-Control yeasts, the yeast walls were removed by Snailase and genome DNA were isolated using FastPure Cell/Tissue DNA Isolation Mini Kit. Then the regions occurred copy number loss were amplified by



PCR using Rapid Taq Master Mix (primers were listed by [Table S2](#)). Then the PCR products were purified by Universal DNA Purification Kit and sequenced by Pacbio HiFi sequencing method (sequel II platform). The raw fastq data were analyzed using Flye-2.9.5<sup>54</sup> with the parameters: `-pacbio-raw -g 20k -m 2000`.

### QUANTIFICATION AND STATISTICAL ANALYSIS

The Student's *t* test was used as the default hypothesis test method to analyze the significance of the difference in RR yeast clone numbers in this study. The Wilcoxon test was used to analyze the differences in structural variant breakpoint numbers corresponding to different DNA fragments. Fishers' exact test and Pearson's Chi-squared test were used to analyze the differences in breakpoint positions among different RR yeast clones. For all hypothesis testing methods, *p*-values less than 0.01 or 0.05 were considered as significant difference. *p* < 0.01 was marked by \*\*; *p* < 0.05 was marked by \*; and not significant was marked by ns.