

## ORIGINAL ARTICLE

# Signature of backward replication slippage at the copy number variation junction

Tamae Ohye<sup>1</sup>, Hidehito Inagaki<sup>1</sup>, Mamoru Ozaki<sup>2</sup>, Toshiro Ikeda<sup>3</sup> and Hiroki Kurahashi<sup>1</sup>

Copy number abnormalities such as deletions and duplications give rise to a variety of medical problems and also manifest innocuous genomic variations. Aberrant DNA replication is suggested as the mechanism underlying *de novo* copy number abnormalities, but the precise details have remained unknown. In our present study, we analyzed the del(2)(q13q14.2) chromosomal junction site observed in a woman with a recurrent pregnancy loss. Microarray analyses allowed us to precisely demarcate a 2.8 Mb deletion in this case, which does not appear in the database of human genomic variations. This deletion includes only one brain-specific gene that could not be related to the reproduction failure of the patient. At the junction of the deletion, we found that 11–13-nucleotide sequence, originally located at the proximal breakpoint region, was repeated four times with a single-nucleotide microhomology at the joint between each repeat. The proximal region and the distal region was finally joined with six-nucleotide microhomology. The structure of the junction is consistent with backward replication slippage proposed previously. Our data lend support to the notion that a common DNA replication-mediated pathway generates copy number variation in the human genome.

*Journal of Human Genetics* (2014) 59, 247–250; doi:10.1038/jhg.2014.20; published online 20 March 2014

**Keywords:** backward replication slippage; copy number variation; deletion; DNA replication

## INTRODUCTION

The mechanisms underlying gross chromosomal rearrangements (GCRs) including deletion/duplication, translocation and inversion are still largely unknown. Among the known GCRs, deletions and duplications give rise to a number of medical issues, such as congenital anomalies and intellectual disability that arise via copy number abnormalities of indispensable genes, and also manifest as innocuous polymorphic genomic variations.<sup>1</sup> GCR development is dependent on two intrinsic factors: double-strand breakage (DSB) and its illegitimate repair.<sup>2</sup> In general, DSBs will be correctly repaired by error-free pathways via homologous recombination. However, when DSBs arise within low-copy-repeat regions or segmental duplications, template anomalies may occur during DSB repair leading to chromosomal deletions or duplications. A subset of non-random deletions/duplications is caused by such non-allelic homologous recombination events between two homologous sequences, referred to as low-copy-repeat regions or segmental duplication.<sup>3,4</sup> Programmed DSBs by Spo11 endonuclease will cause meiotic recombination in meiosis I. These non-random deletions/duplications are mainly attributed to non-allelic homologous recombination in meiosis I.<sup>5</sup> On the other hand, most deletions or duplications take place in a random fashion. Deletions have been

believed to arise from random DSBs followed by error-prone repair, such as non-homologous end joining, throughout the cell cycle particularly in G1 phase.<sup>6</sup>

Error-free homologous recombination has been believed to be a major pathway for DSB repair during S/G2 phase because sister chromatids are available.<sup>6</sup> In contrast, recent advances in genomic analyses using microarray or next generation sequencing technology have accumulated sequence information on breakpoints and junctions in random GCRs. The discovery of microhomology accompanied by complex structures at the junctions of copy number abnormalities raised the hypothesis for the involvement of aberrant DNA replication. Such a replication-based mechanism is referred to as fork stalling and template switching or as microhomology-mediated break-induced replication.<sup>7,8</sup> These mechanisms are on the basis of the collapse of the replication fork followed by a restart of DNA synthesis through the invasion by a free DNA end into another replication fork within close proximity.<sup>8</sup> In fact, nearly half of all deletions/duplications have been consistently revealed to carry microhomology at the junction.<sup>9,10</sup> However, the details of the underlying molecular pathway remain unknown in mammals.

In our present study, we characterized the genomic structure of the del(2)(q13q14.2) junction site, which was identified in a woman with

<sup>1</sup>Division of Molecular Genetics, Institute for Comprehensive Medical Science, Fujita Health University, Toyoake, Japan; <sup>2</sup>Division of Genomic Medicine, Department of Advanced Medicine, Medical Research Institute, Kanazawa Medical University, Ishikawa, Japan and <sup>3</sup>Department of Obstetrics and Gynecology, Faculty of Medicine, Kagoshima University, Kagoshima, Japan

Correspondence: Professor H Kurahashi, Division of Molecular Genetics, Institute for Comprehensive Medical Science, Fujita Health University, 1-98 Dengakugakubo, Kutsukake-cho, Toyoake, Aichi 470-1192, Japan.

E-mail: kura@fujita-hu.ac.jp

Received 2 December 2013; revised 7 February 2014; accepted 25 February 2014; published online 20 March 2014

a recurrent pregnancy loss. We provide supportive evidence for the involvement of aberrant DNA replication in the development of the underlying deletion.

## MATERIALS AND METHODS

### Subjects

A Japanese couple underwent cytogenetic examination due to two consecutive pregnancy losses. The karyotype of the male was 46,XY and that of the female was 46,XX,del(2)(q13q14.2). After informed consent was obtained, peripheral blood samples were obtained again from the woman for genomic analysis. No parental sample was obtained. This study was approved by the Ethical Review Board for Human Genome Studies at Fujita Health University (Accession number 86, approved on 12 March 2010).

### Cytogenetic microarray

Cytogenetic microarray analysis was performed using Agilent 244K in accordance with the manufacturer's protocol (Agilent Technologies, Santa Clara, CA, USA). The data were analyzed with the aid of Genomic Workbench 6.5 software (Agilent) and UCSC Human Genome Browser (<http://genome.ucsc.edu>).

### Fluorescence *in situ* hybridization

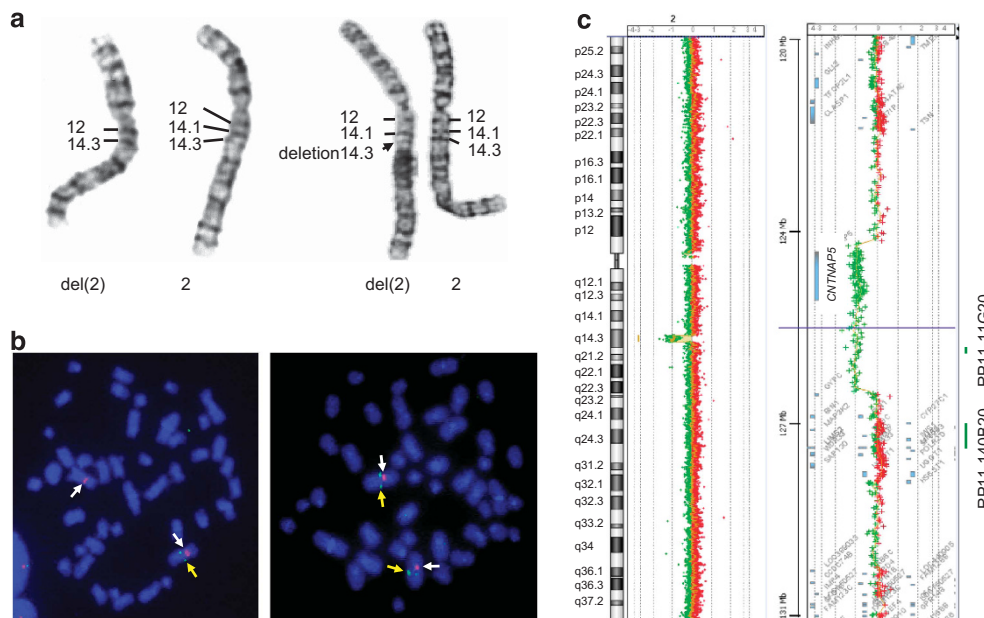
Fluorescence *in situ* hybridization was performed using standard methods. phytohaemagglutinin-stimulated lymphocytes or Epstein-Barr virus-transformed lymphoblastoid cell lines were arrested by exposure to colcemid. Metaphase preparations were then obtained by hypotonic treatment with 0.075 M KCl followed by methanol/acetate fixation. Bacterial artificial clones on 2q14.3, RP11-11G20 (chr2:126,018,973–126,184,807) and 140B20 (chr2:128,035,141–128,559,312), were used as test probes with a chromosome 2 centromere probe (CEP2 SpectrumOrange Probe; Abbott Laboratories, Abbott Park, IL, USA) used as a reference. The probe was labeled by nick translation with digoxigenin-11-dUTP. After hybridization, the probe was detected with DyLight 488 Anti-Digoxigenin/Digoxin. Chromosomes were visualized by counterstaining with 4',6-diamino-2-phenylindole.

### Analysis of junction fragments

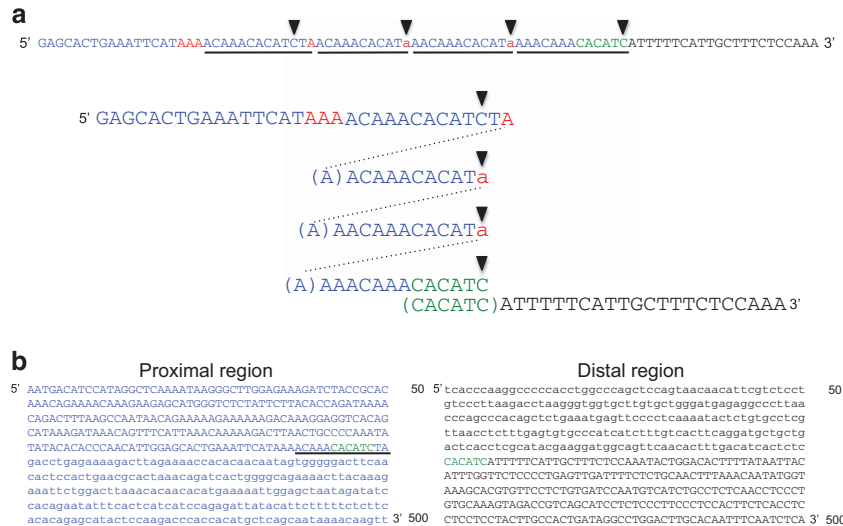
To isolate a junction fragment, standard or long-range PCR was performed using LA Taq (TaKaRa, Shiga, Japan). The PCR conditions were 35 cycles of 10 s at 98 °C and 15 min at 60 °C. PCR primers were designed using sequence data from the human genome database. The primers used for amplification were as follows: del2-3F, 5'-GCTTGCTTTGTTCAACACCCTGAG-3' and del2-5R, 5'-TACTTGTGTGCACCTTCGTTGGTATTC-3'. PCR products were directly sequenced with the PCR primers using the Sanger method. Breakpoint sequences were characterized using the RepeatMasker (<http://www.repeatmasker.org/>) and the non-B DB (<http://nonb.abcc.ncicrf.gov/apps/site/default>).

## RESULTS

Standard cytogenetic evaluations of the study couple revealed a del(2)(q13q14.2) deletion in the women (Figure 1a). As we did not obtain a parental sample, we could not determine whether this was a *de novo* deletion. To demarcate this deletion and attempt to identify the genes responsible for the recurrent pregnancy loss in this female subject, we performed cytogenetic microarray analyses. We, thereby, identified a 2.8-Mb deletion, arr[hg19] 2q14.3(124,622,589–127,367,440)x1 (Figure 1c), which was not found in the public databases such as Human Genome Variation Database (<https://gwas.biosciencencedbc.jp>) and Database of Genomic Variants (<http://dgv.tcag.ca/dgv/app/home>). It was confirmed by standard fluorescence *in situ* hybridization with a BAC probe to be located at 2q14.3 (Figure 1b). *CNTNAP5* was found to be the only gene in this deleted region. *CNTNAP5* is a brain-specific gene that encodes a protein belonging to the neurexin superfamily of unknown function. The entire *CNTNAP5* gene was lost via the 2.8-Mb deletion. We reevaluated the phenotype of the case and confirmed that the case was a normal healthy female except for the recurrent pregnancy loss. Although some overlapping deletions were identified in the disease-associated structural variant databases such as ISCA (<https://www.iscaconsortium.org>) and DECIPHER (<http://decipher.sanger.ac.uk>),



**Figure 1** Cytogenetic analyses of the female patient examined in this study. **(a)** Partial karyotype showing a normal chromosome 2 and that with an interstitial deletion. The initial analysis showed the karyotype 46,XX,del(2)(q13q14.2), but the re-evaluation after microarray confirmed 46,XX,del(2)(q14.3q14.3). **(b)** Fluorescence *in situ* hybridization analysis of metaphase chromosomes. The yellow arrows indicate signals corresponding to RP11-11G20 (left, green) or 140B20 (right, green) located at 2q14.3. Red signals indicate the centromere of chromosome 2 (white arrows). RP11-11G20 shows a heterozygous deletion while RP11-140B20 is not deleted. **(c)** Cytogenetic array data. The left panel shows the whole chromosome 2 and the right panel shows the detail. The location of the probes are indicated at the right.



**Figure 2** Analyses of the breakpoints and junction of the 2.8-Mb deletion. **(a)** Deletion junction. Nucleotides in blue indicate the sequence of the proximal region, while those in black indicate the distal sequence. The sequences of 11–13-nucleotides repeated four times are underlined. Nucleotides in red or green are those participating in microhomology. Nucleotide positions depicted by arrowheads are occasionally mutated. Those in lowercase are the mutations. **(b)** Sequences of the proximal and distal breakpoint regions. Nucleotides depicted in lowercase are deleted. The six nucleotides in green are those commonly appearing in both proximal and distal regions, and used junction formation as microhomology.

we found no case with a recurrent pregnancy loss. Taken together, these observations led us to the supposition that the deletion might be benign.

To analyze the breakpoint of this deletion at a nucleotide resolution, multiple PCR primers were designed upstream and downstream of the putative breakpoint and long-range PCR was performed using one upstream primer and one downstream primer. One of the PCR primer pairs successfully yielded a PCR product that incorporated the deletion junction. At this junction, we found that a 11–13-nucleotide sequence, originally located at the proximal breakpoint region, was repeated four times with a one-nucleotide microhomology at the junction between each repeat (Figure 2a). Finally, the proximal and distal region was joined with a six-nucleotide microhomology. We found no repeat number variation manifesting as a polymorphism in the general population in the 1000 Genome database (<http://www.1000genomes.org>). Hence, the four copies of the 11–13-nucleotide repeat were a concurrent by-product of the *de novo* emergence of the 2.8-Mb deletion.

We further analyzed the sequence around the proximal and distal breakpoint regions (Figure 2b). The proximal breakpoint region was located within the LINE1 element, while no characteristic sequence was found around the distal breakpoint. We did not identify any non-B DNA motif that could have potentially induced replication fork stalling at either the proximal or distal breakpoint regions.<sup>11</sup>

## DISCUSSION

The female patient suffering from a recurrent pregnancy loss examined in this study was found to carry a 2.8-Mb deletion that included only one gene, *CNTNAP5*. *CNTNAP5* is a brain-specific gene encoding a member of the neurexin superfamily of unknown function. Although the deletion of *CNTNAP5* has been reported in some patients with intellectual disability or autism, the association between this deletion and these disorders is unclear.<sup>12,13</sup> It might be unlikely, however, that the deletion of *CNTNAP5* would affect female reproductive functions and the genetic basis

for the recurrent pregnancy loss of our study patient thus remained uncertain. Such a large deletion as seen in our patient can exist without any phenotypic abnormalities if the genes that are contained in the region in question is dispensable. A similar large deletion, del(2)(q13q14.1), has been reported previously in a woman with no phenotypic abnormalities,<sup>14</sup> although this deleted region does not overlap with the one identified in our current study.

Nearly 50% of reported deletions/duplications carry microhomology at the junctions, suggesting that these GCRs are generated via the replication-related pathways fork stalling and template switching or microhomology-mediated break-induced replication.<sup>9</sup> However, these terms are mostly defined on the basis of phenomenological findings of junction sequence. Single-strand nicks that arise before S-phase entry might trigger microhomology-mediated break-induced replication, but the biological evidence for this is still lacking.<sup>8,15</sup> Arlt *et al.*<sup>16</sup> designed an elegant experiment to demonstrate the involvement of replication stress in the generation of GCRs with microhomology. They cultured cells with aphidicolin and successfully induced *de novo* copy number abnormalities including both deletions and duplications. They also analyzed the junctions of these rearrangements and consistently found microhomology, which is analogous to human copy number abnormalities. Further, these rearrangements with microhomology have been observed even in non-homologous end joining-deficient cell lines.<sup>17</sup> These data may represent direct evidence that replication stress can induce microhomology-mediated GCRs.

Strikingly, we found in our current experiments that 11–13-nucleotide stretches were repeated four times at the junction of the deletion in our female subject. This observation is consistent with serial or backward replication slippage that has been proposed previously.<sup>18,19</sup> In addition, the presence of a base substitution at the same nucleotide in the repeats suggests that some modification of the nucleotides that could impede the progression of a replication fork may be a mechanism underlying the onset of the deletion. It has

been reported that several rounds of invasion, extension and dissociation are repeated in the template switching in break-induced replication.<sup>20</sup> In our current case, microhomology was observed not only between each repeat unit but also between the proximal and distal breakpoints, suggesting that a similar mechanism, that is, a microhomology-mediated restart of replication, finally bypassed the replication impediment leading to the deletion. It is possible that the proximal DNA end could invade a distal breakpoint region as far as 2.8 Mb away, as both regions might be in close proximity in the nucleus and be replicated concurrently.

An unresolved question that remained from our current analyses was the nature of the molecular pathway for DNA damage repair that is utilized in the development of replication stress-induced GCR. The presence of base substitutions within the nascent repeat sequence commonly observed in serial replication slippage might provide clues toward identifying this pathway.<sup>18,19</sup> When the replication fork encounters a damaged base or nucleotide in a leading-strand template, the damaged lesion would generally be bypassed in a homology-dependent manner using a nascent sister chromatid originating from the lagging-strand. However, in case *rad51* is unavailable or in short supply, error-prone translesion synthesis or error-free pathways based on replication fork regression and template switching by forming a chicken-foot structure would be activated. These pathways are mediated by monoubiquitination or polyubiquitination of proliferating cell nuclear antigen, respectively and are referred to as post-replication repair.<sup>21,22</sup> The error-prone translesion synthesis pathway is usually suppressed but another possible mechanism is the error-prone restart of DNA replication proposed recently.<sup>23</sup> When the replication fork stalls at sites of DNA damage, the microhomology-primed restart would be error prone possibly mediated by a DNA polymerase with low-processivity. Increased mutation rates during the replication of repeat regions might result from a similar mechanism.<sup>24,25</sup>

In conclusion, our current analysis of a female patient with recurrent pregnancy loss implicates the post-replication repair pathway as a mechanism underlying copy number variation in mammals. A full elucidation of the molecular pathway leading to serial/backward replication slippage deserves further investigation.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGEMENTS

We thank Dr Hiroshi Kogo and Dr Makiko Tsutsumi for helpful discussions and Mrs Narumi Kamiya for technical assistance. These studies were supported by a grant-in-aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan (21390101 and 24390085) and from the Ministry of Health, Labour and Welfare of Japan (10103465) to HK

- Beckmann, J. S., Estivill, X. & Antonarakis, S. E. Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. *Nat. Rev. Genet.* **8**, 639–646 (2007).
- Kurahashi, H., Bolor, H., Kato, T., Kogo, H., Tsutsumi, M., Inagaki, H. *et al.* Recent advance in our understanding of the molecular nature of chromosomal abnormalities. *J. Hum. Genet.* **54**, 253–260 (2009).
- Shaikh, T. H., Kurahashi, H. & Emanuel, B. S. Evolutionarily conserved low copy repeats (LCRs) in 22q11 mediate deletions, duplications, translocations, and genomic instability: an update and literature review. *Genet. Med.* **3**, 6–13 (2001).
- Stankiewicz, P. & Lupski, J. R. Genome architecture, rearrangements and genomic disorders. *Trends Genet.* **18**, 74–82 (2002).
- Baumer, A., Dutly, F., Balmer, D., Riegel, M., Tükel, T., Krajewska-Walasek, M. *et al.* High level of unequal meiotic crossovers at the origin of the 22q11.2 and 7q11.23 deletions. *Hum. Mol. Genet.* **7**, 887–894 (1998).
- Chapman, J. R., Taylor, M. R. & Boulton, S. J. Playing the end game: DNA double-strand break repair pathway choice. *Mol. Cell.* **47**, 497–510 (2012).
- Lee, J. A., Carvalho, C. M. & Lupski, J. R. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* **131**, 1235–1247 (2007).
- Hastings, P. J., Lupski, J. R., Rosenberg, S. M. & Ira, G. Mechanisms of change in gene copy number. *Nat. Rev. Genet.* **10**, 551–564 (2009).
- Conrad, D. F., Bird, C., Blackburne, B., Lindsay, S., Mamanova, L., Lee, C. *et al.* Mutation spectrum revealed by breakpoint sequencing of human germline CNVs. *Nat. Genet.* **42**, 385–391 (2010).
- Kidd, J. M., Graves, T., Newman, T. L., Fulton, R., Hayden, H. S., Malig, M. *et al.* A human genome structural variation sequencing resource reveals insights into mutational mechanisms. *Cell* **143**, 837–847 (2010).
- Cer, R. Z., Donohue, D. E., Mudunuri, U. S., Temiz, N. A., Loss, M. A., Starnier, N. J. *et al.* Non-B DB v2.0: a database of predicted non-B DNA-forming motifs and its associated tools. *Nucleic Acids Res.* **41**, D94–D100 (2013).
- Ballarati, L., Recalcati, M. P., Bedeschi, M. F., Lalatta, F., Valtorta, C., Bellini, M. *et al.* Cytogenetic, FISH and array-CGH characterization of a complex chromosomal rearrangement carried by a mentally and language impaired patient. *Eur. J. Med. Genet.* **52**, 218–223 (2009).
- Pagnamenta, A. T., Bacchelli, E., de Jonge, M. V., Mirza, G., Scerri, T. S., Minopoli, F. *et al.* Characterization of a family with rare deletions in CNTNAP5 and DOCK4 suggests novel risk loci for autism and dyslexia. *Biol. Psychiatry* **68**, 320–328 (2010).
- Sumption, N. D. & Barber, J. C. A transmitted deletion of 2q13 to 2q14.1 causes non-typic abnormalities. *J. Med. Genet.* **38**, 125–127 (2001).
- Hastings, P. J., Ira, G. & Lupski, J. R. A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS Genet.* **5**, e1000327 (2009).
- Arlt, M. F., Mülle, J. G., Schaibley, V. M., Ragland, R. L., Durkin, S. G., Warren, S. T. *et al.* Replication stress induces genome-wide copy number changes in human cells that resemble polymorphic and pathogenic variants. *Am. J. Hum. Genet.* **84**, 339–350 (2009).
- Arlt, M. F., Rajendran, S., Birkeland, S. R., Wilson, T. E. & Glover, T. W. *De novo* CNV formation in mouse embryonic stem cells occurs in the absence of Xrcc4-dependent nonhomologous end joining. *PLoS Genet.* **8**, e1002981 (2012).
- Chen, J. M., Chuzhanova, N., Stenson, P. D., Férec, C. & Cooper, D. N. Meta-analysis of gross insertions causing human genetic disease: novel mutational mechanisms and the role of replication slippage. *Hum. Mutat.* **25**, 207–221 (2005).
- Chen, J. M., Chuzhanova, N., Stenson, P. D., Férec, C. & Cooper, D. N. Complex gene rearrangements caused by serial replication slippage. *Hum. Mutat.* **26**, 125–134 (2005).
- Smith, C. E., Llorente, B. & Symington, L. S. Template switching during break-induced replication. *Nature* **447**, 102–105 (2007).
- Daigaku, Y., Davies, A. A. & Ulrich, H. D. Ubiquitin-dependent DNA damage bypass is separable from genome replication. *Nature* **465**, 951–955 (2010).
- Ulrich, H. D. & Walden, H. Ubiquitin signalling in DNA replication and repair. *Nat. Rev. Mol. Cell Biol.* **11**, 479–489 (2010).
- Mizuno, K., Miyabe, I., Schalbetter, S. A., Carr, A. M. & Murray, J. M. Recombination-restarted replication makes inverted chromosome fusions at inverted repeats. *Nature* **493**, 246–249 (2013).
- Shah, K. A., Shishkin, A. A., Voineagu, I., Pavlov, Y. I., Shcherbakova, P. V. & Mirkin, S. M. Role of DNA polymerases in repeat-mediated genome instability. *Cell Rep.* **2**, 1088–1095 (2012).
- Saini, N., Zhang, Y., Nishida, Y., Sheng, Z., Choudhury, S., Mieczkowski, P. *et al.* Fragile DNA motifs trigger mutagenesis at distant chromosomal loci in *Saccharomyces cerevisiae*. *PLoS Genet.* **9**, e1003551 (2013).