



HAL
open science

Resection and repair of a Cas9 double-strand break at CTG trinucleotide repeats induces local and extensive chromosomal deletions

Valentine Mosbach, David Viterbo, Stéphane Descorps-Declère, Lucie Poggi, Wilhelm Vaysse-Zinkhöfer, Guy-Franck Richard

► To cite this version:

Valentine Mosbach, David Viterbo, Stéphane Descorps-Declère, Lucie Poggi, Wilhelm Vaysse-Zinkhöfer, et al.. Resection and repair of a Cas9 double-strand break at CTG trinucleotide repeats induces local and extensive chromosomal deletions. PLoS Genetics, 2020, 16 (7), pp.e1008924. 10.1371/journal.pgen.1008924 . pasteur-03657831

HAL Id: pasteur-03657831

<https://pasteur.hal.science/pasteur-03657831>

Submitted on 3 May 2022

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

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



Distributed under a Creative Commons Attribution 4.0 International License


RESEARCH ARTICLE

Resection and repair of a Cas9 double-strand break at CTG trinucleotide repeats induces local and extensive chromosomal deletions

Valentine Mosbach¹ [✉], David Viterbo¹ [✉], Stéphane Descorps-Declère^{1,2} , Lucie Poggi^{1,3}, Wilhelm Vaysse-Zinkhöfer^{1,3} , Guy-Franck Richard¹ ^{*}

1 Institut Pasteur, CNRS, UMR3525, Paris, France, **2** Institut Pasteur, Center of Bioinformatics, Biostatistics and Integrative Biology (C3BI), Paris, France, **3** Sorbonne Universités, Collège doctoral, Paris, France

 These authors contributed equally to this work.

 Current address: Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), UMR7104 CNRS/Unistra, INSERM U1258, Illkirch, France

* gfrichar@pasteur.fr



 OPEN ACCESS

Citation: Mosbach V, Viterbo D, Descorps-Declère S, Poggi L, Vaysse-Zinkhöfer W, Richard G-F (2020) Resection and repair of a Cas9 double-strand break at CTG trinucleotide repeats induces local and extensive chromosomal deletions. *PLoS Genet* 16(7): e1008924. <https://doi.org/10.1371/journal.pgen.1008924>

Editor: Lorraine S. Symington, Columbia University, UNITED STATES

Received: December 6, 2019

Accepted: June 10, 2020

Published: July 16, 2020

Copyright: © 2020 Mosbach et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: V. M. was supported by Fondation Guy Nicolas and Fondation Hardy. W. V.-Z. is the recipient of a PhD fellowship from la Ligue Nationale Contre le Cancer. This work was generously supported by the Institut Pasteur and by the Centre National de la Recherche Scientifique (CNRS). The funders had no role in study design,

Abstract

Microsatellites are short tandem repeats, ubiquitous in all eukaryotes and represent ~2% of the human genome. Among them, trinucleotide repeats are responsible for more than two dozen neurological and developmental disorders. Targeting microsatellites with dedicated DNA endonucleases could become a viable option for patients affected with dramatic neurodegenerative disorders. Here, we used the *Streptococcus pyogenes* Cas9 to induce a double-strand break within the expanded CTG repeat involved in myotonic dystrophy type 1, integrated in a yeast chromosome. Repair of this double-strand break generated unexpected large chromosomal deletions around the repeat tract. These deletions depended on *RAD50*, *RAD52*, *DNL4* and *SAE2*, and both non-homologous end-joining and single-strand annealing pathways were involved. Resection and repair of the double-strand break (DSB) were totally abolished in a *rad50Δ* strain, whereas they were impaired in a *sae2Δ* mutant, only on the DSB end containing most of the repeat tract. This observation demonstrates that Sae2 plays significant different roles in resecting a DSB end containing a repeated and structured sequence as compared to a non-repeated DSB end. In addition, we also discovered that gene conversion was less efficient when the DSB could be repaired using a homologous template, suggesting that the trinucleotide repeat may interfere with gene conversion too. Altogether, these data show that *SpCas9* may not be the best choice when inducing a double-strand break at or near a microsatellite, especially in mammalian genomes that contain many more dispersed repeated elements than the yeast genome.

Author summary

With the discovery of highly specific DNA endonucleases such as TALEN and CRISPR-Cas systems, gene editing has become an attractive approach to address genetic disorders. Myotonic dystrophy type 1 (Steinert disease) is due to a large expansion of a CTG trinucleotide

data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors declare no competing interest.

repeat in the DMPK gene. At the present time, despite numerous therapeutic attempts, this dramatic neurodegenerative disorder still has no cure. In the present work, we tried to use the Cas9 endonuclease to induce a double-strand break within the expanded CTG repeat of the DMPK gene integrated in the yeast genome. Surprisingly, this break induced chromosomal deletions around the repeat tract. These deletions were local and involved non-homologous joining of the two DNA ends, or more extensive involving homologous recombination between repeated elements upstream and downstream the break. Using yeast genetics, we investigated the genetic requirements for these deletions and found that the triplet repeat tract altered the capacity of the repair machinery to faithfully repair the double-strand break. These results have implications for future gene therapy approaches in human patients.

Introduction

Microsatellites are short tandem repeats ubiquitously found in all eukaryotic genomes sequenced so far [1]. Altogether, they cover ~2% of the human genome, a figure similar to the whole protein-coding sequence [2]. Naturally prone to frequent repeat length polymorphism, some microsatellites are also prone to large expansions that lead to human neurological or developmental disorders, such as trinucleotide repeats involved in Huntington disease, myotonic dystrophy type 1 (Steinert disease), fragile X syndrome or Friedreich ataxia [3]. These expansion-prone microsatellites share the common property to form secondary DNA structures *in vitro* [4] and genetic evidence suggest that similar structures may also form *in vivo* [5,6], transiently stalling replication fork progression [7–11]. Among those, CCG/CGG trinucleotide repeats are fragile sites in human cells, forming frequent double-strand breaks when the replication machinery is slowed down or impaired [12]. Similarly, CAG/CTG and CCG/CGG microsatellites are also fragile sites in *Saccharomyces cerevisiae* cells [13,14]. Therefore, microsatellite abundance and the natural fragility of some of them make these repeated sequences perfect targets to generate chromosomal rearrangements potentially leading to cancer.

Double-strand break (DSB) repair mechanisms have been studied for decades in model organisms as well as in human cells and led to the identification of the main genes involved in this process [15]. Many of these advances were made possible by the use of highly specific DNA endonucleases, such as the meganucleases I-SceI or HO [16–18]. Other frequently used methods involved ionizing radiation making genome-wide DSBs [19]. However, the fate of a single double-strand break within a repeated and structured DNA sequence has never been addressed, until recently. In a former work, we used a TALE Nuclease (TALEN) to induce a unique DSB into a long CTG trinucleotide repeat integrated into a *S. cerevisiae* chromosome. TALEN are made of the fusion between a Transcription Activator-like Effector, a *Xanthomonas* family of modular transcription activators (TALE), and the FokI nuclease domain. We showed that 100% of yeast cells in which the TALEN was expressed exhibited a large contraction of the repeat tract, going from an initial length of ~80 CTG triplets to less than 35. *POL32*, *DNL4* and *RAD51* were shown to play no detectable role in repairing this DSB. On the contrary, *RAD50*, *RAD52* and *SAE2* were required for proper repair of the DSB, and a functional Sae2 protein was found to be essential for efficient DSB resection, suggesting that repeat contraction occurred by a single-strand annealing (SSA) process, involving preliminary resection of the break, followed by annealing of the two DSB ends carrying the repeat tract [20,21].

In the present work, we used the *Streptococcus pyogenes* Cas9 endonuclease (*SpCas9*) to induce a DSB within the same long CTG trinucleotide repeat integrated in the yeast genome. The break was made at the 3' end of the repeat tract (Fig 1A), using a guide RNA that targets

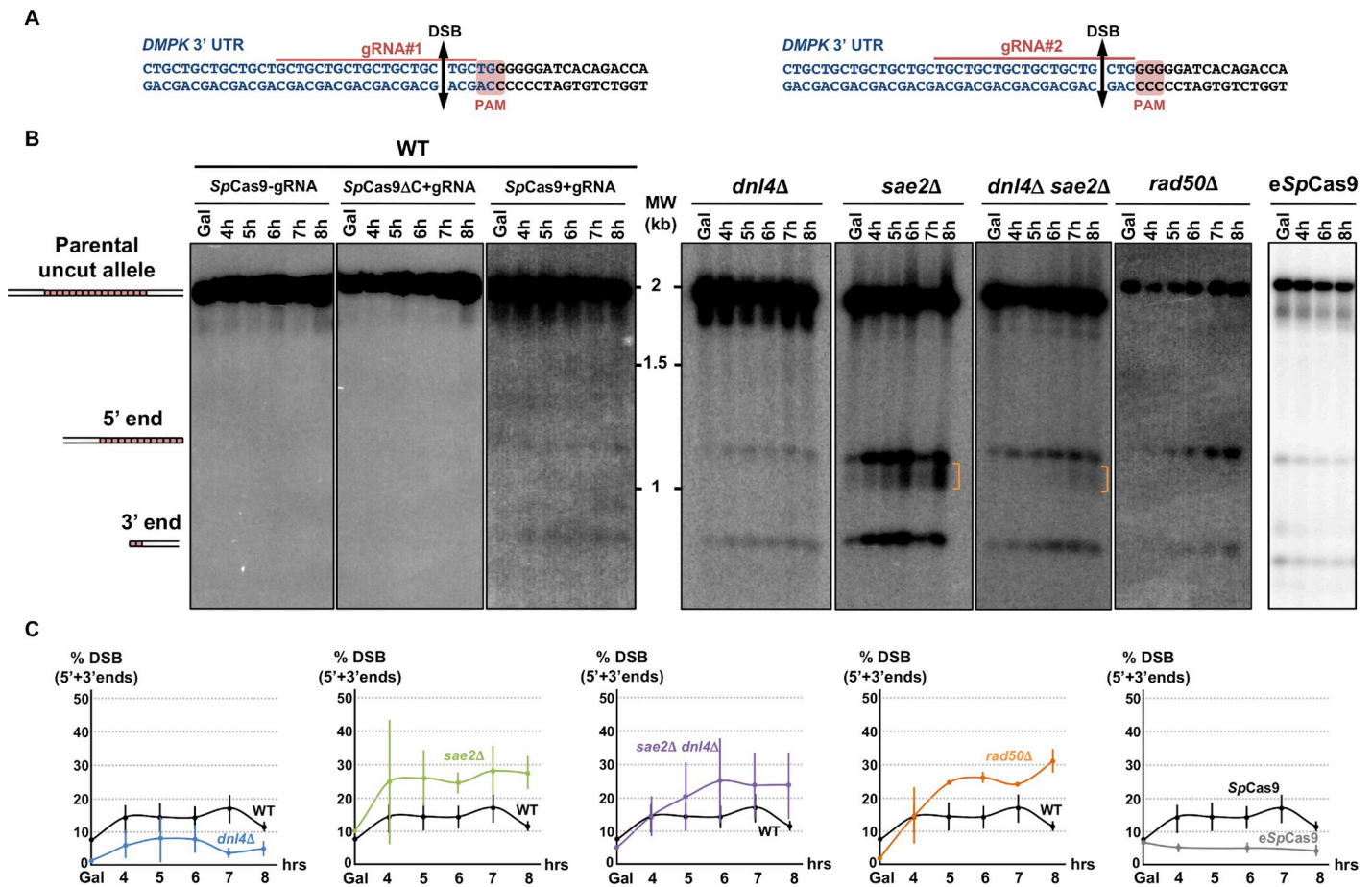


Fig 1. *SpCas9* and *eSpCas9* DSB induction in wild type and mutant strains. **A:** Sequence of the *SUP4::*(CTG)*n* locus. The CTG trinucleotide repeat tract comes from a human DM1 patient and is shown in blue. The flanking non-repeated DNA is in black. For each guide RNA, the PAM, the gRNA sequence as well as the expected DSB site are indicated. **B:** Southern blots of yeast strains during time courses. Lanes labeled *SpCas9*-gRNA and *SpCas9*ΔCter+gRNA are control strains in which no DSB was visible. In the strain expressing both *SpCas9* and the gRNA, two bands are visible in addition to the parental allele (1966 bp). One band corresponds to the 3' end of the DSB containing a small number of triplets (821 bp), the other one corresponds to the 5' end of the DSB containing most of the repeat tract (1145 bp). **C:** Quantification of 5' and 3' DSB signals. For each time points, the total 5' + 3' signals were quantified and plotted as a ratio of the total signal in the lane. Three independent time courses were run in each strain background (except *rad50*Δ and *eSpCas9* for which two time courses were run) and plots show the average of three (or two) time courses. Error bars are 95% confidence intervals.

<https://doi.org/10.1371/journal.pgen.1008924.g001>

the repeat tract. Frequent rearrangements were found in surviving cells, with local deletions as well as more extensive ones involving recombination between retrotransposon LTRs. Survival and repair depended on *RAD50*, *RAD52*, *SAE2* and *DNL4* and double-strand break resection was abolished in *rad50*Δ and *sae2*Δ mutants. A more specific version of the nuclease, Enhanced *SpCas9*, generated the same rearrangements. In addition, we also discovered that gene conversion was less efficient when *SpCas9* was used to induce a DSB within a CTG repeat tract that could be repaired with a homologous template, suggesting that the trinucleotide repeat may interfere with gene conversion too.

Results

A Cas9-induced double-strand break within CTG repeats induces cell death and chromosomal deletions around the repeat tract

In previous work, we showed that a TALEN targeted at the CTG trinucleotide repeat from the human *DMPK* gene 3' UTR integrated in a yeast chromosome, was extremely efficient at

contracting the repeat tract below the pathological length [20,21]. In order to determine whether the CRISPR-Cas9 system could be used in the same manner, a plasmid-borne *Streptococcus pyogenes* Cas9 nuclease (*SpCas9*) was expressed in *Saccharomyces cerevisiae* from a *GALI*-inducible promoter [22]. The same plasmid also carried a CTG guide RNA (hereafter named gRNA#1) under the control of the constitutive *SNR52* promoter. The PAM used in this experiment was the TGG sequence located right at the border between the CTG tract and non-repeated DNA (Fig 1A). As controls, we used the same *SpCas9*-containing plasmid without the gRNA or a frameshift mutant of the *SpCas9* gene resulting in a premature stop codon (*SpCas9* Δ NdeI) and the gRNA#1. The same genetic assay as previously described was used [20,21]. It is based on a modified suppressor tRNA gene (*SUP4*) in which a CTG trinucleotide repeat was integrated. The length of the CTG repeat at the start of the experiment was determined to be approximately 80 triplets. Four hours after transition from glucose to galactose medium, two faint bands were visible on a Southern blot, corresponding to the 5' and 3' ends of the *SpCas9* DSB. No signal was detected in control strains (Fig 1B). The DSB was quantified to be present in ca. 10%-15% of the cells at any given time point and remained the same for the duration of the time course. No evidence for repeat tract contraction was visible. Survival to the *SpCas9* break was low (17.9% \pm 4%), as calculated from CFU on galactose plates over CFU on glucose plates (Fig 2, see Materials & Methods). Surviving colonies were picked, total

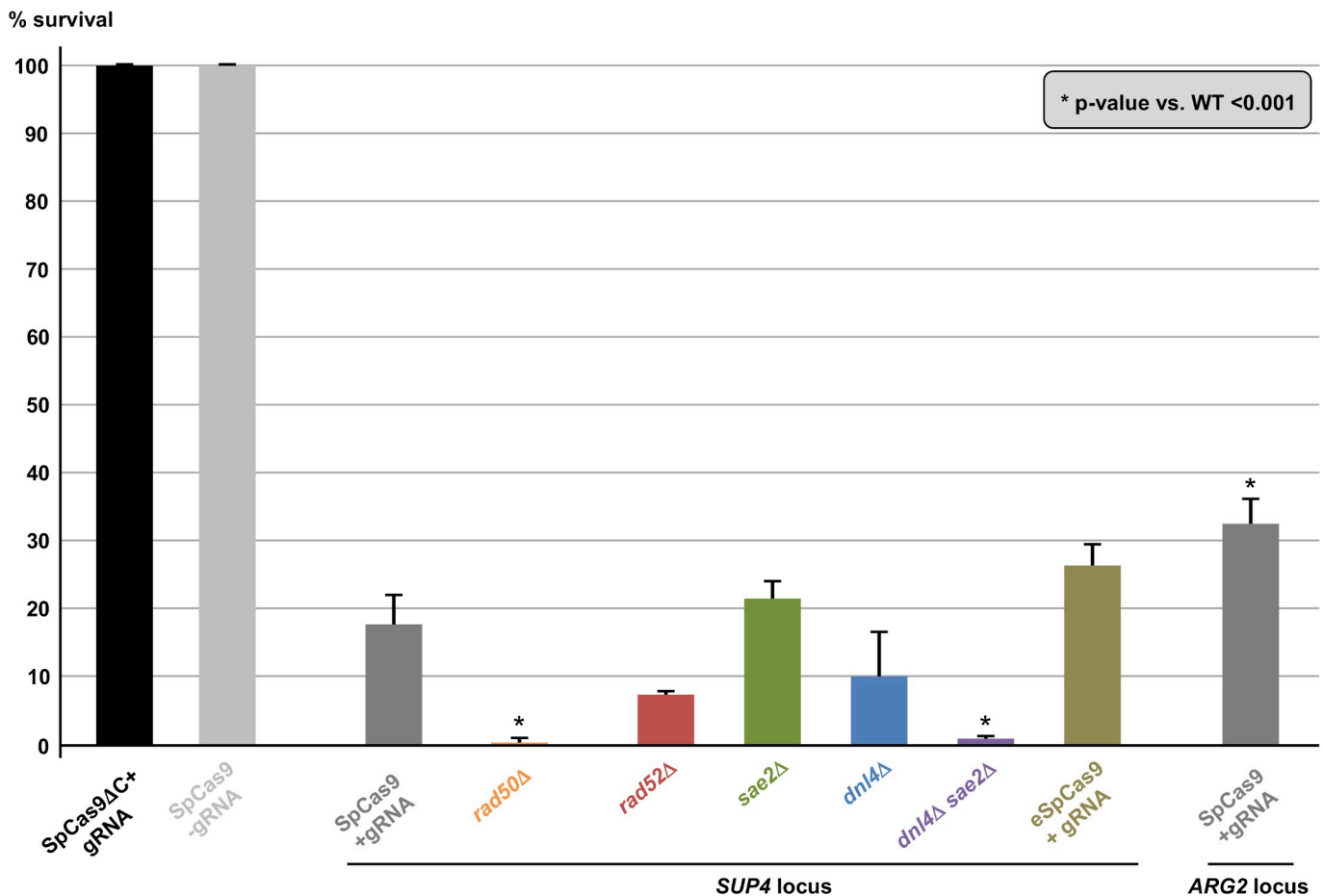


Fig 2. Yeast survival to Cas9 induction. For each strain, the same number of cells were plated on galactose and glucose plates and the survival was expressed in CFU number on galactose plates over CFU number on glucose plates. The mean and the 95% confidence interval are plotted for each strain. Significant t-test p-values when compared to wild-type *SpCas9* survival are indicated by asterisks. See Materials & Methods for statistics.

<https://doi.org/10.1371/journal.pgen.1008924.g002>

genomic DNA was extracted and the *SUP4::CTG* locus was analyzed by Southern blot. Patterns observed were remarkably different among clones, most of them showing bands of aberrant molecular weight, either much larger or much shorter than the repeat tract. In some clones, a total absence of signal suggested that the probe target was deleted and in other cases weakness of the signal was compatible with a partial deletion (Fig 3A). To understand these

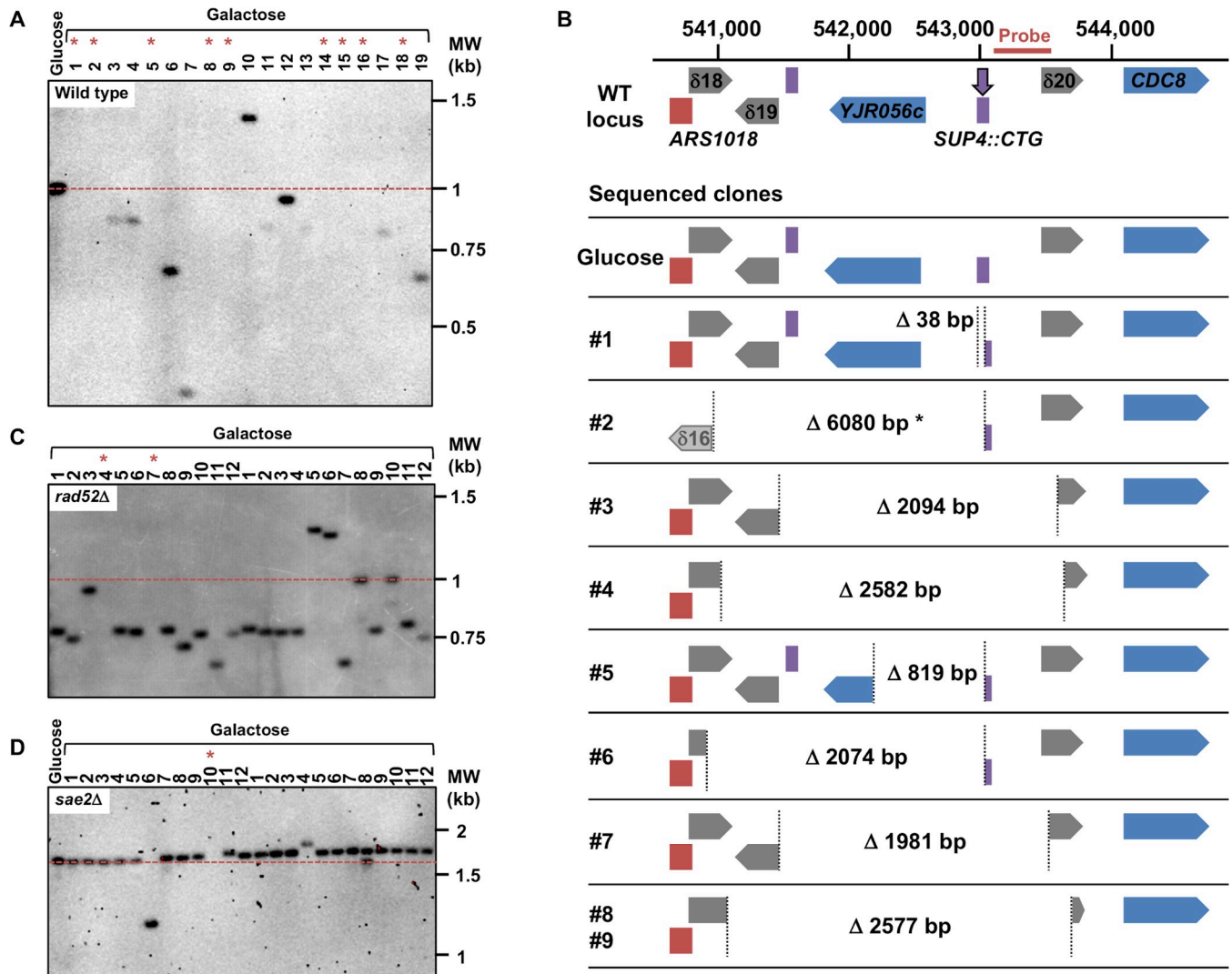


Fig 3. Chromosomal rearrangements following *SpCas9* induction. **A:** Southern blot of genomic DNA at the *SUP4* locus in the wild-type strain. The probe hybridizes ~300 bp downstream the repeat tract (see Fig 3B). The dotted red line shows the initial length of the CTG repeat tract. The lane labeled "Glucose" contains a clone in which Cas9 was not induced. Lanes numbered #1 through #19 contain independent clones in which Cas9 was induced. Asterisks point to lanes in which no signal was detected, meaning that the probe containing sequence was deleted. Note that signal intensities varies among lanes, showing that the probe did not fully bind to its target sequence, due to its partial deletion. **B:** Some examples of chromosome rearrangements following Cas9 induction in the wild-type strain. The genomic locus surrounding *SUP4* is shown on top, *ARS1018* is drawn in red, delta elements are in grey, protein-coding genes are colored in blue and tRNA genes in purple. The DSB (vertical purple arrow) is induced within *SUP4::(CTG)_n*. Chromosome coordinates are indicated above and the probe used for hybridization is represented by an horizontal red bar. The locus sequence was retrieved from the Saccharomyces Genome Database (<http://yeastgenome.org/>, genome version R64-2-1, released 18th November 2014). Under the reference locus are cartooned the different chromosomal structures observed in some of the survivors. A yeast colony that was grown in glucose was also sequenced as a control. For each clone, vertical dotted lines represent junctions of rearrangements observed, with deletion sizes indicated in base pairs. Asterisk: clone #2 showed a complex rearrangement with a local inverted duplication involving the $\delta 16$ LTR and the 3' end of the *KCH1* gene 5 kb upstream *SUP4*. Two clones (#8 and #9) exhibit exactly the same chromosomal deletion at precisely the same nucleotides. Note that *CDC8* is an essential gene. **C:** Southern blot of genomic DNA at the *SUP4* locus in the *rad52* Δ strain. Legend as for Fig 3A. **D:** Southern blot of genomic DNA at the *SUP4* locus in the *sae2* Δ strain. Legend as for Fig 3A. Note that for this Southern blot genomic DNA was digested with EcoRV (instead of Ssp I, see Methods), therefore the expected CTG repeat length was around 1.8 kb, instead of 1 kb.

<https://doi.org/10.1371/journal.pgen.1008924.g003>

abnormal patterns, the genome of nine surviving clones were totally sequenced by paired-end Illumina. As a control, one clone in which *SpCas9* had not been induced was also sequenced. In all nine cases, a deletion around the repeat tract was found, extending from a few nucleotides to several kilobases (Fig 3B). Some of the deletions involved flanking Ty1 retrotransposon LTRs, and in one case (clone #2), a complex event between a distant LTR ($\delta 16$) and the $\delta 20$ LTR close to the repeat tract was detected. Following this discovery, total genomic DNA was extracted from more surviving colonies, analyzed by Southern blot and deletion junctions were amplified by PCR and Sanger sequencing. Two different sets of primers were used to amplify the junction, su47/su48 allowing to amplify local deletions around the repeat tract, whereas su23/su42 were used to amplify larger deletions between Ty1 LTRs (Fig 4). Surviving colonies were classified into 12 different types, according to the *SUP4::CTG* locus after *SpCas9* induction: type I corresponded to a colony in which the repeat tract was unchanged (or was slightly expanded), type II to a colony in which a repeat contraction occurred, type III-V corresponded to local deletions around the repeat tract, types VI-XI to more extensive deletions, and type XII to a complex event between another δ element and $\delta 20$. A few examples of junction sequences are shown in S1 Fig.

Chromosomal deletions are under the control of *RAD50*, *RAD52*, *DNL4* and *SAE2*

We next decided to investigate the role of several genes known to be involved in DSB repair on chromosomal deletions generated by the *SpCas9* nuclease. In a *rad52 Δ* strain, in which all homologous recombination was abolished, survival decreased but was not significantly different from wild type ($7.6\% \pm 0.7\%$, Fig 2). Molecular analysis of the survivors by Southern blot showed that deletions seemed to be less extensive than in the wild-type strain, fewer lanes showing a partial or total absence of signal (Fig 3C). Junction sequencing confirmed that deletions between Ty LTRs were lost (Type VI events), except for two cases in which the deletion occurred through annealing of eight or nine nucleotides and was therefore *RAD52* independent (Fig 4 and S1 Fig, clones #C4 and #C7). This result showed that about 50% of colonies growing on galactose plates survived the DSB by *RAD52*-dependent homologous recombination between two LTR elements flanking the trinucleotide repeat tract.

The possible role of non-homologous end-joining (NHEJ) in the observed deletions was also addressed by deleting the gene encoding yeast Ligase IV (*DNL4*). In the *dnl4 Δ* strain, the level of detected DSBs was slightly lower than in wild type (Fig 1B and 1C). Survival was decreased, but not significantly different from wild type ($10.5\% \pm 6.3\%$, Fig 2). Molecular analysis of the survivors showed that local deletions were totally lost, whereas extensive deletions involving Ty LTR represented 84% of all events (Fig 4 and S1 Fig). Hence, we concluded that all local deletions were NHEJ dependent.

In a recent work, we showed that *SAE2* was essential to repair a DSB induced by a TALEN within a long trinucleotide repeat. In its absence, unrepaired breaks accumulated and DSB resection was lost on the trinucleotide repeat-containing end [20]. We therefore tested the effect of a *sae2 Δ* mutation on a *SpCas9* DSB in the same experimental system. Southern blot analysis of repair intermediates showed that DSB ends accumulated twice as much in the *sae2 Δ* mutant as compared to wild type (Fig 1B and 1C). In addition, a smear was detected below the 5' DSB end (Fig 1B, orange bracket), hallmark of an incomplete resection triggering a repair defect [23]. Survival was similar to wild type ($21.5\% \pm 2.9\%$, Fig 2). Southern blot analysis of surviving colonies displayed very little size changes as compared to uninduced controls (Fig 3D). However, sequencing showed that the most frequent event was an insertion (or sometimes a small deletion) of one to eight nucleotides between the PAM and the repeat tract

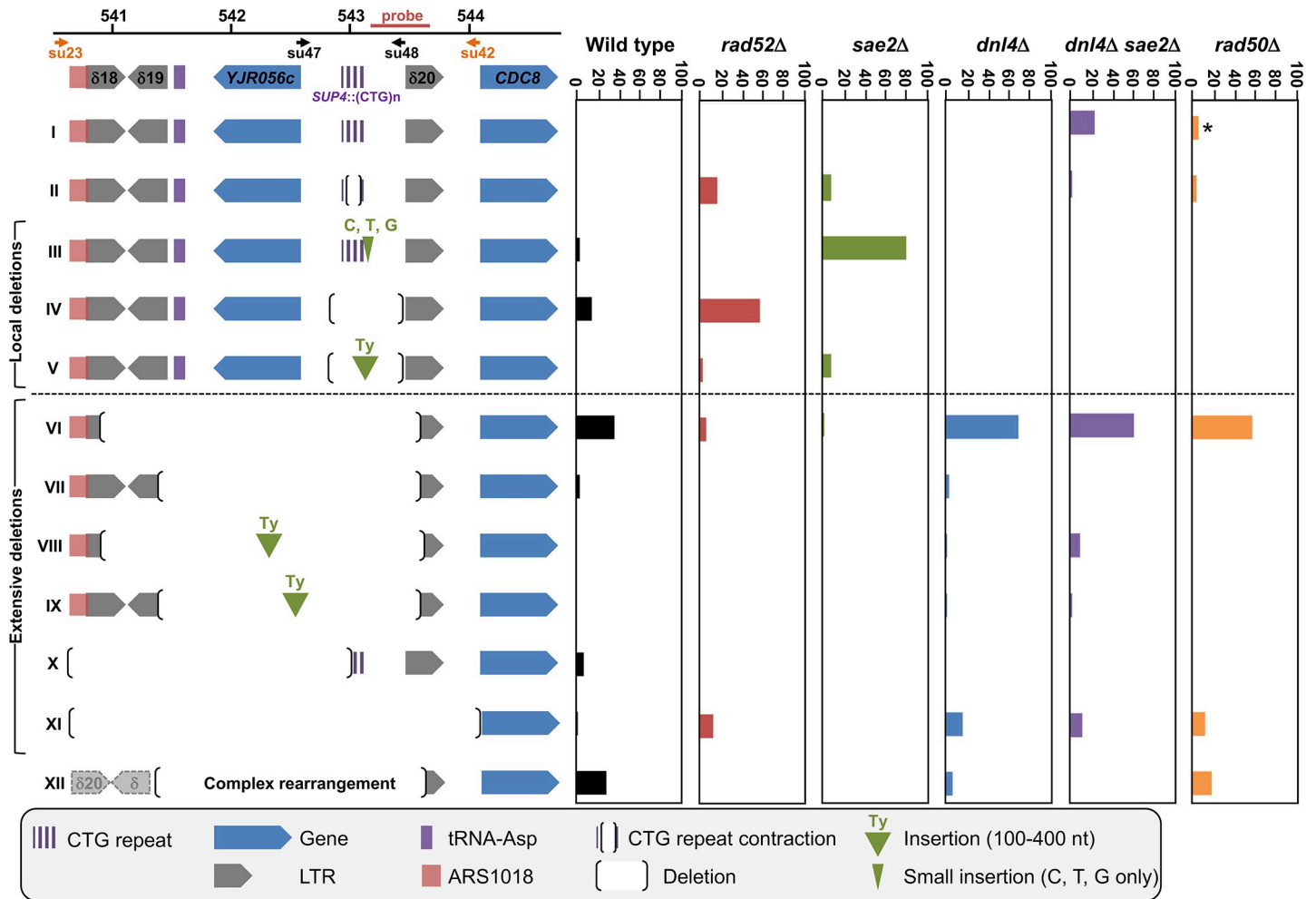


Fig 4. Summary of chromosomal rearrangements observed in wild-type and mutant strains, following SpCas9 induction. **Left:** The twelve different possible outcomes following SpCas9 induction are shown, subdivided in local and extensive deletions (see text for details). The SUP4 locus is pictured and shows the position of each genetic element on yeast chromosome X. The probe used on Southern blots is shown, as well as both primer couples used to amplify the locus. In order to assess a given clone to a rearrangement type, the following rules were followed: i) when a band was detected by Southern blot, primers su47 and su48 were used to amplify the locus and sequence it. These events corresponded to types I-V. The absence of a PCR product indicated that primer su48 genomic sequence was probably deleted and therefore primers su23 and su42 were used to amplify and sequence the locus. These were classified as types IV-V events; ii) when no band was detected by Southern blot, primers su 23 and su42 were directly used to amplify and sequence the locus. These events were classified as types VI-X and XII. When no PCR product was obtained, it meant that at least one of the two primers genomic sequence was probably deleted and these events were classified as type XI. Note that this last category may also contain rare -but possible- chromosomal translocations that ended up in putting each primer in a separate chromosome, making unobtainable the PCR product. The extent of type XI deletions cannot go downstream the su42 primer, since the CDC8 gene is essential. **Right:** The proportion of each type or event recovered is represented for wild type and mutants. The asterisk near Type I events in the rad50Δ strain indicates that some of them were small expansions (see text). Altogether, 262 surviving clones were sequenced, distributed as follows: WT: 51, rad52Δ: 29, dnl4Δ: 61, sae2Δ: 32, dnl4Δ sae2Δ: 47, rad50Δ: 42.

<https://doi.org/10.1371/journal.pgen.1008924.g004>

(Type III events, Fig 4). These local insertions represented 78% of all survivors, whereas only one Ty LTR recombination (Type VI) was detected (S1 Fig). This result showed that in the absence of SAE2, long range deletions were lost, probably due to the inability to resect the DSB into single-stranded DNA prone for homologous recombination.

The double mutant sae2Δ dnl4Δ was also built and showed an additive effect on survival, with a significant 30-fold reduction in CFU on galactose plates (0.6%±0.9%, Fig 2). This result proved that in the absence of one of the two genes repair could occur by the other pathway, but absence of both genes was almost lethal to yeast cells receiving a SpCas9 DSB. Southern analysis showed that DSB levels were similar to sae2Δ levels (ca. 24% after 8 hrs versus 28% for

sae2Δ), showing that *SAE2* was epistatic to *DNL4*. The smear corresponding to resection defects was also visible (Fig 1B, orange bracket). Interestingly, 21% of survivors exhibited zero to two triplets lost, which could be attributed to natural microsatellite instability. These were classified as Type I events and were specific of the *sae2Δ dnl4Δ* double mutant (Fig 4 and S1 Fig). It is possible that given the low survival rate, cells in which *SpCas9* and/or the gRNA was mutated were positively selected during the time course in liquid culture and were therefore subsequently recovered on galactose plates. Remarkably, to the exception of the Type I events hereabove mentioned, all but one event corresponded to extensive deletions around the repeat tract, similarly to the single *dnl4Δ* mutant.

Finally, in a *rad50Δ* strain, the DSB accumulated over the duration of the time course at levels similar to *sae2Δ* mutants (Fig 1B and 1C). No smear was detected in this strain background, suggesting that the *sae2Δ* resection defect was specific of this mutant and did not involve the integrity of the MRX-Sae2 complex. Survival was very low in this strain background ($0.3\% \pm 0.4\%$), significantly different from wild type but not from the *sae2Δ dnl4Δ* double mutant (Fig 2). Survivor analysis showed a few repeat contractions (Type II), but most events were deletions between LTRs (Type VI), large deletions (Type XI) or complex rearrangements (Type XII), a pattern not significantly different from what was observed with the *sae2Δ dnl4Δ* double mutant (Fig 4). However, uniquely present in this strain background, were found three repeat expansions, two of them associated with the insertion of a 'C' in the first triplet preceding the PAM, most probably inhibiting Cas9 recognition and cutting (classified as Type I events in Fig 4, see S1 Fig). This suggests that repeat expansions are more frequent in a *rad50Δ* mutant in the presence of a DSB, as previously observed [24]. Such expansions were not recovered in the *sae2Δ dnl4Δ* double mutant, showing that this strain phenotype does not recapitulate exactly the *rad50Δ* phenotype, or that the rather limited number of survivors analyzed was not sufficient to detect a small number of expansions.

In conclusion, when a *SpCas9* DSB was induced into a long CTG trinucleotide repeat, cell survival was low and depended on *RAD50*, *RAD52*, *SAE2* and *DNL4*. Two classes of repair events were found: local deletions under the control of *RAD50* and *DNL4* and therefore the NHEJ pathway, and extensive deletions under the control of *SAE2* and *RAD52*. In addition, the deletion of *RAD50* almost completely recapitulated the *sae2Δ dnl4Δ* double mutation, except that the smear was not visible on Southern blots and a few expansions were recovered.

Enhanced *SpCas9* generates the same chromosomal deletions as *SpCas9*

Over the last four years, several mutants of the widely used *SpCas9* have been engineered or selected by genetic screens. *SpCas9*-HF1 and e*SpCas9* were built to exhibit less off-target DSBs [25,26], HypaCas9 was made to be even more accurate [27], Sniper-Cas9 also showed reduced off-target effects [28], while evoCas9 was selected in yeast for improved specificity [29]. We decided to explore the possibility that chromosomal deletions observed in our experimental system were partly due to the fact that *SpCas9* exhibited a high off-target activity on long CTG trinucleotide repeat tract, perhaps by generating more than one DSB within the repeat tract, or within the surrounding loci. In order to test this hypothesis, Enhanced *SpCas9* (e*SpCas9*) was expressed in yeast, along with the same guide RNA as previously (gRNA #1, Fig 1A). Survival was slightly higher than with *SpCas9* ($26.3\% \pm 3.0\%$), but not significantly different (t test p-value = 0.06). DSB end accumulation was lower than *SpCas9* (Fig 1B and 1C). Molecular analysis of surviving yeast cells did not show any statistical difference between types of deletions observed with e*SpCas9* as compared to *SpCas9* (Chi2 p-value = 0.14) (Fig 5 and S1 Fig). A second guide RNA (gRNA#2) was designed, so that the DSB would be made two nucleotides

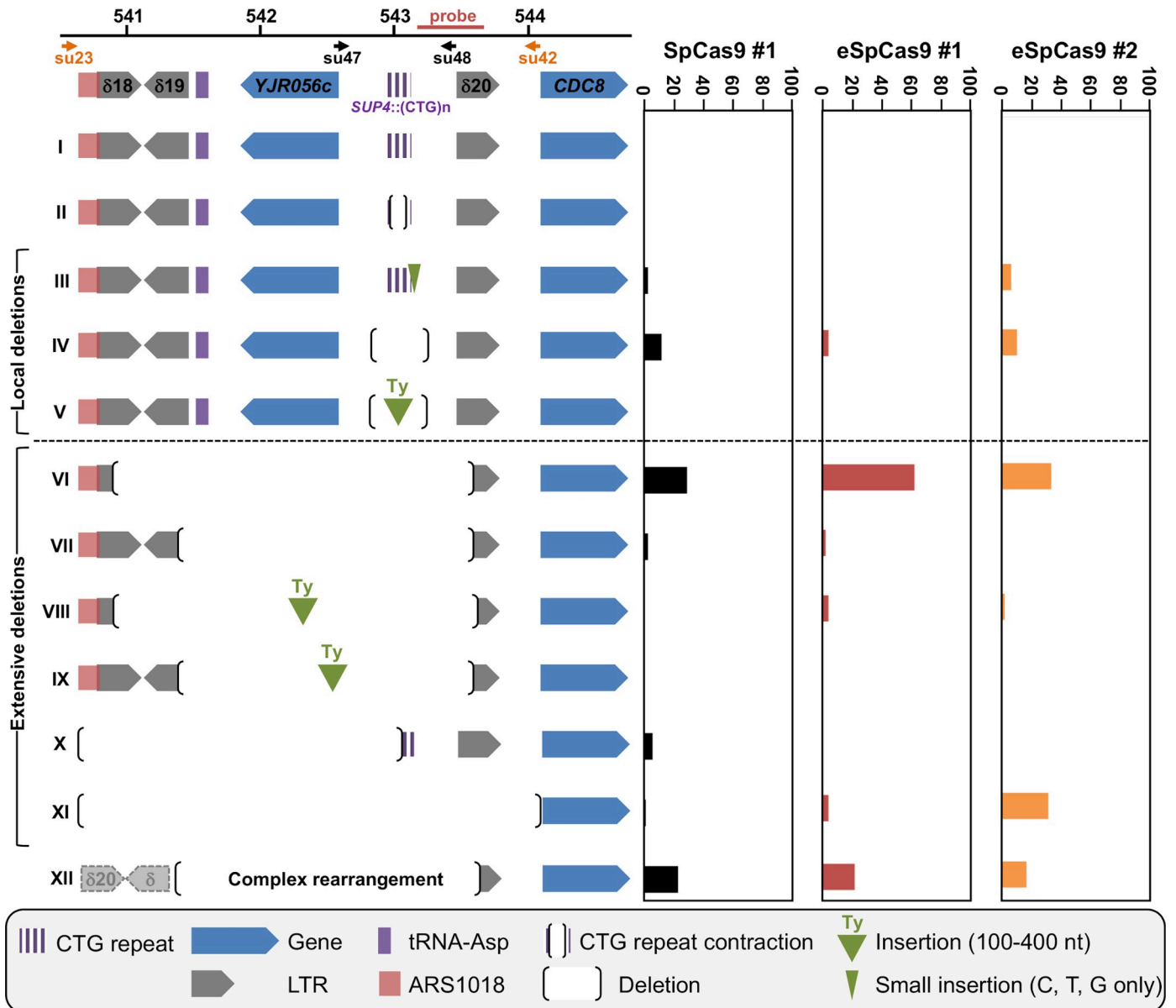


Fig 5. Summary of chromosomal rearrangements observed in following *SpCas9* and enhanced *SpCas9* (e*SpCas9*) inductions. See Fig 4 for legend. e*SpCas9* #1 and #2 refer respectively to guide RNAs #1 and #2 described in Fig 1A. The following number of surviving clones were analyzed: *SpCas9*: 51, e*SpCas9* #1: 49, e*SpCas9* #2: 48. See text for details.

<https://doi.org/10.1371/journal.pgen.1008924.g005>

closer to the repeat tract end (Fig 1A). Interestingly, the number of deletions involving a LTR (Types VI-IX) was lower than with gRNA#1 (35% with gRNA#2 vs 63% with gRNA#1) but the proportion of very large deletions (Type XI) significantly increased from 4% to 31% (Chi2 p-value = $1.6 \cdot 10^{-3}$). We concluded that moving the DSB cut site two nucleotides toward non-repeated DNA increased the outcome of very large deletions. Altogether, these results show that using a more specific version of *SpCas9* did not decrease chromosomal rearrangements, suggesting that deletions seen with *SpCas9* were probably not due to extra off-target DSBs within the CTG repeat tract or the surrounding loci.

Gene conversion efficacy is decreased when a Cas9 DSB is made within a long CTG trinucleotide repeat

Gene conversion is a very efficient DSB-repair mechanism in *S. cerevisiae*. We previously showed that a single DSB induced by the *I-SceI* meganuclease in a yeast chromosome was efficiently repaired using a CTG repeat-containing homologous template as a donor [24,30,31]. In order to determine whether a Cas9-induced DSB within a CTG repeat was properly repaired by the recombination machinery, we reused a similar experimental system in which two copies of the *SUP4* allele were present on yeast chromosome X, one containing a (CTG)₆₀ repeat tract and the other copy containing an *I-SceI* recognition site (Fig 6A). In this ectopic gene conversion assay, 80.2%±2.3% of yeast cells survived after an *I-SceI* DSB and 100% of survivors were repaired by gene conversion using the ectopic *SUP4*::CTG copy as a donor [31]. When *SpCas9* was induced in the same yeast strain along with gRNA#1, only 32.6%±3.8% of CFU formed on galactose plates (Fig 2). Molecular analysis of surviving cells showed that 89% (34 out of 38) repaired by ectopic gene conversion, as expected, and now contain two *I-SceI* recognition sites, one in each *SUP4* copy (Fig 6B, GC events). However, one expansion event was also detected, as well as one local deletion (Type IV) and two events involving a deletion and a DNA insertion (Type V). Intriguingly, the DNA insertion was a 211 bp piece of DNA from the *YAK1* gene, located 158 kilobases upstream the *ARG2* locus, on chromosome X left arm. This gene contains a long and imperfect CAG/CTG repeat within its reading frame, like many yeast genes [32–35]. An unusual recombination event occurred between the *YAK1* CAG/CTG repeat and the *ARG2* repeat, leading to a chimeric repeat (Fig 6C). This rearrangement may be

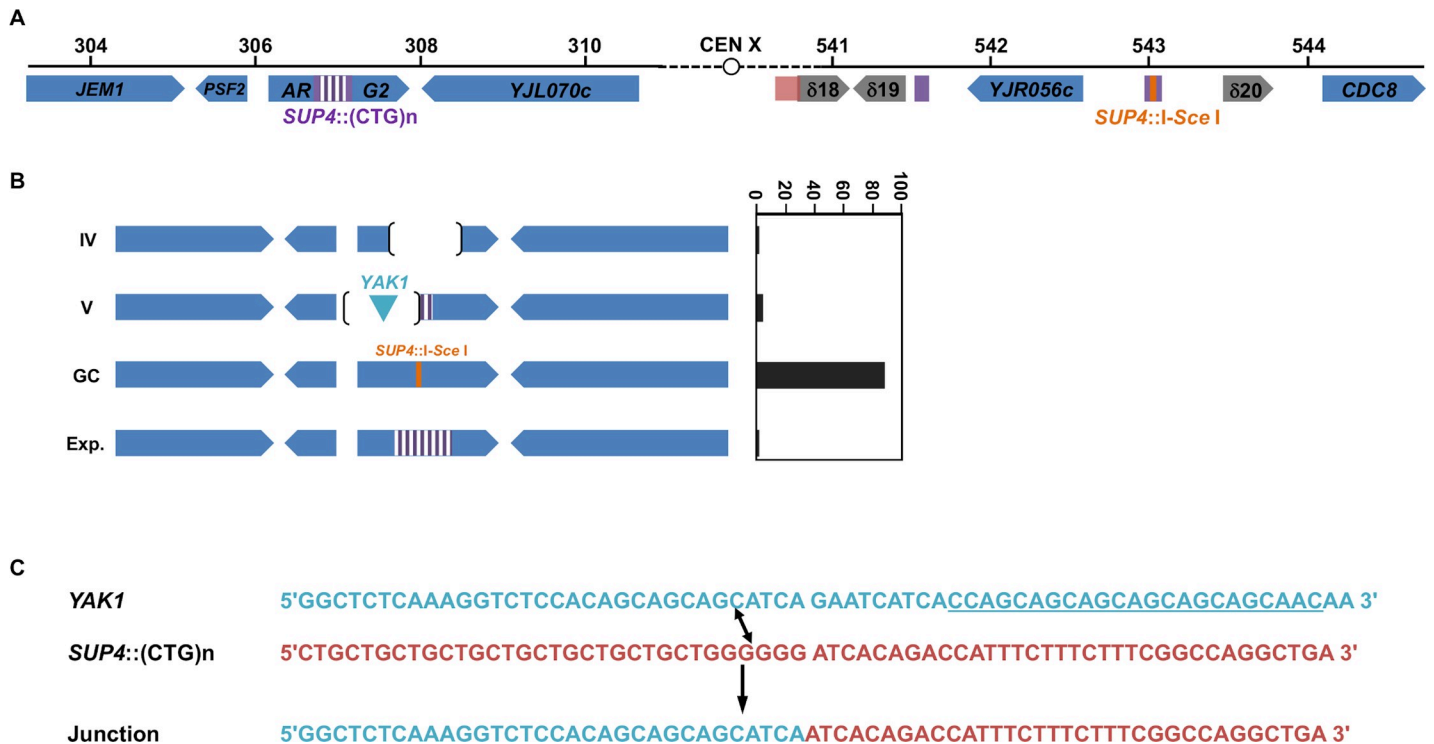


Fig 6. Chromosomal rearrangements observed at the ARG2 locus, following *SpCas9* induction. A: *ARG2* and *SUP4* loci drawn to scale. A 2.6 kb piece of DNA containing 1.8 kb of the *SUP4* locus in which a CTG repeat was integrated, as well as the *TRP1* selection marker were integrated at *ARG2* [31]. The *TRP1* gene is not represented here but is centromere-proximal located. B: Types of rearrangements observed. Types IV and V deletions are explained in Fig 5. GC: gene conversion with *SUP4*::I-SceI. Exp.: CTG repeat expansion. C: Type V rearrangements involving the *YAK1* gene. The off-target in *YAK1* identified by CRISPOR is underlined. The blue sequence from *YAK1* recombined with the red sequence from *SUP4* to give a hybrid molecule called "Junction".

<https://doi.org/10.1371/journal.pgen.1008924.g006>

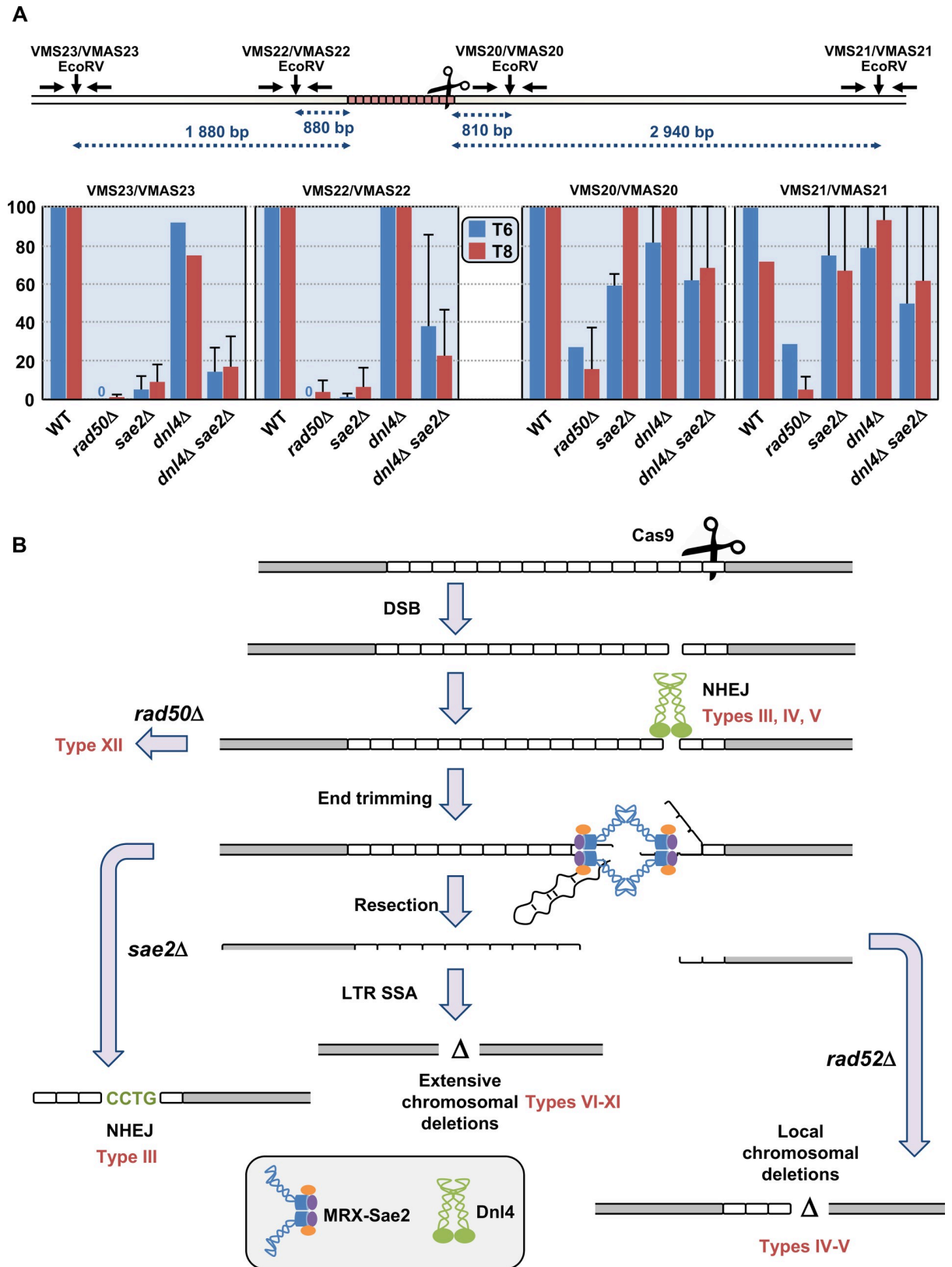
the result of an off-target DSB generated by *SpCas9* within the *YAK1* repeat, or an abnormal recombination event between the two CTG repeats following *SpCas9* induction. Using the CRISPOR *in silico* tools, off-target sites were examined for guide RNA #1 [36]. Two off-targets with zero mismatches were predicted, in the *NGR1* and *SGF73* genes. However, since they are using a non-canonical PAM (TGA), the Cutting Frequency Determination score (CFD score, ref. [37]) was very low at these two sites, suggesting that they would be poor substrate for *SpCas9* with this gRNA (S1 Table). When one mismatch was allowed, two hits were found in the *YAK1* gene, one of them with a canonical NGG PAM exhibiting the best CFD score of all predicted off-target sites (S1 Table). We therefore decided to check whether an off-target DSB could be induced within the *YAK1* gene by *SpCas9*. A time course was performed in conditions in which the nuclease was non-induced or induced, and the resulting Southern blot was hybridized with a *YAK1* specific probe. No evidence for a band that could correspond to a DSB at this locus could be seen. Faint signals were detected, both in non-induced and induced conditions, at molecular weights that did not fit the expected DSB size (S2 Fig). We concluded that, if an off-target DSB was made by *SpCas9* at the *YAK1* locus, it was too rare to be detected by Southern blot, and presumably could not influence cell survival, nor be sufficient to trigger frequent ectopic recombination events with the *SUP4* locus.

Resection of a Cas9-induced double-strand break

Quantitative PCR experiments were performed in order to determine the resection level in strains in which Cas9 was induced. The nuclease generates a DSB in the very last CTG triplets of the repeat tract (Fig 1A). Therefore, the 5' end of the break contains most of the 80 triplets whereas the 3' end contains only two triplets. This asymmetry allows to compare resection of a repeated and structured DNA end versus non-repeated DNA, concomitantly and in the same experimental setting. We took advantage of the convenient position of four *EcoRV* restriction sites, two on each side of the DSB, at different distances from the break (Fig 7A). Primers were designed in such a way that *EcoRV* digested DNA could not be PCR amplified. However, if DNA resection reached an *EcoRV* site, the resulting single-stranded DNA became resistant to digestion and therefore susceptible to amplification. In wild-type cells after eight hours, resection of the Cas9 DSB was always 100% at all *EcoRV* sites, except at the 3' distal site in which it was a little lower, around 70% (Fig 7A). In *dnl4Δ* cells, resection was not statistically different from wild type. In the *rad50Δ* mutant, DSB resection was totally abolished on the 5' end of the break that contains most of the repeat tract and severely impaired on the other end, showing that the MRX-Sae2 complex was essential on both DSB ends. Interestingly, the *sae2Δ* mutant exhibited a resection defect on the 5' end of the break but not on the other side. This was also true for the double mutant *sae2Δ dnl4Δ*. All these data prove that: i) Ligase IV plays no role in DSB resection; ii) Sae2 is essential to resect a long CTG trinucleotide repeat but is dispensable to resect a non-repeated DSB end.

Genome-wide mutation spectrum in cells expressing *SpCas9*

When carefully looking at deletion borders in haploid strains in which *SpCas9* was induced, they were found to be more extensive on the 5' side of the break than on the 3' side (Fig 4). This suggests that larger 3' deletions encompassing the essential gene *CDC8* or its promoter may not have been recovered because they would be lethal, probably counting for some of the lethality observed. In order to check this hypothesis, we expressed *SpCas9* in diploids containing *SUP4::(CTG)_n* repeat tracts on both homologues. In these cells, both chromosomes could be cut by the nuclease. We quantified by qPCR *CDC8* copy number in six independent diploid survivors. In all cases, it was reduced by half as compared to a control qPCR on another



resection as compared to the total DSB amount detected on Southern blots are shown at 6 hours (in blue) and 8 hours (in red), along with standard deviations. **B:** Mechanistic model for chromosomal deletions following a Cas9-induced DSB. See text for details. Resulting deletion types are indicated in red near each pathway.

<https://doi.org/10.1371/journal.pgen.1008924.g007>

chromosome (S3A Fig). This showed that in these cells, only one of the two *CDC8* alleles was present, suggesting that the other was often deleted during DSB repair. In order to check if the whole chromosome could have been lost, we also amplified a region near the *JEM1* gene, on the other chromosomal arm, near the *ARG2* gene. Surviving clones showed a significantly higher signal, compatible with the presence of two chromosomes (Mann-Whitney-Wilcoxon rank test, p -value = 10^{-3}). Therefore, it was concluded that the mortality observed in haploid cells expressing *SpCas9* was at least partly due to the frequent deletion of the essential *CDC8* gene, but did not induce significant chromosome loss.

In order to detect possible off-target mutations, independent haploid and diploid colonies in which *SpCas9* had been induced were deep-sequenced, using Illumina paired-end technology. In diploid cells, two nucleotide substitutions were detected out of five independent clones when *SpCas9* was repressed (S3B Fig). When the nuclease was induced, six mutations were detected out of 18 sequenced clones, a similar proportion. One 36-bp deletion was found in the *FLO11* minisatellite and two deletions of one repeat unit were found in AT dinucleotide repeats, but no mutation was found in any other CAG/CTG repeat tract. Altogether, we concluded that *SpCas9* expression in diploid cells did not significantly increase genome-wide mutation frequency. The genome of 10 independent haploid cells in which *SpCas9* was induced was also completely sequenced. Eight mutations were detected among six of these survivors, all of them being nucleotide substitutions in non-repeated DNA (S3B Fig). This is statistically not different from what was observed in diploids (Fisher exact test p -value = 0.12). We concluded that besides chromosomal deletions around the *SUP4* locus observed in these haploids, *SpCas9* did not induce other mutations in yeast cells, to a level detectable with the present sequencing approach.

Discussion

SpCas9-induced DSB repair within CTG repeats generates chromosomal deletions

In previous work, in which we induced a DSB within a CTG repeat using a dedicated TALEN, 100% of surviving yeast colonies repaired the break by contracting the repeat tract (S4 Fig and ref. 21). These contractions occurred by single-strand annealing, and depended on *RAD52*, *RAD50* and *SAE2*, but was independent of *LIG4*, *POL32* and *RAD51* [20]. It was therefore striking and completely unexpected that a DSB made by the *SpCas9* nuclease (or by its more specific mutant version, *eSpCas9*) at exactly the same location within the very same CTG trinucleotide repeat induced frequent chromosomal deletions around the repeat tract and almost no repeat contraction.

With the TALEN, repeat contraction was proposed to be an iterative phenomenon, involving several rounds of cutting and contraction until the repeat tract was too short for the two TALEN arms to dimerize and induce a DSB [20,21]. A similar outcome was expected with *SpCas9*, iterative rounds of cutting and contraction could occur until the remaining CTG repeat tract would be too short for the gRNA to bind and induce a DSB. But this was not observed here, surprisingly proving that a *SpCas9*-induced DSB was differently repaired from a TALEN-induced DSB targeting the same exact repeated sequence. Previous terminal-transferase mediated PCR of four TALEN-induced DSB showed that the break left 2–11 bp of

homology, probably due to variable positioning of the left TALEN arm on the repeated sequence [20]. In comparison, Cas9 should not have any binding flexibility and should leave 5 bp of homology each time it cuts. We therefore think this is unlikely to explain the difference between the two nucleases, since even though the TALEN sometimes leave very little homology, no chromosomal rearrangement was observed. Reasons for this discrepancy may include different DSB ends (4 nucleotides 5' overhangs with the TALEN, blunt ends with Cas9), different substrate-enzyme kinetics of the TALEN as compared to Cas9, a role for the guide RNA in maintaining both ends together after cutting, or differences in checkpoint activation, these hypotheses being neither exhaustive nor mutually exclusive. These possibilities are now being investigated. Note that the use of PolQ-mediated micro homologies for end joining is very efficient in human cells and similar strategies might be more effective for contracting repeats in patient cells.

Spontaneous homologous recombination events between delta elements surrounding *SUP4* were already described by the past by Rothstein and colleagues [38]. Recombination between $\delta 18/\delta 20$ or $\delta 19/\delta 20$ was less frequent than between $\delta 16/\delta 20$ or $\delta 17/\delta 19$. This was the opposite in our experiments. This is probably due to the way recombination was triggered. In the Rothstein *et al.* article, recombination was spontaneous, therefore favoring LTRs with the highest sequence identity ($\delta 16/\delta 20$ or $\delta 17/\delta 19$). In our case, the initiating event was a DSB always at the same position, between $\delta 19$ and $\delta 20$. Resection occurred until the closest regions of homology were revealed, hence favoring $\delta 18/\delta 20$ or $\delta 19/\delta 20$ over recombination with the more distant $\delta 16$ or $\delta 17$. In addition, in the present case, Cas9-induced deletions also involved microhomology sequences or no homology at all, suggesting that repair depended on the initiating damage (replication-induced single-strand nicks vs. nuclease-induced double-strand breaks). Our data are also reminiscent of a previous work in which spontaneous deletions around the *URA2* gene were classified in seven different classes, six of them harboring microhomologies at their junctions and one showing no obvious homology [39].

In recent work, using a GFP reporter system in human cells, Cinesi and colleagues showed that *SpCas9* induced contractions as well as expansions of CTG trinucleotide repeats, whereas the nickase mutant *SpCas9-D10A* only induced contractions. By small pool-PCR analysis, a four-fold increase in the rate of CTG repeat tract contractions was observed when the nickase was expressed with a CTG-carrying gRNA, as compared to the expression of the nuclease alone ([40], Fig 2C). The sp-PCR experiment was not performed when the double-strand endonuclease *SpCas9* was expressed, so it is not possible to determine whether a similar increase in contractions would be observed. When individual clones were sequenced, local deletions of the triplet repeat tract were found in three clones out of 17 (17.6%) when a CAG-carrying gRNA was used, and two clones out of 11 (18.2%) when a CTG-carrying gRNA was expressed ([40], S3 Fig). These numbers were not statistically different from what was observed in the absence of the *SpCas9-D10A* nickase ([40], S1 Fig). There is no report of repeat tract sequencing when *SpCas9* was expressed, so it is not possible to know if the double-strand endonuclease would make more (or less) deletions in this experimental system. In our present experiments, eight clones out of 51 (15.7%) analyzed contained a local deletion (Types IV events, Fig 4) similar to those described by Cinesi and colleagues. However, it must be noted that in their work, only the top 1% brightest GFP-positive cells were sequenced, whereas no selection was used in our experiments; surviving colonies were randomly picked and analyzed at the molecular level. Therefore, the absolute frequency of local deletions cannot be reliably compared between the two experimental setups.

It must be noted that an approach using the *SpCas9-D10A* nickase was successfully implemented to delete the CAG trinucleotide repeat involved in Huntington disease, by making two single-strand nicks, upstream and downstream the repeat tract [41].

In another work looking at the effect of a Cas9-induced DSB at the *LYS2* locus in *S. cerevisiae*, the authors found frequent *POLA4*-dependent small insertions (1–3 bp) in 42–68% of the survivors (depending on the PAM used) and local deletions (1–17 bp) in the remaining cases. However, given that there is no transposon or transposon remnant in the close proximity of the *LYS2* locus, the authors could not retrieve LTR deletions [42]. This strongly suggests that deletions observed heavily depend on the surrounding chromosomal location where the DSB is made.

CTG trinucleotide repeats interfere with SpCas9-triggered gene conversion

When an *I-SceI* DSB was induced within a *SUP4* allele, the break could be repaired by gene conversion with a CTG repeat-containing homologous donor at the *ARG2* locus. All yeast cells repaired by gene conversion with the donor, generating repeat contractions and expansions in the process [31]. Here, the exact reverse reaction was induced, the break was made within CTG repeats and repaired with a non-repeated sequence. DSB repair was much less efficient, since only 32.6% of the cells survived (Fig 2) and less specific since 10% of the repair events were unfaithful recombination (Fig 6B). This shows that when a Cas9 DSB was made into a CTG repeat, gene conversion was partially impaired, either by the repeat tract or by the Cas9 protein, or by both. Note that, in past experiments, when the DSB was made in the *I-SceI* recognition site, less than 10 nucleotides needed to be resected on each side of the break before homology with the other *SUP4* allele could be reached. In our present experiments, when the DSB was made within the CTG repeat tract, 64 nucleotides needed to be resected on one side and 304 on the other side of the break before homology with the other *SUP4* allele could be reached. This could also explain the better survival observed in the former case.

Ligase IV and Sae2 are respectively driving local and extensive chromosomal deletions

Yeast Ligase IV is encoded by the *DNL4* gene and is the enzyme used to ligate DSB ends during non homologous end-joining [43]. It was previously shown that *RAD50* and *SAE2* were essential to resect and process a TALEN-induced DSB but a *DNL4* deletion had no effect on break processing, cell survival or repair efficacy [20]. On the contrary, repair of a Cas9, an HO or an *I-SceI* DSB at the *MAT* locus, in the absence of any homologous donor cassette, was shown to be dependent on the product of the *DNL4* gene [42,44]. *SpCas9* DSB repair has also been studied in human cells in the presence of a drug (NU7441), acting as a chemical inhibitor of non-homologous end-joining. In these conditions, the frequency of single-base insertions and small deletions decreased whereas larger deletions increased, suggesting that these repair events occurred by an alternative end-joining mechanism (alt-EJ/MMEJ) involving micro-homologies flanking the DSB [45,46]. Here, we showed that when *DNL4* was inactivated, local deletions were totally lost. However, survival was not significantly decreased because yeast cells could repair the DSB using LTR recombination, generating extensive deletions around the repeat tract (Fig 4). Supporting this model, the absence of any resection defect in the *dnl4Δ* mutant demonstrated that in the absence of end-joining, resection may take place very efficiently to repair the DSB by homologous recombination, using flanking homologies.

SAE2 is associated to the *MRE11-RAD50-XRS2* complex, whose roles are multiple during DSB repair [47] and it was proposed to encode an endonuclease activity essential to process DNA hairpins [48], as well as to resect *I-SceI* double-strand breaks [49]. We previously showed that it was essential to resect a TALEN-induced DSB end containing a long CTG trinucleotide repeat, but less important to resect the non-repeated end [20]. In the present experiments, extensive deletions involving LTR elements were lost in a *sae2Δ* mutant, and 97% of yeast cells

repaired the DSB by local deletions, most of them resulting in insertions or deletions between the PAM and the gRNA sequence (Fig 4), inactivating *SpCas9* capacity to induce another DSB. Small insertions of a few nucleotides were also frequently detected following *SpCas9* DSB induction at the VDJ locus in human B cells [50] or at the *MAT* locus in *S. cerevisiae* [42]. However, in our experiments, all nucleotides inserted were C, T or G, all three encoded by the gRNA. No insertion of an adenosine residue was found out of 28 insertions sequenced (S1 Fig). This intriguing observation suggests the possibility that the gRNA could be used as a template to repair the DSB, as it was demonstrated that a single-stranded RNA could be used to repair an HO-induced DSB into the *LEU2* gene [51].

Although *DNL4* and *SAE2* trigger different types of chromosomal deletions and none of the single mutants significantly decreased survival, the *dnl4Δ sae2Δ* double mutant abolished repair, like the *rad50Δ* mutant, since only 0.6% of the cells survived (Fig 2), showing the synthetic effect of both mutations. However, repair events in the double mutant were similar to those observed in *dnl4Δ* (Fig 4). It is possible that other nucleases, like *EXO1* or *DNA2*, could perform long range resection in the absence of *SAE2* [49,52], allowing the occurrence of extensive deletions (Types VI-XI).

In order to exclude the possibility that differences between wild type and mutants could be due to lower or higher expression of Cas9 in different backgrounds, the amount of Cas9 protein in each strain as compared to glucose-6-phosphate dehydrogenase (G6PDH, the product of the *ZWF1* gene in budding yeast) was determined by Western blots. G6PDH is one of the most abundant proteins in budding yeast, with ca. 14,000 molecules per cell (*Saccharomyces Genome Database*, <https://www.yeastgenome.org/locus/S000005185>). Cas9 level showed a small 2-fold decrease in *dnl4Δ*, *sae2Δ* and *dnl4Δ sae2Δ* strains, as compared to wild type, but no difference was observed in the *rad50Δ* mutant (S5 Fig). We concluded that the amount of *SpCas9* in all strains was comparable to G6PDH. Therefore, phenotypic differences observed in mutants could not be due to a much higher or much lower expression of *SpCas9* in these strains, as compared to wild type.

All these results are compatible with a model in which a *SpCas9* DSB was tentatively repaired by NHEJ first (Fig 7B). In the absence of *RAD50*, the MRX-Sae2 complex could not assemble, resection could not occur, NHEJ was impossible and DSB ends were lost, triggering high mortality. If *DNL4* was inactivated, resection proceeded normally and when reaching flanking LTRs, repair occurred by *RAD52*-mediated SSA. In the absence of this gene, the break was repaired by *RAD52*-independent local deletions. When *SAE2* was inactivated, resection was impeded on the 5' DSB end, leading to resection defects observed as smears on Southern blots. Mutagenic NHEJ was favored, leading to local insertions and deletions. It is unknown whether Sae2 would play the same essential role on other secondary structure-forming trinucleotide repeats, like GAA or CGG triplets, or if its activity is specific to CTG triplets, hence of a structure rather than a repeat, but this important question is currently under investigation.

Materials and methods

Yeast strains and plasmids

All mutant strains were built from strain GFY6162-3D by classical gene replacement method [53], using *KANMX4* or *HIS3* as marker (S2 Table). *KANMX4* cassettes were amplified from the EUROSCARF deletion library, using primers located 1kb upstream and downstream the cassette. *VMS1/VMAS1* were used to amplify *rad52Δ::KANMX*, *VMS2/VMAS2* were used to amplify *rad51Δ::KANMX*, *VMS3/VMAS3* were used to amplify *pol32Δ::KANMX*, *VMS4/VMAS4* were used to amplify *dnl4Δ::KANMX*, *VMS6/VMAS6* were used to amplify *rad50Δ::KANMX* and *SAE2up/SAE2down* were used to amplify *sae2Δ::KANMX* (S3 Table). *VMY350*

and VMY352 strains were respectively used to construct VMY650 and VMY352 by mating-type switching, as follows: the pJH132 vector [54] carrying the HO endonuclease under the control of an inducible *GALI-10* promoter was transformed in the haploid *MAT α* strains. After 5h of growth in lactate medium, HO expression was induced by addition of 2% galactose (final concentration) and grown for 1.5 hour. Cells were then plated on YPD and mating type was checked three days later by crosses with both *MAT α* and *MAT α* tester strains.

For *SpCas9* inductions, addgene plasmid #43804 containing the nuclease under the control of the *Gall* promoter and the *LEU2* selection marker was digested with *HpaI* and cloned into yeast by homology-driven recombination [55] with a single PCR amplified fragment containing the *SNR52* promoter, the gRNA#1 and the *SUP4* terminator, using primers SNR52Left and SNR52Right (S3 Table) to give plasmid pTRi203. A frameshift was then introduced in this plasmid by *NdeI* digestion followed by T4 DNA polymerase treatment and religation of the plasmid on itself, to give plasmid pTRi206. In this plasmid, the *SpCas9* gene is interrupted by a stop codon after amino acid Ile₁₆₁. The haploid GFY6162-3D strain (or its mutant derivatives), was subsequently transformed with pTRi203 or pTRi206 and transformants were selected on SC-Leu. The plasmid containing Enhanced *SpCas9* (version 1.1, Addgene #71814, Slaymaker et al., 2016) was a generous gift of Carine Giovannangeli from the *Museum National d'Histoire Naturelle*. The e*SpCas9* gene was amplified using primers LP400 and LP401 (S3 Table) and cloned into yeast cells in the Addgene#43804 plasmid digested with *BamHI*, by homology-driven recombination, with 34-bp homology on one side and 40-bp homology on the other side [55], to give plasmid pLPX11. For the gRNA#1, plasmid pLPX11 was digested with *HpaI* and cloned into yeast by homology-driven recombination [55] with a single PCR amplified fragment containing the *SNR52* promoter, the gRNA#1 and the *SUP4* terminator, as above to give plasmid pTRi207. For the gRNA#2, a guide RNA cassette was ordered from ThermoFisher (GeneArt), flanked by *EcoRI* sites and was cloned in pRS416 [56] using standard procedures to give plasmid pLPX210.

In silico simulations of off-target sites

To assess the number of off-target sites for *SpCas9* in *Saccharomyces cerevisiae*, online tools were used. CRISPOR is a software that evaluates the specificity of a guide RNA through an alignment algorithm that maps sequences to a reference genome to identify putative on- and off-target sites [57]. To predict off-target sites, the online tool sequentially introduces changes in the sequence of the gRNA and checks for homologies in the specified genome [36]. The Cutting Frequency Determination score (CFD score) relies on several criteria to assess off-target probabilities: nucleotide deletions, insertions, mismatches, as well as the position and identity of the mismatch(es) [37].

Cas9 inductions

Before nuclease induction, Southern blot analyses were conducted on several independent subclones to select one containing ca. 80 CTG triplets. For Cas9 inductions, yeast cells were grown overnight at 30°C in liquid SC-Leu medium, then washed with sterile water to remove any trace of glucose. Cells were split in two cultures, half of the cells were grown in synthetic -Leu medium supplemented with 2% galactose (final concentration) and the other half were grown in synthetic -Leu medium supplemented with 2% glucose (final concentration). Around 4×10^8 cells were collected at different time points (T = 0, 4, 5, 6, 7 and 8 hours) and killed by addition of sodium azide (0.01% final). Cells were washed with water, and frozen in dry ice before DNA extraction. To determine survival to Cas9 induction, 24 hours after the T0 time point, cells were diluted to an appropriate concentration, then plated on SC-Leu plates

containing either 20 g/l glucose or galactose. After 3–5 days of growth at 30°C, ratio of CFU on galactose plates over CFU on glucose plates was considered to be the survival rate.

Double-strand break analysis and quantification

Total genomic DNA (4 µg) of cells collected at each time point was digested for 6h with EcoRV (40 U) (NEB) loaded on a 1% agarose gel (15x20 cm) and run overnight at 1 V/cm. The gel was vacuum transferred in alkaline conditions to a Hybond-XL nylon membrane (GE Healthcare) and hybridized with two randomly-labeled probes specific of each side of the repeat tract, upstream and downstream the *SUP4* gene [58]. After washing, the membrane was overnight exposed to a phosphor screen and signals were read and quantified on a FujiFilm FLA-9000. For the *YAK1* Southern blot, DNA was digested with BamHI and the probe was a 719 bp probe amplified with YAK1f and YAK1r primers (S3 Table).

SUP4 locus analysis after Cas9 induction

Several colonies from each induced or repressed plates were picked, total genomic DNA (4 µg) was extracted with Zymolyase, digested for 6h by SspI (20 U) (NEB), loaded on a 1% agarose gel (15x20 cm) and run overnight at 1V/cm. The gel was vacuum transferred in alkaline conditions to a Hybond-XL nylon membrane (GE Healthcare) and hybridized with a randomly-labeled PCR fragment specific of a region downstream the *SUP4* gene, amplified from the su8-su9 primer couple (S3 Table). After washing, the membrane was overnight exposed on a phosphor screen and signals were revealed on a FujiFilm FLA-9000. Genomic DNA of each clone for which a signal was detected by Southern blot was subsequently amplified with su47-su48 primers and sequenced using su47 (S3 Table). Genomic DNA of clones for which no signal was detected by Southern blot of no PCR product was obtained with su47-su48 were subsequently amplified with su23-su42 primers and sequenced using su42 (S3 Table). Sanger sequencing was performed by GATC biotech.

Western blots

Liquid cultures of each strain were grown to exponential phase in synthetic glucose medium or synthetic galactose medium without leucine, in order to maintain the Cas9 plasmid. Proteins were extracted on 2×10^8 cells in 200 µl Laemmli solution with 100 µl glass beads. Proteins were separated on a 10% acrylamide gel in standard conditions and blotted to a nitrocellulose membrane (Optitran BA-S 83 reinforced NC, Schleicher & Schuell). For Cas9 detection, a monoclonal HRP-conjugated mouse antibody was used (Abcam [7A9-3A3], ab202580, 1/1000 dilution). Note that blocking was achieved in 10 mM Tris-HCl pH8.0, NaCl 150 mM, 0.05% Tween 20, 3% dry milk (instead of the regular 5%). For G6PDH detection, the primary antibody was a polyclonal rabbit antibody (Sigma-Aldrich (A9521), 1/100 000 dilution). A secondary goat anti-rabbit antibody conjugated to horseradish peroxidase was used for detection of G6PDH (Thermo Scientific, 0.16 µg/ml final concentration). Quantification was performed using a ChemiDoc MP Imager (Bio-Rad) with the dedicated Image Lab software. The molecular weight marker used was the Precision Plus Protein Kaleidoscope marker (Bio-Rad).

Analysis of Cas9-induced DSB end resection by qPCR

A real-time PCR assay, using primer pairs flanking EcoRV sites 0.81 kb and 2.94 kb away from the 3' end of the CTG repeat tract (VMS20/VMAS20 and VMS21/VMAS21 respectively) and 0.88 kb and 1.88 kb away from the 5' end of the CTG repeat tract (VMS22/VMAS22 and VMS23/VMAS23 respectively), was used to quantify end resection. Another pair of primers

was used to amplify a region of chromosome X near the *ARG2* gene [11], to serve as an internal control of DNA amount (JEM1f-JEM1r). Genomic DNA of cells collected at T = 0h, T = 6h and T = 8h was split in two fractions, incubated at 80°C for 10 minutes in order to inactivate any remaining active DNA nuclease, then one fraction was used for EcoRV digestion and the other one for a mock digestion in a final volume of 15 µl. Samples were incubated for 5h at 37°C, then the enzyme was inactivated for 20 min at 80°C. DNA was subsequently diluted by adding 55 µl of ice-cold water, and 4 µl was used for each real-time PCR reaction in a final volume of 25 µl. PCRs were performed with the Absolute SYBR Green Fluorescein mix (Thermo Scientific) in the Mastercycler S realplex (Eppendorf), using the following program: 95°C 15min, 95°C 15sec, 55°C 30 sec, 72°C 30 sec repeated 40 times, followed by a 20 min melting curve. Reactions were performed in triplicates and the mean value was used to determine the amount of resected DNA, using the following formula: raw resection = $2/(1+2^{\Delta Ct})$ with $\Delta Ct = C_{t,EcoRV} - C_{t,mock}$. Relative resection values were calculated by dividing raw resection values by the percentage of DSB quantified at the corresponding time point.

The same protocol was used to determine the relative amount of *CDC8* and chromosome X in surviving clones after Cas9 induction, except that total genomic DNA was not digested prior to real-time PCR. Primer couples VMS23-VMAS23 were used to amplify *CDC8* and JEM1f-JEM1r for chromosome X left arm. Primers Chromo4_f and Chromo4_r were used to amplify a region of chromosome IV as an internal control for total DNA amount. See S3 Table for all primer sequences.

Library preparation for deep-sequencing

Approximately 10 µg of total genomic DNA was extracted and sonicated to an average size of 500 bp, on a Covaris S220 (LGC Genomics) in microtubes AFA (6x16 mm) using the following setup: Peak Incident Power: 105 Watts, Duty Factor: 5%, 200 cycles, 80 seconds. DNA ends were subsequently repaired with T4 DNA polymerase (15 units, NEBiolabs) and Klenow DNA polymerase (5 units, NEBiolabs) and phosphorylated with T4 DNA kinase (50 units, NEBiolabs). Repaired DNA was purified on two MinElute columns (Qiagen) and eluted in 16 µl (32 µl final for each library). Addition of a 3' dATP was performed with Klenow DNA polymerase (exo-) (15 units, NEBiolabs). Home-made adapters containing a 4-bp unique tag used for multiplexing, were ligated with 2 µl T4 DNA ligase (NEBiolabs, 400,000 units/ml). DNA was size fractionated on 1% agarose gels and 500–750 bp DNA fragments were gel extracted with the Qiaquick gel extraction kit (Qiagen). DNA was PCR amplified for 12 cycles with Illumina primers PE1.0 and PE2.0 and Phusion DNA polymerase (1 unit, Thermo Scientific). Six PCR reactions were pooled for each library, and purified on a Qiagen purification column. Elution was performed in 30 µl and DNA was quantified on a spectrophotometer and on agarose gel.

Analysis of paired-end Illumina reads

Multiplexed libraries were loaded on a HiSeq2500 (Illumina), 110 bp paired-end reads for haploids and 260 bp paired-end reads for diploids were generated. Reads quality was evaluated by FastQC v0.10.1 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Reads were mapped along S288C chromosome reference sequence (*Saccharomyces Genome Database*, release R64-2-1, November 2014), using the paired-end mapping mode of BWA v0.7.4-r385 with default parameters [57]. The output SAM files were converted and sorted to BAM files using SAMtools v0.1.19-44428cd [59]. The command *IndelRealigner* from GATK v2.4-9 [60] was used to realign the reads. Duplicated reads were removed using the option “*MarkDuplicates*” implemented in Picard v1.94 (<http://picard.sourceforge.net/>). Reads uniquely mapped

to the reference sequence with a minimum mapping quality of 20 (Phred-scaled) were kept. Mpileup files were generated by SAMtools without BAQ adjustments. SNPs and INDELS were called by the options “*mpileup2snp*” and “*mpileup2indel*” of *Varscan2* v2.3.6 [61] with a minimum depth of 10 reads for haploids and 20 reads for diploids. Average read coverage was 255X for diploid cells ($\sigma = 187X$) and 190X for haploids ($\sigma = 43X$). Diploid strains are homozygous except for selection markers and some specific loci like *MAT* and *SUP4*. Therefore, *de novo* heterozygous mutations should represent 50% of reads, on the average. Taking that into account, lower and upper thresholds for variant allele frequency were respectively set between 30% and 70% in diploids. For haploids, the threshold for minimum variant allele frequency was set at 70%. Mutations less than 10 bp away from each other were discarded to avoid mapping problems due to paralogous genes or repeated sequences. To assess microsatellite mutations, we only retained reads uniquely anchored at least 20 bp on each side of the microsatellite [62]. All detected mutations were manually examined using the IGV software (version 2.3.77), and compared between all sequenced libraries for interpretation. All the scripts used in order to process data are available on github (<https://github.com/sdeclere/nuclease>). All Illumina sequences were uploaded in the European Nucleotide Archive (ENA), accession number PRJEB16068.

Statistical analyses

All analyses were performed using the R package (version 3.6.3) [63]. Survival rates after DSB induction were compared using the t-test (Fig 2). When comparing rates after DSB induction at *SUP4* in wild type vs mutant strains, the Bonferroni correction for multiple testing was applied. Deletions and rearrangement types between different strains were compared using the Chi2 test.

Supporting information

S1 Fig. Sequences at the left and right of junctions in rearranged haploid clones. Junctions were deduced from Illumina read mapping (when available) and confirmed by subsequent PCR and Sanger sequencing. Nucleotides in red are those used to anneal each DSB end, and are therefore present in only one copy in the genomic sequence. The extent of calculated deletions (Δ) is indicated in parentheses. Nucleotides in red in parentheses correspond to small deletions. Nucleotides in green correspond to insertions. The length of Ty insertions is indicated along with the LTR it comes from. Nucleotides in purple (*SpCas9* at the *ARG2* locus) correspond to the *I-SceI* site (see text). Nucleotides in light blue correspond to homeologies between the left and right junction sequences that were lost after deletion (the junction sequence shows the nucleotide in blue, not the one in red). Nucleotides in light blue in parentheses correspond to homeologies that were removed during the deletion. Note that extended homologies between LTRs does not always allow to determine the exact breakpoint with a high precision.

(PDF)

S2 Fig. Southern blot at the *YAK1* locus. The time course was run in non-induced (glucose) and induced (galactose) conditions, as previously. The uncut locus is clearly visible as a 2612 bp band, but no signal can be seen at the expected size for a DSB (2294 bp). Two fuzzy bands present in both conditions and corresponding to faint cross-hybridizations are indicated by asterisks. Note that even when the blot was overexposed no signal could be detected at the expected DSB size.

(PDF)

S3 Fig. Genome- wide mutation spectrum observed in haploid and diploid cells following Cas9 induction. **A:** Real-time PCR quantification of *CDC8* and *JEM1* amounts relative to an internal control on chromosome IV, in diploid cells in which Cas9 was induced. Half the amount of *CDC8* product was detected in each clone analyzed. This was significantly different from the amount of product amplified from the *JEM1* gene located on the other chromosome X arm. **B:** Illumina results for diploid and haploid cells. For each clone, the number of mutations detected is shown. Substit.: nucleotide substitution; Indel: insertion or deletion; Indel micro.: insertion or deletion of one repeat unit in a microsatellite. The asterisk corresponds to a 36 bp deletion in the *FLO11* minisatellite (36 bp repeat).
(PDF)

S4 Fig. Fate of a DSB made within a trinucleotide repeat tract using different endonucleases. The DSB induced by I-SceI (**A**), a TALEN (**B**) or SpCas9 (**C**). In each case, survivors were separated in three different categories: deletions (local or large) around the repeat tract, removing partially or totally the repeat (left), repeat contraction without other mutation (right) or all other kinds of rearrangements (middle). Note that in A, five triplets flank the repeat tract on each side, whereas in B and C the break is made at the end of a long (80 triplets) repeat tract, leaving 1–4 triplets downstream and the remaining upstream the DSB.
(PDF)

S5 Fig. Western blots. **A:** Western blots. Proteins were extracted in non-induced (Glu) and induced (Gal) conditions for wild type and each mutant strain. Glucose-6-phosphate deshydrogenase (G6PDH) was used as a loading control. **B:** Ratios of Cas9 over G6PDH signals, for each strain in each condition.
(PDF)

S1 Table. Off-target sites in the yeast genome, ranked by decreasing CFD score, using the CRISPOR tool.
(XLS)

S2 Table. List of strains used in the present study.
(DOCX)

S3 Table. List and sequence of primers used in this study.
(DOCX)

Acknowledgments

We thank Carine Giovannangeli for the generous gift of the Enhanced SpCas9 plasmid.

Author Contributions

Conceptualization: Guy-Franck Richard.

Formal analysis: Valentine Mosbach, David Viterbo, Lucie Poggi, Wilhelm Vaysse-Zinkhöfer.

Funding acquisition: Guy-Franck Richard.

Investigation: Valentine Mosbach, David Viterbo, Lucie Poggi, Wilhelm Vaysse-Zinkhöfer.

Methodology: Valentine Mosbach, Guy-Franck Richard.

Project administration: Guy-Franck Richard.

Software: Stéphane Descorps-Declère.

Supervision: Guy-Franck Richard.

Writing – original draft: Guy-Franck Richard.

Writing – review & editing: Valentine Mosbach, Guy-Franck Richard.

References

1. Richard G-F, Kerrest A, Dujon B. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiol Mol Biol Rev.* 2008; 72: 686–727. <https://doi.org/10.1128/MMBR.00011-08> PMID: 19052325
2. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature.* 2004; 431: 931–945. <https://doi.org/10.1038/nature03001> PMID: 15496913
3. Orr HT, Zoghbi HY. Trinucleotide repeat disorders. *Annu Rev Neurosci.* 2007; 30: 575–621. <https://doi.org/10.1146/annurev.neuro.29.051605.113042> PMID: 17417937
4. Gacy AM, Goellner G, Juranic N, Macura S, McMurray CT. Trinucleotide repeats that expand in human disease form hairpin structures in vitro. *Cell.* 1995; 81: 533–540. [https://doi.org/10.1016/0092-8674\(95\)90074-8](https://doi.org/10.1016/0092-8674(95)90074-8) PMID: 7758107
5. Liu G, Chen X, Bissler JJ, Sinden RR, Leffak M. Replication-dependent instability at (CTG)_x (CAG) repeat hairpins in human cells. *Nat Chem Biol.* 2010; 6: 652–9. <https://doi.org/10.1038/nchembio.416> PMID: 20676085
6. Axford MM, Wang YH, Nakamori M, Zannis-Hadjopoulos M, Thornton CA, Pearson CE. Detection of Slipped-DNAs at the Trinucleotide Repeats of the Myotonic Dystrophy Type I Disease Locus in Patient Tissues. *PLoS Genet.* 2013; 9: 1–13. <https://doi.org/10.1371/journal.pgen.1003866> PMID: 24367268
7. Anand RP, Shah KA, Niu H, Sung P, Mirkin SM, Freudenreich CH. Overcoming natural replication barriers: differential helicase requirements. *Nucleic Acids Res.* 2012; 40: 1091–105. <https://doi.org/10.1093/nar/gkr836> PMID: 21984413
8. Nguyen JHG, Viterbo D, Anand RP, Verra L, Sloan L, Richard G-F, et al. Differential requirement of Srs2 helicase and Rad51 displacement activities in replication of hairpin-forming CAG/CTG repeats. *Nucleic Acids Res.* 45: 4519–4531. <https://doi.org/10.1093/nar/gkx088> PMID: 28175398
9. Pelletier R, Krasilnikova MM, Samadashwily GM, Lahue R, Mirkin SM. Replication and expansion of trinucleotide repeats in yeast. *Mol Cell Biol.* 2003; 23: 1349–57. <https://doi.org/10.1128/mcb.23.4.1349-1357.2003> PMID: 12556494
10. Samadashwily G, Raca G, Mirkin SM. Trinucleotide repeats affect DNA replication in vivo. *Nat Genet.* 1997; 17: 298–304. <https://doi.org/10.1038/ng1197-298> PMID: 9354793
11. Viterbo D, Michoud G, Mosbach V, Dujon B, Richard G-F. Replication stalling and heteroduplex formation within CAG/CTG trinucleotide repeats by mismatch repair. *DNA Repair.* 2016; 42: 94–106. <https://doi.org/10.1016/j.dnarep.2016.03.002> PMID: 27045900
12. Sutherland GR, Baker E, Richards RI. Fragile sites still breaking. *Trends Genet.* 1998; 14: 501–506. [https://doi.org/10.1016/s0168-9525\(98\)01628-x](https://doi.org/10.1016/s0168-9525(98)01628-x) PMID: 9865156
13. Balakumaran BS, Freudenreich CH, Zakian VA. CGG/CCG repeats exhibit orientation-dependent instability and orientation-independent fragility in *Saccharomyces cerevisiae*. *Hum Mol Genet.* 2000; 9: 93–100. <https://doi.org/10.1093/hmg/9.1.93> PMID: 10587583
14. Freudenreich CH, Kantrow SM, Zakian VA. Expansion and length-dependent fragility of CTG repeats in yeast. *Science.* 1998; 279: 853–856. <https://doi.org/10.1126/science.279.5352.853> PMID: 9452383
15. Haber J E. *Genome stability.* Summer Scholl. New York: Garland Science; 2014.
16. Fairhead C, Dujon B. Consequences of unique double-stranded breaks in yeast chromosomes: death or homozygosis. *Mol Gen Genet.* 1993; 240: 170–180. <https://doi.org/10.1007/BF00277054> PMID: 8355651
17. Haber JE. In vivo biochemistry: physical monitoring of recombination induced by site-specific endonucleases. *BioEssays.* 1995; 17: 609–620. <https://doi.org/10.1002/bies.950170707> PMID: 7646483
18. Plessis A, Perrin A, Haber JE, Dujon B. Site-specific recombination determined by I-Sce I, a mitochondrial group I intron-encoded endonuclease expressed in the yeast nucleus. *Genetics.* 1992; 130: 451–460. PMID: 1551570
19. Nelms BE, Maser RS, MacKay JF, Lagally MG, Petrini JHJ. In Situ Visualization of DNA Double-Strand Break Repair in Human Fibroblasts. *Science.* 1998; 280: 590–592. <https://doi.org/10.1126/science.280.5363.590> PMID: 9554850

20. Mosbach V, Poggi L, Viterbo D, Charpentier M, Richard G-F. TALEN-induced double-strand break repair of CTG trinucleotide repeats. *Cell Rep.* 2018; 22: 2146–2159. <https://doi.org/10.1016/j.celrep.2018.01.083> PMID: 29466740
21. Richard G-F, Viterbo D, Khanna V, Mosbach V, Castelain L, Dujon B. Highly specific contractions of a single CAG/CTG trinucleotide repeat by TALEN in yeast. *PLoS ONE.* 2014; 9: e95611. <https://doi.org/10.1371/journal.pone.0095611> PMID: 24748175
22. DiCarlo JE, Norville JE, Mali P, Rios X, Aach J, Church GM. Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res.* 2013; 41: 4336–43. <https://doi.org/10.1093/nar/gkt135> PMID: 23460208
23. Chen H, Lisby M, Symington LS. RPA coordinates DNA end resection and prevents formation of DNA hairpins. *Mol Cell.* 2013; 50: 589–600. <https://doi.org/10.1016/j.molcel.2013.04.032> PMID: 23706822
24. Richard G-F, Goellner GM, McMurray CT, Haber JE. Recombination-induced CAG trinucleotide repeat expansions in yeast involve the MRE11/RAD50/XRS2 complex. *EMBO J.* 2000; 19: 2381–2390. <https://doi.org/10.1093/emboj/19.10.2381> PMID: 10811629
25. Kleinstiver BP, Pattanayak V, Prew MS, Tsai SQ, Nguyen NT, Zheng Z, et al. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature.* 2016; 529: 490–495. <https://doi.org/10.1038/nature16526> PMID: 26735016
26. Slaymaker IM, Gao L, Zetsche B, Scott DA, Yan WX, Zhang F. Rationally engineered Cas9 nucleases with improved specificity. *Science.* 2016; 351: 84–88. <https://doi.org/10.1126/science.aad5227> PMID: 26628643
27. Chen JS, Dagdas YS, Kleinstiver BP, Welch MM, Sousa AA, Harrington LB, et al. Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. *Nature.* 2017; 550: 407–410. <https://doi.org/10.1038/nature24268> PMID: 28931002
28. Lee JK, Jeong E, Lee J, Jung M, Shin E, Kim Y, et al. Directed evolution of CRISPR–Cas9 to increase its specificity. *Nat Commun.* 2018; 9: 3048. <https://doi.org/10.1038/s41467-018-05477-x> PMID: 30082838
29. Casini A, Olivieri M, Petris G, Montagna C, Reginato G, Maule G, et al. A highly specific SpCas9 variant is identified by *in vivo* screening in yeast. *Nat Biotechnol.* 2018; 36: 265–271. <https://doi.org/10.1038/nbt.4066> PMID: 29431739
30. Richard G-F, Dujon B, Haber JE. Double-strand break repair can lead to high frequencies of deletions within short CAG/CTG trinucleotide repeats. *Mol Gen Genet.* 1999; 261: 871–882. <https://doi.org/10.1007/s004380050031> PMID: 10394925
31. Richard G-F, Cyncynatus C, Dujon B. Contractions and expansions of CAG/CTG trinucleotide repeats occur during ectopic gene conversion in yeast, by a MUS81-independent mechanism. *J Mol Biol.* 2003; 326: 769–782. [https://doi.org/10.1016/s0022-2836\(02\)01405-5](https://doi.org/10.1016/s0022-2836(02)01405-5) PMID: 12581639
32. Field D, Wills C. Abundant microsatellite polymorphism in *Saccharomyces cerevisiae*, and the different distributions of microsatellites in eight prokaryotes and *S. cerevisiae*, result from strong mutation pressures and a variety of selective forces. *Proc Natl Acad Sci USA.* 1998; 95: 1647–1652. <https://doi.org/10.1073/pnas.95.4.1647> PMID: 9465070
33. Malpertuy A, Dujon B, Richard G-F. Analysis of microsatellites in 13 hemiascomycetous yeast species: mechanisms involved in genome dynamics. *J Mol Evol.* 2003; 56: 730–741. <https://doi.org/10.1007/s00239-002-2447-5> PMID: 12911036
34. Richard G-F, Dujon B. Distribution and variability of trinucleotide repeats in the genome of the yeast *Saccharomyces cerevisiae*. *Gene.* 1996; 174: 165–174. [https://doi.org/10.1016/0378-1119\(96\)00514-8](https://doi.org/10.1016/0378-1119(96)00514-8) PMID: 8863744
35. Richard G-F, Hennequin C, Thierry A, Dujon B. Trinucleotide repeats and other microsatellites in yeasts. *Res Microbiol.* 1999; 150: 589–602. [https://doi.org/10.1016/s0923-2508\(99\)00131-x](https://doi.org/10.1016/s0923-2508(99)00131-x) PMID: 10672999
36. Haeussler M, Schönig K, Eckert H, Eschstruth A, Mianné J, Renaud J-B, et al. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol.* 2016; 17. <https://doi.org/10.1186/s13059-016-1012-2> PMID: 27380939
37. Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR–Cas9. *Nat Biotechnol.* 2016; 34: 184–191. <https://doi.org/10.1038/nbt.3437> PMID: 26780180
38. Rothstein R, Helms C, Rosenberg N. Concerted deletions and inversions are caused by mitotic recombination between delta sequences in *Saccharomyces cerevisiae*. *Mol Cell Biol.* 1987; 7: 1198–1207. <https://doi.org/10.1128/mcb.7.3.1198> PMID: 3550432

39. Welcker AJ, de Montigny J, Potier S, Souciet JL. Involvement of very short DNA tandem repeats and the influence of the RAD52 gene on the occurrence of deletions in *Saccharomyces cerevisiae*. *Genetics*. 2000; 156: 549–557. PMID: [11014805](#)
40. Cinesi C, Aeschbach L, Yang B, Dion V. Contracting CAG/CTG repeats using the CRISPR-Cas9 nickase. *Nat Commun*. 2016; 7: 13272. <https://doi.org/10.1038/ncomms13272> PMID: [27827362](#)
41. Dabrowska M, Juzwa W, Krzyzosiak WJ, Olejniczak M. Precise Excision of the CAG Tract from the Huntingtin Gene by Cas9 Nickases. *Front Neurosci*. 2018; 12. <https://doi.org/10.3389/fnins.2018.00075> PMID: [29535594](#)
42. Lemos BR, Kaplan AC, Bae JE, Ferrazzoli AE, Kuo J, Anand RP, et al. CRISPR/Cas9 cleavages in budding yeast reveal templated insertions and strand-specific insertion/deletion profiles. *Proc Natl Acad Sci*. 2018; 115: E2040–E2047. <https://doi.org/10.1073/pnas.1716855115> PMID: [29440496](#)
43. Wilson TE, Grawunder U, Lieber MR. Yeast DNA ligase IV mediates non-homologous DNA end joining. *Nature*. 1997; 388: 495–498. <https://doi.org/10.1038/41365> PMID: [9242411](#)
44. Frank-Vaillant M, Marcand S. NHEJ regulation by mating type is exercised through a novel protein, Lif2p, essential to the Ligase IV pathway. *Genes&Development*. 2001; 15: 3005–3012.
45. Charpentier M, Khedher AHY, Menoret S, Brion A, Lamribet K, Dardillac E, et al. CtlP fusion to Cas9 enhances transgene integration by homology-dependent repair. *Nat Commun*. 2018; 9: 1133. <https://doi.org/10.1038/s41467-018-03475-7> PMID: [29556040](#)
46. van Overbeek M, Capurso D, Carter MM, Thompson MS, Frias E, Russ C, et al. DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks. *Mol Cell*. 2016; 63: 633–646. <https://doi.org/10.1016/j.molcel.2016.06.037> PMID: [27499295](#)
47. Haber JE. The many interfaces of Mre11. *Cell*. 1998; 95: 583–586. [https://doi.org/10.1016/s0092-8674\(00\)81626-8](https://doi.org/10.1016/s0092-8674(00)81626-8) PMID: [9845359](#)
48. Lengsfeld BM, Rattray AJ, Bhaskara V, Ghirlando R, Paull TT. Sae2 Is an Endonuclease that Processes Hairpin DNA Cooperatively with the Mre11/Rad50/Xrs2 Complex. *Mol Cell*. 2007; 28: 638–651. <https://doi.org/10.1016/j.molcel.2007.11.001> PMID: [18042458](#)
49. Mimitou EP, Symington LS. Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing. *Nature*. 2008; 455: 770–4. <https://doi.org/10.1038/nature07312> PMID: [18806779](#)
50. So CC, Martin A. DSB structure impacts DNA recombination leading to class switching and chromosomal translocations in human B cells. *PLOS Genet*. 2019; 15: e1008101. <https://doi.org/10.1371/journal.pgen.1008101> PMID: [30946744](#)
51. Storici F, Bebenek K, Kunkel TA, Gordenin DA, Resnick MA. RNA-templated DNA repair. *Nature*. 2007; 447: 338–341. <https://doi.org/10.1038/nature05720> PMID: [17429354](#)
52. Zhu Z, Chung WH, Shim EY, Lee SE, Ira G. Sgs1 helicase and two nucleases Dna2 and Exo1 resect DNA double-strand break ends. *Cell*. 2008; 134: 981–94. <https://doi.org/10.1016/j.cell.2008.08.037> PMID: [18805091](#)
53. Orr-Weaver TL, Szostak JW, Rothstein RJ. Yeast transformation: a model system for the study of recombination. *Proc Natl Acad Sci U S A*. 1981; 78: 6354–6358. <https://doi.org/10.1073/pnas.78.10.6354> PMID: [6273866](#)
54. Holmes AM, Haber JE. Double-strand break repair in yeast requires both leading and lagging strand DNA polymerases. *Cell*. 1999; 96: 415–424. [https://doi.org/10.1016/s0092-8674\(00\)80554-1](https://doi.org/10.1016/s0092-8674(00)80554-1) PMID: [10025407](#)
55. Muller H, Annaluru N, Schwerzmann JW, Richardson SM, Dymond JS, Cooper EM, et al. Assembling large DNA segments in yeast. *Methods Mol Biol Clifton NJ*. 2012; 852: 133–150. https://doi.org/10.1007/978-1-61779-564-0_11 PMID: [22328431](#)
56. Sikorski RS, Hieter P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics*. 1989; 122: 19–27. PMID: [2659436](#)
57. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma Oxf Engl*. 2009; 25: 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324> PMID: [19451168](#)
58. Viterbo D, Marchal A, Mosbach V, Poggi L, Vaysse-Zinkhöfer W, Richard G-F. A fast, sensitive and cost-effective method for nucleic acid detection using non-radioactive probes. *Biol Methods Protoc*. 2018; 3. <https://doi.org/10.1093/biomethods/bpy006> PMID: [32161800](#)
59. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25: 2078–9. <https://doi.org/10.1093/bioinformatics/btp352> PMID: [19505943](#)
60. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011; 43: 491–8. <https://doi.org/10.1038/ng.806> PMID: [21478889](#)

61. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* 2012; 22: 568–76. <https://doi.org/10.1101/gr.129684.111> PMID: 22300766
62. Fungtammasan A, Ananda G, Hile SE, Su MS-W, Sun C, Harris R, et al. Accurate typing of short tandem repeats from genome-wide sequencing data and its applications. *Genome Res.* 2015 [cited 13 Oct 2016]. <https://doi.org/10.1101/gr.185892.114> PMID: 25823460
63. Millot G. Comprendre et réaliser les tests statistiques à l'aide de R. 2nd ed. Brussels: de boeck; 2011.