R-loops and regulatory changes in chronologically ageing fission yeast cells drive non-random patterns of genome rearrangements

David A. Ellis[1,5], Félix Reyes-Martín[2], María Rodríguez-López[1], Cristina Cotobal[1], Xi-Ming Sun[3,4], Daniel C. Jeffares[1,6], Samuel Marguerat[3,4], Víctor A. Tallada[2], Jürg Bähler[1,*]

[1] Institute of Healthy Ageing and Department of Genetics, Evolution & Environment, University College London, London WC1E 6BT, U.K.

[2] Centro Andaluz de Biología del Desarrollo, Universidad Pablo de Olavide/Consejo Superior de Investigaciones Científicas, 41013 Seville, Spain.

[3] MRC London Institute of Medical Sciences (LMS), Du Cane Road, London W12 0NN, U.K.

[4] Institute of Clinical Sciences (ICS), Faculty of Medicine, Imperial College London, Du Cane Road, London W12 0NN, U.K.

[5] Current address: Ear Institute, University College London, London, WC1X 8EE, U.K.

[6] Current address: Department of Biology, University of York, York YO10 5DD, U.K.

* Corresponding author: Jürg Bähler, University College London, Gower Street, Darwin Building – GEE, London WC1E 6BT, United Kingdom

Email: j.bahler@ucl.ac.uk; phone +44(0)203-108-1602

Keywords: chronological lifespan, genomic instability, *Schizosaccharomyces pombe,* double-strand break repair, R-loop

Abstract

Aberrant repair of DNA double-strand breaks can recombine distant pairs of chromosomal breakpoints. Such chromosomal rearrangements are a hallmark of ageing and compromise the structure and function of genomes. Rearrangements are challenging to detect in non-dividing cell populations, because they reflect individually rare, heterogeneous events. The genomic distribution of *de novo* rearrangements in non-dividing cells, and their dynamics during ageing, remain therefore poorly characterized. Studies of genomic instability during ageing have focussed on mitochondrial DNA, small genetic variants, or proliferating cells. To gain a better understanding of genome rearrangements during cellular ageing, we focused on a single diagnostic measure – DNA breakpoint junctions – allowing us to interrogate the changing genomic landscape in non-dividing cells of fission yeast (*Schizosaccharomyces pombe*). Aberrant DNA junctions that accumulated with age were associated with microhomology sequences and R-loops. Global hotspots for age-associated breakpoint formation were evident near telomeric genes and linked to remote breakpoints on the same or different chromosomes, including the mitochondrial chromosome. An unexpected mechanism of genomic instability caused more local hotspots: age-associated reduction in an RNA-binding protein could trigger R-loop formation at target loci. This finding suggests that biological processes other than transcription or replication can drive genome rearrangements. Notably, we detected similar signatures of genome rearrangements that accumulated in old brain cells of humans. These findings provide insights into the unique patterns and potential mechanisms of genome rearrangements in non-dividing cells, which can be triggered by ageing-related changes in gene-regulatory proteins.

Introduction

Cellular processes like transcription and replication can trigger DNA lesions such as double-strand breaks (DSBs) (Gaillard and Aguilera 2016; Heyer 2015; Aguilera and García-Muse 2013; Skourti-Stathaki and Proudfoot 2014). A sensitive sequencing approach has revealed DSBs at hotspots in mouse brain cells, linked to transcribed genes with neuronal functions (Wei et al. 2016), suggesting that the physiological context can affect the landscape of DSBs. DSBs are normally repaired by homologous recombination or by non-homologous end-joining – two pathways which protect chromosomes from aberrant structural variations (Soutoglou et al. 2007; Difilippantonio et al. 2000; Haber 2016). Under certain physiological conditions, e.g. when the regular DNA-repair pathways are compromised, alternate DNA end-joining processes take over, often involving short homologous sequences (microhomologies) that are typically unmasked through DNA-end resection from the DSBs (Ottaviani et al. 2014; Decottignies 2007, 2013). Microhomology-mediated end-joining events can link chromosomal breakpoints that are normally far apart or even on different chromosomes (Bétermier et al. 2014; Villarreal et al. 2012). Such events lead to genome rearrangements such as inversions, duplications, translocations or deletions, which can considerably affect the function of genomes. The patterns of genome rearrangements are shaped by the particular mechanisms of their formation and by the fitness effects they exert on the cell.

Ageing has been associated with both an increase in DSBs (Lombard et al. 2005; White and Vijg 2016) and a decline in the efficiency and accuracy of DNA repair (Gorbunova and Seluanov 2016; White and Vijg 2016). Accordingly, increased genomic instability and chromosomal rearrangements are well-known hallmarks of ageing (Szilard 1959; Curtis 1963; Strehler 1986; Hasty et al. 2003; López-Otín et al. 2013; Vijg and Suh 2013). Impaired non-homologous end-joining leads to accelerated ageing in human patients and mouse models, and microhomology-mediated end-joining increases with age (Vaidya et al. 2014). Genome re-sequencing studies during ageing have been limited to mitochondrial DNA (Williams et al. 2013; Kennedy et al. 2013), small genetic variants (Gangloff et al. 2017;

Makarenko et al. 2020) and duplications (Maestroni et al. 2017), or proliferating cells (Hu et al. 2014). No systematic approaches have been applied to identify heterogeneous, rare chromosomal rearrangements in non-dividing, somatic cells (Quispe-Tintaya et al. 2016; White and Vijg 2016; Laurie et al. 2012; Jacobs et al. 2012).

Processes affecting ageing are remarkably conserved from yeast to human, including both genetic and environmental factors (López-Otín et al. 2013; Kaeberlein 2010). The fission yeast, *Schizosaccharomyces pombe*, is a potent model for cellular ageing; we and others have explored effects of nutrient limitation, signalling pathways and genetic variations on chronological lifespan in *S. pombe* (Ellis et al. 2018; Roux et al. 2009; Rallis et al. 2013, 2014). Chronological lifespan is defined as the time a cell survives in a quiescent, non-dividing state, which models post-mitotic ageing of somatic metazoan cells (López-Otín et al. 2013; Kaeberlein 2010). Quiescent *S. pombe* cells feature distinct DNA-damage responses (Mochida and Yanagida 2006; Ferreira and Cooper 2004; Maestroni et al. 2017) and distinct mutational forces that can promote genetic diversity (Makarenko et al. 2020; Gangloff et al. 2017). Here we interrogate aberrant genomic DNA-junction sequences in non-dividing *S. pombe* cells, revealing distinct signatures of ageing-associated chromosomal rearrangements and suggesting their mechanistic underpinning. Similar patterns of rearrangements are also evident in ageing human brain cells.

Results and Discussion

Microhomology-associated genome rearrangements specifically increase in ageing yeast cells. Previously, we sequenced eight chronologically ageing pools of *S. pombe* cells, generated from advanced intercross lines of strains Y0036 (Clément-Ziza et al. 2014) and DY8531 (Hu et al. 2015), and analysed changes in the standing genetic variation as a function of age to identify longevity-associated quantitative trait loci (Ellis et al. 2018) (Materials and Methods). Here we report the striking new structural variation that arose in these cellular pools during ageing. Structural variant calling software usually requires

support from multiple sequence reads (Ye et al. 2009; Wang et al. 2011; Emde et al. 2012; Sindi et al. 2012; Rausch et al. 2012; Jiang et al. 2012; Trappe et al. 2014; Layer et al. 2014; Kimura and Koike 2015; Jeffares et al. 2017). Whilst this is useful for reducing false positives, these algorithms will only identify variations present in multiple cells in a population. To identify the rare, heterogeneous variations arising spontaneously in different non-dividing cells, we stringently filtered split reads that joined sequences from two distant genomic sites (Fig. S1) (Ellis et al. 2018). Such split reads represent potential breakpoint junctions of genome rearrangements, which lead to new sequence combinations (Fig. 1A). We identified 776,174 such junctions from 225,554,047 total reads across eight replicate pools sampled at six time points.

Several lines of evidence indicate that these breakpoint junctions are not artefacts of sequence library preparation but represent *in vivo* genome rearrangements. First, fewer junctions were present within coding regions than would be expected by chance (Fig. 1B). This bias may reflect selection against intra-genic rearrangements that disrupt gene function. Second, modelling showed that the free DNA ends available for junction formation were not proportionally represented in the observed juxtapositions, i.e. sequences represented by higher read depth were not more likely to feature in breakpoint junctions (Fig. S2). Third, a larger age-associated increase was evident in intra-chromosomal junctions than in inter-chromosomal junctions, and among intra-chromosomal events, junctions joining neighbouring regions were preferred over those joining more distal regions (Fig. S3; Note S1). We would not expect such bias from a bioinformatics artefact. Fourth, age-associated junctions were characterised by separate repair signatures at breakpoints compared to signatures suggestive of false positives, of which there were far fewer (see below and Note S1). A drawback of this approach is that the juxtaposed regions forming a junction are analysed without supporting information from other reads or breakpoints, and the exact nature and extent of the structural variations thus remains unknown.

During chronological ageing, breakpoint junctions strongly increased relative to the total number of mapped reads in each sample, particularly from Day 2 onwards (Fig. 1C).

This increase was most pronounced for junctions involving nuclear DNA only, but was also evident within mitochondrial DNA and between nuclear and mitochondrial DNA (Fig. 1C). Work from others supports the notion that DSBs in nuclear DNA can be repaired with mitochondrial DNA (Hu et al. 2014; Ricchetti et al. 1999; Decottignies 2005). The breakpoint junctions featured different types of sequence rearrangements: single-base insertions not present at either joined region, blunt junctions directly linking two regions, or microhomologies of up to 20 bases shared between both joined regions (Fig. 1D). The blunt junctions and the junctions with single-base insertions or single-base microhomologies did not increase with age. These junctions might have been formed by a distinct mechanism before ageing and/or they could reflect artefacts (Note S1). In stark contrast, junctions featuring 2-20 bases of microhomology did markedly accumulate with age (Fig. 1D). This signature indicates that these ageing-associated rearrangements occur by microhomology-mediated end-joining. The observed size distribution, with a peak at 5-6 bases, might reflect a trade-off between the length of microhomology available near DSBs and the benefit of longer homology for end-joining repair. Interestingly, rearrangements with similar patterns of microhomology seem also to be enriched in cancer cells (Wala et al. 2017). We conclude that genome-wide rearrangements, represented by breakpoint junctions featuring microhomologies, accumulate as a function of the age of non-dividing cells.

Using motif discovery, we found seven long sequence motifs enriched at microhomology-mediated junctions (Fig. S4A). Motifs of known transcription factors from fission and budding yeast, which are shorter than the longer sequence motifs discovered, showed significant homology within these longer motifs (Fig. S4B). These transcription factors are involved in nutrient starvation and other stress responses or in cell-cycle control. This result suggests that specific transcription factors are associated with regions near chromosomal junctions, raising the possibility that they are involved in triggering DSBs and microhomology-mediated rearrangements.

Similar patterns of genome rearrangements accumulate in old human brain cells. Non-dividing yeast cells are a model for the post-mitotic ageing of somatic human cells such as the long-lived cells of the brain (Magrassi et al. 2013; Kaeberlein 2010). To check whether similar ageing-associated rearrangements also occur in human cells, and to validate our method in an independent system, we applied our junction calling pipeline to published sequencing data of young and old human brain tissue (Williams et al. 2013). We found a subtle increase in junctions in older brain cells (Fig. 2A), but the coverage and sample number in this data set were low (Materials and Methods). As in fission yeast (Fig. 1D), these junctions were associated with microhomology (Fig. 2A). The microhomology-associated junctions in human brain cells showed a bimodal distribution: a large population featuring similar microhomology lengths to fission yeast (peaking around 4-6 bases), and a less abundant population featuring longer microhomology (median ~16 bases). Interestingly, in cancer genome sequences a transition in the probability of junction formation occurs at around 11 bases of microhomology (Wala et al. 2017); the authors suggest that this transition reflects a shift in repair mechanisms from microhomology-mediated end-joining to single-strand annealing. Notably, simulated data showed that there were fewer junctions in coding regions of human brain cells than would be expected by chance (Fig. 2B). As for fission yeast (Fig. 1B), this finding likely reflects selection against rearrangements that interfere with gene function, either through cell death or active culling of unfit cells (Merino et al. 2015). These results raise the possibility that similar patterns of ageing-related DNA rearrangements occur in both yeast and in human brain cells.

Local and global hotspots for genome rearrangements in ageing yeast cells. Junction formation over time represents a complex, multi-dimensional process: each junction is comprised of two juxtaposed sequences from any two genomic regions; independent junctions can recurrently form between the same two regions (Fig. 3A, green), or between one ‘hotspot’ region and different other regions (Fig. 3A, blue & red); and junction-formation can be either age-dependent (Fig. 3A, red) or not (Fig. 3A, blue). To visualize the yeast

junctions and identify age-associated patterns in any rearrangements, we determined ratios of the number of junctions in the oldest cells to the corresponding number in the youngest cells. Ageing-associated junctions formed preferentially at two distinct types of hotspot: 1) those enriched for local, intra-chromosomal junctions (mostly within 20kb), and 2) those enriched for more global, inter-chromosomal junctions (Fig. 3B). The local hotspots were more abundant but less pronounced than the global hotspots (Fig. 3B). These local hotspots likely reflect spatial constraints on junction formation, with neighbouring DNA being a more likely repair substrate than distal DNA (Fig. S3). At global hotspots, on the other hand, junctions between nearby regions of DNA were under-represented, contrary to expectation and the situation at local hotspots (Fig. S3). These findings suggest that distinct mechanisms operate at global and local hotspots.

We identified three strong global hotspots featuring numerous connections with other, typically remote, sequences throughout the genome. These global hotspots were located near the right end of Chromosome II and near both ends of Chromosome III, the latter being the sites of ribosomal DNA (rDNA) (Fig. 3B). Given that these global hotspots occur in repetitive regions near chromosome ends (Toda et al. 1984; Wood et al. 2002; Mandell et al. 2004), they might simply reflect the large number of repeated sequences. However, if the copy number of these repeated sequences remained constant during ageing, the ratio of junctions between young and old cells should still reflect ageing-associated changes (Fig. S5A). We therefore checked for changes in the ratio of repeat copy numbers between young and old cells at global hotspots. This analysis showed that repeat sequences at hotspots did not increase, but actually decreased with age (Fig. S5B). Thus, if anything, we under-estimated the prevalence of the global hotspots as sites for age-associated junction formation. Work in other systems has shown that copy numbers of rDNA repeats decrease with age, and instability in this region is linked with ageing and longevity (Spivey et al. 2017; Kobayashi 2014; Lu et al. 2018). The decrease of rDNA copy numbers with age could be linked to junction formation in these regions: work in primate kidney cells has demonstrated differential repair of cellular DNA sequences (Zolan et al. 1982), reminiscent of the

heterogeneity we observe for age-associated rearrangements. We conclude that ageing-associated rearrangements occur preferentially at either local or global hotspots.

Possible causes for global hotspots downstream of *thl2*. To better understand the global hotspots, we analysed the positions of junctions relative to genome annotations. The first hotspot was downstream of a *tlh2* gene (Fig. 4A; left), encoding a RecQ family DNA helicase. Copies of *tlh2* reside on all four sub-telomeres of Chromosomes I and II (Mandell et al. 2004). The *S. pombe* reference genome assembly only includes one *tlh2* gene copy, and it seems likely that the other three copies also feature hotspots. Initially discovered in *E. coli* (Nakayama et al. 1984), RecQ helicases are highly conserved. Notably, mutations in two human RecQ helicases, BLM and WRN, lead to premature ageing through Bloom or Werner syndromes (Bohr 2008) and another, RECQL5, alleviates transcriptional stress (Saponaro et al. 2014). The other two hotspots were near the 5.8S, 18S and 28S ribosomal RNA genes (Fig. 4A; middle and right). At all hotspots, junctions were enriched at the 3'-ends of these genes (Fig. 4A). Ribosomal RNA genes are highly transcribed and common sites of transcription and replication stress (Gaillard and Aguilera 2016; Aguilera and García-Muse 2013). Moreover, rDNA repeats can become unstable during ageing and cause cell death (Spivey et al. 2017; Lu et al. 2018; Kobayashi 2014).

A re-analysis of RNA-seq data from non-dividing cells (Atkinson et al. 2018) revealed a subtle increase in *tlh2* expression during chronological ageing (Fig. S6A). To test whether increased transcription at *tlh2* leads to increased junction formation at this hotspot, we generated a strain that overexpresses *tlh2.* We then sequenced the genome of this strain, along with a wild-type control, in early stationary phase when *tlh2* expression is normally low. The proportion of junctions downstream of *tlh2* was higher in the *tlh2* overexpression strain compared to wild-type (Fig. S6B). Thus, overexpression of *tlh2* is sufficient to trigger increased breakpoint junctions, possibly reflecting rearrangements owing to transcriptional stress. Moreover, the *tlh2* overexpression strain was substantially shorter-lived than wild-type cells (Fig. S6C). These results suggest that increased transcription of *tlh2* and/or

increased levels of the Tlh2 protein cause rearrangements that affect cell survival and longevity.

Given the subtle and variable increase in *tlh2* expression based on bulk RNA-seq data (Fig. S6A), we wondered whether *tlh2* might be transcribed heterogeneously during ageing (variegation), with rearrangements being limited to a few cells featuring high *tlh2* expression. Single-molecule RNA Fluorescence *In Situ* Hybridization (smFISH) experiments showed heterogeneous expression in cells overexpressing *tlh2*, while no clear *tlh2* expression was evident in wild-type cells during ageing (Fig. 4B). Inspection of the mapped RNA-seq reads from cell populations (Atkinson et al. 2018), however, revealed a pronounced RNA peak in the rearranged region downstream of *tlh2,* and this peak greatly increased in older cells (Fig. 4C). RNA annealed to DNA, known as R-loops, are common sites of genome instability: they can trigger DSBs through collisions with the transcription or replication machineries or through active processing by nucleotide excision-repair nucleases (Santos-Pereira and Aguilera 2015; Skourti-Stathaki and Proudfoot 2014; García-Rubio et al. 2018a); they can also directly interfere with DNA repair (Amon and Koshland 2016). Recent work in budding yeast has shown that sirtuins inhibit R-loop formation and associated genome instability, and the absence of sirtuins leads to a genome-wide increase in nascent mRNAs (Feldman and Peterson 2019). Along with other chromatin remodelling factors, the *S. pombe* sirtuin Sir2 modulates *tlh2* expression by modifying chromatin at the telomeres (Barrales et al. 2016). Accordingly, the smFISH experiment showed that *tlh2* was de-repressed in *sir2* deletion cells (Fig. 4B). To test whether the *tlh2* locus is prone to R-loop accumulation, we applied a DNA-RNA immunoprecipitation followed by sequencing (DRIP-seq) experiment for the profiling of R-loops (Ginno et al. 2012). Indeed, R-loops were strongly enriched downstream of *tlh2* (Fig. 4C). Together these results suggest that R-loops may trigger the rearrangements at these global hotspots whose transcriptional activity depends on chromatin-remodelling factors such as Sir2.

Ageing-related changes in RNA-binding proteins and R-loops may promote genome rearrangements. Given the bias towards 3'-ends of genes at global hotspots, we analysed the position of all breakpoint junctions relative to coding regions (Materials and Methods). In agreement with Fig. 1B, junctions were under-enriched in coding regions (Fig. 5A). Moreover, significantly more junctions occurred in 3'-untranslated regions (UTRs) of genes than in 5'-UTRs, similar to the situation at the global hotspots. Our pipeline showed no bias towards AT- or GC-rich regions (Materials and Methods). This finding suggests that the ends of genes are particularly prone to rearrangements. Next, we identified genes whose 3'-UTRs contained more junctions than would be expected by chance, given their length (Fig. 5B). Functional enrichment analysis using AnGeLi (Bitton et al. 2015) showed that, of the 148 enriched genes, 17 produced transcripts that are targets of Scw1, representing a significant enrichment (FDR-corrected $p < 0.0001$). Scw1 is an RNA-binding protein (RBP) (Hasan et al. 2014; Jin and McCollum 2003; Karagiannis et al. 2002) and an orthologue of the human polypyrimidine-tract-binding proteins RBPMS and RBPMS2. RBPMS was initially identified in an effort to find the Werner Syndrome gene (Shimamoto et al. 1996), suggesting intriguing parallels with the RecQ helicase Tlh2 at the hotspot discussed above.

Scw1 negatively regulates its target RNAs, possibly by binding to a motif in their 3'-UTRs (Hasan et al. 2014). To examine how an RBP might affect DNA rearrangements, we looked for any ageing-dependent changes in Scw1 substrate binding (Materials and Methods). These analyses suggested that Scw1 loses its affinity for some RNA targets with age (Fig. S7A), but it does not switch binding substrate from RNA to DNA (Fig. S7B). In fact, the protein levels of Scw1 markedly decreased in ageing cells (Fig. 5C). The budding yeast orthologue of Scw1, Whi3, aggregates during replicative ageing (Schlissel et al. 2017), so it is possible that the decrease in Scw1 reflects protein aggregation. In any case, this result suggests that Scw1, rather than promoting rearrangements at its targets, is preventing them in young cells.

How might an age-associated functional loss of Scw1 trigger rearrangements? We hypothesized that R-loops might again be involved. The presence of RBPs on nascent

transcripts can inhibit the formation of R-loops by preventing RNA annealing to the single-stranded DNA template (Huertas and Aguilera 2003; Aguilera 2002; Li and Manley 2005; Wahba et al. 2011). To test whether loss of Scw1 promotes R-loop formation, we quantified R-loops in *scw1* deletion (*scw1*Δ) and wild-type cells using R-loop immunostaining and fluorescence quantification. We validated the technique using RNase H mutants that are known to show increased R-loop formation as a positive control (Zhao et al. 2018) and RNase H treatment as a negative control (Fig. S8). For a physiologically relevant comparison, we used proliferating and early stationary phase cells, because older wild-type cells naturally lack Scw1 (Fig. 5C). Indeed, the *scw1*Δ cells contained more R-loops than wild-type cells (Fig. 5D). We applied DRIP-seq to directly test the effect of Scw1 on R-loop formation at the 227 published target genes of Scw1 (Hasan et al. 2014). This experiment showed that such R-loops slightly, but significantly increased in the absence of Scw1 (Fig. 5E). An increase remained even when the 17 Scw1 targets we identified (Fig. 5B) were not included (one sample t-test: T=-2.46, p=0.014). This finding suggests that the R-loops increased across all or most Scw1 targets. We conclude that the absence of Scw1 is sufficient for increased R-loop formation at its target genes. We propose that diminished Scw1 function in old cells leads to exposed 3'-ends of nascent target mRNAs that can re-anneal with complementary sequences in their template DNA (Fig. 5F). The resulting R-loops may then trigger genome instability and associated rearrangements downstream of the coding regions of Scw1 target genes. Thus, age-associated changes in the expression of an RBP with a role in post-transcriptional gene regulation can lead to non-random junction distribution via increased genome instability at its targets.

Conclusions. Our sensitive analyses of deeply sequenced yeast genomes uncover the rare and diverse events of chromosomal rearrangements that specifically accumulate during ageing of non-dividing cells. Our results provide evidence for widespread genome rearrangements accumulating during ageing of non-dividing fission yeast, associated with microhomology sequences near the breakpoints. Similar rearrangements are also evident in

older human brain cells. Junctions indicative of rearrangements show a non-random distribution, both in terms of the regions that are joined (local DNA is often preferred to distal DNA) and the biological importance of the region (fewer junctions occur in coding regions). These rearrangements occur in global or local hotspots and may be triggered by non-random patterns of DNA-RNA interactions. We find that the age-associated decline in the RNA regulatory protein Scw1 and associated R-loop formation can trigger breakpoint formation at target genes. Other RBPs that are modulated during ageing might act similarly to Scw1 in triggering genome rearrangements at other hotspots. These findings highlight that physiological changes can fuel cell- and condition-specific genome rearrangements with a non-random distribution.

Materials and Methods

Strains used in this study. The ageing pools of yeast used for the main experiment are described elsewhere (Ellis et al. 2018). Briefly, an inter-crossed population fission yeast derived from the parental strains Y0036 and DY8531 (Hu et al. 2015; Brown et al. 2011; Jeffares et al. 2015) was inoculated into eight separate 2L flasks of liquid yeast extract supplemented (YES) medium, and grown until the optical density reached a plateau. Samples were then snap frozen in liquid nitrogen every 24hrs for the next five days. Scw1Δ and Scw1-TAP strains have been published elsewhere (Hasan et al. 2014). Samples for RIP-chip and ChIP-seq of Scw1-TAP were taken at 0, 2 and 4 days after cells reached a plateau in optical density. The Tlh2OE strain was generated for this study using a PCR-based approach (Bähler et al. 1998) to introduce an *nmt1* promoter (Forsburg 1993) in front of *tlh2*. Experiments using this strain, and experiments for chromosomal spreads, were carried out in Edinburgh minimal medium (EMM). For any lifespan experiments in EMM, cultures were grown until the optical density of cells reached a plateau, when cells were spun down and re-suspended in EMM without glucose. All transgenic strains were confirmed by PCR. The Sir2 deletion strain was obtained from the Bioneer fission yeast deletion library (Kim et al. 2010).

DNA extraction, library preparation and sequencing. Genomic DNA was extracted using a standard phenol chloroform procedure (Murray et al. 2016). After precipitation, DNA pellets were re-suspended in TE buffer and treated with RNase A (Qiagen), before being mechanically sheared to ~200bp (Covaris AFA). Sheared DNA was passed through PCR purification columns (QIAquick, Qiagen), and the fragment size distribution was checked using a 2100 Bioanalyzer (Agilent). Libraries were prepared using NEBNext Ultra kits (NEB) according to the manufacturer's standard protocol – this procedure included dA-tailing (Note S1). After individual quantification (Qubit) and quality control (2100 Bioanalyzer), all forty-eight libraries were pooled. The pool was then sequenced using 126nt paired-end reads on the Illumina HiSeq 2500 (SickKids, Canada).

Read alignment and junction filtration. After performing initial checks in FastQC (Andrews 2010), reads were aligned to the *S. pombe* reference genome (accessed May 2015) (Wood et al. 2002) using default parameters in BWA-MEM (Li and Durbin 2009; Li 2013) (v0.7.12). Bam files were sorted, and PCR duplicates removed using Samtools (Li et al. 2009) (v0.1.19). Split reads were then obtained using a simple shell script integrating Samtools (Li et al. 2009) (v1.2). Note that, when assessing mitochondria-mitochondria split reads, the circularity of the mtDNA needed to be accounted for by ignoring those that aligned to both the start and end of the reference contig. Split reads were then filtered in a custom python script. To pass filtration, both alignments of each split read had to adhere to the following criteria: a minimum length of 40bp, a mapping quality score of 60 (the maximum score given by BWA-MEM), no clipped alignments at both ends (Suzuki et al. 2011). Information on the alignments of each split read was obtained from the CIGAR strings of the initial soft-clipped read, as any information contained in subsequent hard-clipped reads is redundant. The 100bp sequencing reads from human data (Williams et al. 2013) were re-mapped to the hg38n assembly of the human genome (accessed May 2016) using the same pipeline. The original purpose of this data set was to compare somatic mutations

in young and old mitochondrial DNA, and the sequenced libraries were mtDNA-enriched (Williams et al. 2013). However, there was still a considerable amount of coverage at the nuclear chromosomes, although less so than in the fission yeast data set.

To assess the pipeline, we generated one hundred random 100bp sequences of the *S. pombe* reference genome using BioPython (Cock et al. 2009) and output them as separate vcf files (the file format required by Mason (Holtgrewe 2010) – see below). Each vcf file also contained a random location at which the fragment should be inserted. These fragments were then inserted into the reference genome to produce separate fasta files (using GATK (McKenna et al. 2010) and Mason (Holtgrewe 2010)). Each fasta file was used to generate 1x of simulated reads. Once all reads had been simulated, they were mixed with reads simulated using the standard reference genome (with no insertions) to give a proportion of split reads similar to that obtained in the real sequence files. The read alignment and junction filtration pipeline described above was then applied, and the number of simulated junctions that were recovered was counted. Although sensitivity was low (14/100 simulated rearrangements were recovered), there were no false positives, showing that this filtration is conservative but robust. Using a two-sample Wilcox test, a comparison between the GC-content of all simulated junctions (N=100) and recovered junctions (N=18) showed that there was no significant bias in our pipeline toward calling junctions in AT- or GC-rich regions (W=992, p=0.5). All sequence data are available in the European Nucleotide Archive under the study accession PRJEB30570.

Modelling the expected rearrangement distribution. To calculate the expected number of DNA fusions between each chromosome in a random admixture of DNA ends, the number of reads mapping to each chromosome (all samples and timepoints considered simultaneously) was obtained using Samtools (Li et al. 2009) (v1.2) and converted to a proportion of the total number of reads. In Perl, these proportions were used to calculate the number of junctions that would be expected for each combination of chromosomes, were the DNA ends to join randomly. One thousand simulations were performed.

Measurement of microhomology at junctions. For each time point, bam files from all repeats were merged using Samtools (v1.2) (Li et al. 2009). Split reads were then obtained and filtered as above. Using python, junctions were first categorised as follows: those whose alignments share no sequence homology and have no non-homologous sequence between them; those whose alignments share no sequence homology and contain a non-homologous insertion between them; those whose alignments share an overlapping region of sequence homology (note that this homology is not necessarily perfect and may contain InDels). After categorising each junction, the length of any non-homologous or homologous sequence was recorded.

Motif Enrichment. For each junction classed as microhomology-mediated, we used BioPython (Cock et al. 2009) to collect the 100bp surrounding sequence. These 100bp sequences were compiled into a multi-sequence fasta file and submitted to MEME (Bailey et al. 2009) for motif discovery. Only motifs with e-values lower than 0.05 were considered. Using TomTom (Bailey et al. 2009), these sequences were compared to databases of known motifs for DNA-binding protein in fission and budding yeast. The most significant hits are shown in Fig. S8B.

Identification of rearrangement hotspots. Junction location files obtained after split read filtration were merged to create one combined file for each time point. A custom python script was used to divide each chromosome into 20kb windows and categorise each junction into a window based on its two alignments. The number of junctions at all windows was then output as separate chromosome-chromosome matrices for each time point. To see how much of an increase there was at each window, day five matrices were divided by their corresponding matrix at day zero (the end and start of the experiment, respectively). To quantify global hotspots, the average ratio across all windows at each 20kb bin of each chromosome was calculated.

Coverage analysis. For the analysis in Fig. S4, the coverage at every position in the genome was obtained for each merged bam file using Bedtools (Quinlan et al. 2010) (v2.22.1). This per-base coverage was then used to obtain the median coverage at each 20kb bin. To get a score for how much each bin had changed in read depth, the median at day 5 for each bin was divided by the median at day zero.

Junction intersection analyses. To see if junctions were enriched at coding regions, bed files for the locations of each feature were obtained. The coordinates of PomBase-annotated (Wood et al. 2012; McDowall et al. 2015) coding regions were obtained from an *S. pombe* gff3 file (v31); the coordinates of Havana-annotated coding regions were obtained from a human gtf file (v85). For each set of real junctions analysed, an equally sized set of randomly located junctions was simulated using a custom python script. For example, in our yeast analyses there were eight repeats, which meant eight separate simulated sets for any comparison. Only nuclear junctions were used and simulated for all analyses. The intersection of each junction set with a given feature set was then made using Bedtools (Quinlan et al. 2010) (v2.22.1). The proportion of intersecting junctions in each real repeat was then compared to the proportion of intersecting junctions in each simulated repeat.

To analyse the distribution of junctions at the start and end of genes, a custom python script was used to collect CDS start and end positions based on their strand and coordinates from a bed file (see above). For each repeat, junction alignments were then collected at the 500bp surrounding these positions, and their relative positions inside or outside the gene end were recorded.

Gene enrichment. To get the number of junctions in each 3' UTR, 3' UTR coordinates were obtained from an *S. pombe* gff3 file (see above) and converted into bed format. The number of junctions at each 3' UTR were then counted and plotted against the length of that UTR. Any gene with more junctions than twice the standard deviation of all UTRs were considered

recurrently rearranged. Gene enrichment analysis was performed in AnGeLi (Bitton et al. 2015).

Chromosomal spreads and R-loop immunostaining. R-loop immunostaining and fluorescence quantification was performed as described (Hoyos-Manchado et al. 2017), with the exception of cell lysis, where we mechanically broke cells in liquid nitrogen instead of using enzymatic lysis. For negative control samples, we added RNase H (Roche 10786357001) at 3u/100µl and incubated for two hours at 37°C. Nuclei were stained with DAPI at 3µg/mL in 50% glycerol. Images were obtained using a spinning disk confocal microscope (Yokogawa CSU-X1 head mounted on Olympus body), CoolSnap HQ2 camera (Roper Scientific), 100X Plan Apochromat and a 1.4 NA objective (Olympus). Images presented correspond to maximal projections of 10 slides' stacks using Image J open software (Schneider et al. 2012).

RIP-chip experiments. Two experimental repeats were performed for three timepoints: Days 0, 2 and 4 after cells reached stationary phase. RIP-chip of Scw1-TAP was performed as described (Hasan et al. 2014), except for the following modifications: immunoprecipitation was carried out using monoclonal antibodies against protein A (Sigma); the lysis buffer contained 10mg/ml heparin (sigma H7405), 1 mM PMSF and 1:100 protease inhibitor cocktail (Sigma P8340); and magnetic beads containing the immunoprecipitate were resuspended in 50 µl of wash buffer containing 1 mM DTT, 1 unit/ml of SuperaseIN (Ambion 2696) and 30 units/ml of AcTev protease (Life Technologies 12575015). The solution with the beads was incubated for 1.5 h at 18°C and the supernatant recovered, and RNA extracted using PureLink RNA micro columns (Life Technologies), according to the manufacturer's instructions. The RNA was eluted from the column in 14 µl and used for labelling without amplification. For microarrays, fluorescently labelled cDNA was prepared from total RNA and immunoprecipitated RNA from the RIP-chip using the SuperScript Plus Direct cDNA Labelling System (Life Technologies) as described by the manufacturer, except

for the following modifications: 10 µg of total RNA was labelled in a reaction volume of 15 µl. We then used 0.5 µl of 10× nucleotide mix with labelled nucleotide (1/3 of the recommended amount), and added 1 µl of a home-made dNTP mix (0.5 mM dATP, 0.5 mM dCTP, 0.5 mM dGTP, 0.3 mM dTTP) to the reaction. All other components were used at the recommended concentrations. Note that these changes are essential to prevent dye-specific biases. Labelled cDNAs were hybridised to oligonucleotide microarrays manufactured by Agilent as described (Rallis et al. 2013). Microarrays were scanned with a GenePix 4000A microarray scanner and analysed with GenePix Pro 5.0 (Molecular Devices). A dye swap was performed for all repeats. Any gene with data missing at any probe or in either repeat were not included in the analysis. Data was median-centred for analysis and any genes whose expression was highly un-correlated across repeats or did not appear at all timepoints were not considered. The RIP-chips data have been submitted to ArrayExpress under accession number E-MTAB-7618.

ChIP-seq experiments. ChIP-seq assays were performed as described (Cotobal et al. 2015). For experiments using Scw1-TAP, ChIP-seq samples were collected at the same timepoints as for the RIP-chip experiments (see above). ChIP-seq libraries were sequenced using an Illumina MiSeq instrument. SAMtools (Li and Durbin 2009) and BED tools (Quinlan et al. 2010) were used for sequence manipulation. Peak calling was performed using GEM (Guo et al. 2012). The ChIP-seq data are available in the European Nucleotide Archive under the study accession PRJEB30570.

Single-molecule RNA Fluorescence *In Situ* Hybridization. We performed smFISH, imaging and quantification on formaldehyde-fixed cells as described (Sun et al. 2020). Stellaris probes were designed and synthetized by Biosearch Technologies (Petaluma, CA). The *rpb1* probes were labelled with Quasar 670 and the *tlh2* probes with CAL Fluor Red 610. Probe sequences are provided in Table S1. Cells were mounted in ProLong Gold antifade mountant with DAPI (Molecular Probes) and imaged on a Leica TCS Sp8, using a

63x/1.40 oil objective. Optical z sections were acquired (z-step size 0.3 microns) for each scan to cover the depth of the cells.

DRIP-seq experiments. To detect R-loops, we snap-froze cells exponentially growing in YES. Samples for DRIP-seq were prepared based on published protocols (Wahba et al. 2016; García-Rubio et al. 2018b; Hartono et al. 2018), with minor variations. The main difference was in the method for nucleic acid purification. Briefly, cells were thawed on ice and re-suspended in the same buffer used for ChIP-seq chromatin extraction. Cells were homogenized using a Fast-prep instrument, and chromatin was precipitated by centrifugation. Chromatin was re-suspended in Qiagen buffer G2 (from genomic extraction kit) with Proteinase K and incubated for 40 min at 55°C. Nucleic acids were purified by phenol:chloroform extraction and precipitation, digested with S1 nuclease and sonicated. Half of the sample was treated with RNAseH to use as control. R-loop-IP was performed using Dynabeads-mouse M280 pre-coated with S9.6 antibody (Millipore MABE1095). DNA was then eluted and purified for library preparation with NEBNext Ultra kits (NEB).

For data analysis, the depth at every base was collected using SAMtools and divided by the total number of mapped reads for each sample to normalise for differences in sequencing depth. To compare the R-loop signal at Scw1 target genes between wt and scw1Δ, we used a Python script to collect the mean normalised coverage for each target gene, including both those that were identified by Hasan and colleagues as upregulated in scw1Δ and those that were enriched in Scw1-TAP RIP-chip (Hasan et al. 2014).

Acknowledgements

We thank Caia Duncan and Juan Mata for Scw1 strains, Siôn L. Williams for providing human sequence data, Pawan Dhami (Genomics and Genome Engineering Facility funded by the Cancer Research UK-UCL Centre) for help with sequencing, and Mimoza Hoti for assistance with some of the sequencing runs. This research was funded by a BBSRC-DTP studentship to D.A.E. (London Interdisciplinary Doctoral Programme) [grant number

BB/M009513/1] and by a Wellcome Trust Senior Investigator Award to J.B. [grant number 095598/Z/11/Z].

FIGURE LEGENDS

Figure 1. Breakpoint junctions increase in non-dividing, ageing cells.

(A) Spontaneous rearrangements that occur in non-dividing cells should be heterogeneously distributed throughout the population and identifiable in DNA sequencing data.

(B) Proportion of junctions expected to fall in coding regions *vs* proportion that do (based on simulated data; Materials and Methods). Points represent biological (yellow) or *in-silico* (grey) repeats. One-sided two sample Mann-Whitney U to test whether observed is less than expected ($U=0$, $p < 0.001$, $N=8$). Note that proportion of simulated junctions in coding regions is similar to proportion of genome reported to be coding (excluding introns (Wood et al. 2002)).

(C) Number of breakpoint Junctions Per Mapped Read (JPMR) passing filter (Materials and Methods) as a function of cellular age. Cells were cultured in rich medium until the optical density no longer increased, indicating that cells were no longer dividing, and used as day zero for the chronological ageing timecourse. Samples were taken from the cultures at 0, 1, 2, 3, 4 and 5 days. JPMR are shown relative to day zero ($N=8$, confidence interval=68%). Ochre line: total junctions, genome wide. Grey lines from top to bottom: subset of junctions formed among fragments of nuclear DNA (nDNA); between fragments of mitochondrial (mtDNA) and nDNA; among fragments of mtDNA.

(D) Histograms showing number of junctions with various lengths of microhomology (right of dotted lines), blunt joints (between dotted lines), or non-homologous insertion (left of dotted lines) at the breakpoint and during cellular ageing. Histograms at different days are overlaid on top of each other, with early timepoints in blue through to late timepoints in red (see colour legend). Right: schemes for reads with each type of junction.

Figure 2. Similar patterns of genome rearrangement in ageing human brain cells as in ageing yeast cells.

Analysis was performed on sequence data from putamen samples of neurologically normal Caucasian males (Williams et al. 2013), six of whom were young (blue; 19, 24, 28, 28, 29 and 30 year old donors) and six of whom were old (red; 67, 71, 78, 83, 85 and 89 year old donors). The 100bp reads from these samples, which were crudely enriched for mtDNA, were re-mapped to the whole human genome and filtered (Materials and methods).

(A) Histograms showing number of junctions with various lengths of microhomology (right of dotted lines), blunt joints (between dotted lines), or non-homologous insertions (left of dotted lines) at breakpoint. Histograms at different ages are overlaid on top of each other, with samples from young donors in blue and older donors in red (as in A).

(B) Proportion of junctions expected to fall in coding regions *vs* proportion that do (based on simulated data; Materials and Methods). One-sided two sample Mann-Whitney U to test whether observed is less than expected ($U=3$, $p < 0.001$). Simulated proportions approached the 1.5% proportion of the genome reported as coding (Consortium 2012) when more junctions were simulated (data not shown).

Figure 3. Local and global hotspots for age-associated junction formation.

(A) Illustration of complexity of junction formation over time. In this example, junctions are represented as a link between two points in any of the three *S. pombe* chromosomes (depicted around the edge in Mbp). The number of junctions in a region are stacked up in the outer ring. Blue: although a region may have many junctions at a late time point (bottom), these junctions may not be the result of ageing as they could have been present earlier (top). Red: true age-associated hotspots will show an age-associated increase. Some rearrangements may occur recurrently between the same two regions (green) or one region with many others (red).

(B) Heatmap showing the ratio of junction counts between Days 5 and 0 (T5/T0; 'old-young ratio'), in 20kb windows across the genome. An element in the heatmap depicting two

regions with high levels of age-associated breakpoint formation will have a higher old-young ratio and appear darker. The average number of breakpoints at each window, genome-wide, is plotted in grey at the top. An intra-chromosomal region of Chromosome II is blown up in bottom right; local hotspots are exemplified by region II:2940000-3080000, where the old-young ratio exceeds 25 in many bins along the top edge of the heatmap. Global hotspots are reflected by the dark diagonal emanating from the right end of Chromosome II and both ends of Chromosome III.

Figure 4. Bodies of RNA associated with DNA at global hotspots.

(A) Scatter plot showing junctions at each of the global hotspots identified near telomeres (Fig. 3). Coordinates in ~20kb region surrounding each hotspot are shown on X axis, and coordinates at Chromosomes I-III are shown on Y axis. Annotations including pseudogenes (grey), *tlh2* (red) and rRNA genes (black) are shown beneath X axis.

(B) smFISH experiment using probes against the housekeeping control gene *rbp1* (green, left) and *tlh2* (purple, middle and right), with DAPI-stained nuclei (blue, right). Non-dividing wild-type (wt) cells do not appear to show any *tlh2* expression after 2 days (top row) or 4 days (second row) of chronological ageing, proliferating cells overexpressing *tlh2* (tlh2OE) show strong, but highly heterogeneous expression, while proliferating *sir2* deletion mutants (*sir2Δ*) show homogeneous high expression in all cells.

(C) Top: RNA mapping to an unannotated region downstream of *tlh2* in young (100% cell viability) and old (50% cell viability) stationary-phase cells. Each sample's read coverage is normalised to the total number of mapped reads for that sample, showing the mean of two replicates (reanalysed from ref. 67). Bottom: DRIP-seq in proliferating cells using the S9.6 antibody shows that R-loops accumulate at the same region.

Figure 5. In absence of Scw1, R-loops may contribute to local hotspots.

(A) Top: sliding mean (10bp windows) junction counts around coding regions (CDS) genome wide. Bottom: mean ± standard error of junction counts in 250bp bins on either side of start and end of coding regions, genome wide. Comparisons between bins were performed with paired t-tests (5' out *vs* 5' in: $p<0.0001$; 5' out *vs* 3' out: $p<0.001$; 5' in *vs* 3' in: n.s.; 3' in *vs* 3' out: $p<0.0001$).

(B) Scatter plot of 3' UTR length *vs* number of junctions in that UTR. Each point represents a gene. The X axis is log scale to spread the data. Dotted lines show two standard deviations above population mean (solid line). All genes above dotted line were used for gene enrichment analysis. Scw1 targets greater than twice population standard deviation are shown in orange.

(C) Western blot of Scw1-TAP at different days of cellular ageing. Coomasie blue was used as loading control (not shown).

(D) Left: rapidly proliferating and early stationary phase cells (equivalent to Day 0 in C), which normally express Scw1, were analysed for R-loop formation in presence (wt) and absence (*scw1Δ*) of Scw1. Right: example images from chromosomal spreads; DNA stained with DAPI (blue) and R-loops by S9.6 antibody(Boguslawski et al. 1986) (yellow), with (+) or without (-) RNase H treatment (which removes R-loops). Scale bar: 10µm.

(E) Normalised read coverage at Scw1 target genes after DRIP-seq of *scw1Δ* and wt cells using the S9.6 antibody (ochre). The difference between strains is significant (one sample t-test: $T=-2.52$, $p=0.012$). Samples after the addition of RNase H are shown as a control (blue).

(F) Model for how reduced Scw1 levels might lead to genome rearrangement at its targets. In young cells, Scw1 is present and binds its targets at their 3' UTRs. In old cells, Scw1 is absent and thus leaves nascent RNA of its targets free to anneal with template ssDNA.

Figure S1. Pipeline used for identification of breakpoint junctions. Raw sequencing reads (126 nt paired-end, Illumina) were aligned to the reference genome with BWA-MEM. Alignment files were then converted to BAM format and sorted before having PCR duplicates removed in Samtools. Next, Bash commands (Awk and Grep) were used to extract split reads from these files. A custom Python script was used to filter these split reads to obtain a robust set of junctions representing two juxtaposed DNA locations, and to characterise their microhomology use across the junction.

Figure S2. Proportion of different junction types is not proportional to available DNA during library preparation. Pie charts showing numbers of junctions between mitochondrial and nuclear (green), mitochondrial and mitochondrial (purple), and nuclear and nuclear (orange) DNA in simulated data (left, based on read depth at mitochondrial DNA and nuclear chromosomes) and in measured data (right).

Figure S3. Local DNA is preferred to distal DNA in intra-chromosomal rearrangements, but less so at global hotspots. Kernel density and box plots showing the distribution of distances between juxtaposed pieces of DNA for any junction where one juxtaposed piece of DNA is at a global hotspot (bottom; light blue) and all other intra-chromosomal junctions at Day 5 (top; dark blue). These two groups were compared with a two-sample Wilcoxon rank sum test ($p<0.0001$).

Figure S4. Motifs at microhomology-mediated junctions.

A: All motifs discovered by MEME with e-values <0.05. Regions which aligned with known core sequence motifs of DNA-binding proteins in *S. cerevisiae* and *S. pombe* are underlined in grey and black, respectively (see B).

B: Table showing the top hits for known *S. cerevisiae* (grey rows) and *S. pombe* (blank rows) DNA-binding protein motifs that align to the discovered motifs in A. An ID of the long motif from part A is detailed in the ID column, with the shorter aligning motif in the motif column.

Figure S5. **Repeats are not responsible for extreme junction formation at hotspots.**

A: Scheme showing how, in a model where repeats do not expand, the number of repeats at a locus should not affect the early:late ratio of junction numbers.

B: Bee swarm plot showing the early:late ratio of read depth at each 20kb window, including the hotspots (red). A read depth ratio greater than 0 would indicate an increase in read depth (i.e. repeat expansion).

Figure S6. Transcriptional activity at the tlh2 locus may have a minor role in junction formation at global hotspots.

(A) Reads Per Kilobase Million (RPKM) of *tlh2* at the start of the ageing time course (100% cell viability) and when cell viability dropped to 50%. Two biological repeats of the time course were performed (each point corresponds to a repeat).

(B) Left: Scheme for calculation of proportion of junctions downstream of *tlh2* (grey region; red crosses). Right: Results from three independent repeats with wild type (WT) and *tlh2* overexpression cells (tlh2OE). One-sided two sample Mann-Whitney U to test whether tlh2OE is greater than WT (U=9, p=0.04).

(C) Chronological lifespan of tlh2OE (red) compared to WT (grey). Lines show means of three repeats ± 68% confidence intervals.

Figure S7. During ageing, Scw1 binds less to some canonical RNA targets, and does not to bind DNA.

A: RIP-ChIP data showing log2-fold enrichment in Scw1 pull downs of all canonical target RNAs (81) with similar behaviour in both repeats and data at each time point. Two repeats were sampled at Day 0 (top), Day 2 (middle) and Day 4 (bottom rows). Clusters that show reduced enrichment with age are highlighted in purple.

B: Example ChIP-seq of Scw1-TAP. Plots show normalised read depth (mean of two repeats) at three Scw1 targets (red) and three non-targets (blue) in cells aged for two days.

For this visualisation, we chose genes which do not overlap any other annotations and matched targets with non-targets of similar gene and 3' UTR lengths.

Figure S8. R-loop immunostaining controls in exponential culture.

A: Representative images of nuclear spreads from wild-type and *rnh1*Δ *rn1201*Δ double mutant cells, with the latter known to feature increased R-loop formation (Zhao et al. 2018). Nuclei are stained with DAPI (blue) and R-loops are detected with the S9.6 antibody (green). To verify that the green signal comes from R-loops, control slides (right hand panels) were treated with RNAse H before adding the primary antibody. Scale bar 10 µm.

B: R-loop signal quantification. Chi-square: $p < 0.0001$, $N > 50$.

Note S1

Arguments against breakpoint junctions forming *in-vitro*.

In addition to the selection against junctions in coding sequences (Fig. 1B), three pieces of evidence argue against the formation of these junctions *in vitro* after DNA extraction. First, during sequence library preparation, dA nucleotides were incorporated at the 3' ends of DNA fragments before ligation of the sequencing adapters (with complementary dT overhangs). We therefore expected that false positive concatemers occurring *in vitro* during sequence library preparation should feature insertion of an A at the breakpoint. On the other hand, genuine junctions may feature different signatures at the breakpoint, depending on how they were repaired *in vivo*. Whilst we found a population of single-base insertions (of which A or T insertions could reflect dA-tailing during library preparation), these junctions did not increase at all with age, and age-associated junction formation was associated with microhomology at the breakpoint (Fig. 1D). Second, modelling, based on the read depth for each body of DNA (e.g. average depth at Chromosome II), showed that the number of junctions between each type of event (e.g. mtDNA-Chromosome I) was not proportional to the relative number of free DNA ends in the library, with far fewer junctions between mitochondrial DNA and nuclear DNA than would be expected (Fig. S2). Third, the ratio of junction counts between old and young cells was higher for junctions that formed within chromosomes than those that formed between chromosomes (Fig. 3B), showing that age-associated junction formation was more prevalent within chromosomes. Furthermore, of those junctions that formed within chromosomes, nearby DNA was preferred to distant DNA (Fig. S3). This pattern echoes recent findings in the cancer genome where the frequency of junction formation between two pieces of DNA is approximately inversely proportional to the distance between them (Wala et al. 2017).

References

Aguilera A. 2002. The connection between transcription and genomic instability. *EMBO J* **21**: 195–201.

Aguilera A, García-Muse T. 2013. Causes of Genome Instability. *Annu Rev Genet*. **47**: 1-32.

Amon JD, Koshland D. 2016. RNase H enables efficient repair of R-loop induced DNA damage. *Elife* **5**: e20533.

Andrews S. 2010. FastQC: A Quality Control tool for High Throughput Sequence Data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (Accessed September 16, 2017).

Atkinson SR, Marguerat S, Bitton DA, Rodríguez-López M, Rallis C, Lemay J-F, Cotobal C, Malecki M, Smialowski P, Mata J, et al. 2018. Long noncoding RNA repertoire and targeting by nuclear exosome, cytoplasmic exonuclease, and RNAi in fission yeast. *RNA* **24**: 1195–1213.

Bähler J, Wu J-Q, Longtine MS, Shah NG, Mckenzie III A, Steever AB, Wach A, Philippsen P, Pringle JR. 1998. Heterologous modules for efficient and versatile PCR-based gene targeting in Schizosaccharomyces pombe. *Yeast* **14**: 943–951.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37**: W202–W208.

Barrales RR, Forn M, Georgescu PR, Sarkadi Z, Braun S. 2016. Control of heterochromatin localization and silencing by the nuclear membrane protein Lem2. *Genes Dev*. **30**: 133-48.

Bétermier M, Bertrand P, Lopez BS, Hendrickson E, Ramsden D. 2014. Is non-homologous end-joining really an inherently error-prone process? ed. S. Jinks-Robertson. *PLoS Genet* **10**: e1004086.

Bitton DA, Schubert F, Dey S, Okoniewski M, Smith GC, Khadayate S, Pancaldi V, Wood V, Bähler J. 2015. AnGeLi: a tool for the analysis of gene lists from fission yeast. *Front Genet* **6**: 330.

Boguslawski SJ, Smith DE, Michalak MA, Mickelson KE, Yehle CO, Patterson WL, Carrico RJ. 1986. Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids. *J Immunol Methods* **89**: 123–30.

Bohr VA. 2008. Rising from the RecQ-age: the role of human RecQ helicases in genome maintenance. *Trends Biochem Sci* **33**: 609–20.

Brown WRA, Liti G, Rosa C, James S, Roberts I, Robert V, Jolly N, Tang W, Baumann P, Green C, et al. 2011. A Geographically Diverse Collection of Schizosaccharomyces pombe Isolates Shows Limited Phenotypic Variation but Extensive Karyotypic Diversity.

ed. J.C. Fay. *G3 (Bethesda)* **1**: 615–26.

Clément-Ziza M, Marsellach FX, Codlin S, Papadakis MA, Reinhardt S, Rodríguez-López M, Martin S, Marguerat S, Schmidt A, Lee E, et al. 2014. Natural genetic variation impacts expression levels of coding, non-coding, and antisense transcripts in fission yeast. *Mol Syst Biol* **10**: 764.

Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, et al. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**: 1422–1423.

Consortium TEP. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57–74.

Cotobal C, Rodríguez-López M, Duncan C, Hasan A, Yamashita A, Yamamoto M, Bähler J, Mata J. 2015. Role of Ccr4-Not complex in heterochromatin formation at meiotic genes and subtelomeres in fission yeast. *Epigenetics Chromatin* **8**: 28.

Curtis HJ. 1963. Biological Mechanisms Underlying the Aging Process. *Science (80-)* **141**. http://science.sciencemag.org/content/141/3582/686 (Accessed June 14, 2018).

Decottignies A. 2013. Alternative end-joining mechanisms: a historical perspective. *Front Genet* **4**: 48.

Decottignies A. 2005. Capture of extranuclear DNA at fission yeast double-strand breaks. *Genetics* **171**: 1535–48.

Decottignies A. 2007. Microhomology-Mediated End Joining in Fission Yeast Is Repressed by Pku70 and Relies on Genes Involved in Homologous Recombination. *Genetics* **176**: 1403–1415.

Difilippantonio MJ, Zhu J, Chen HT, Meffre E, Nussenzweig MC, Max EE, Ried T, Nussenzweig A. 2000. DNA repair protein Ku80 suppresses chromosomal aberrations and malignant transformation. *Nature* **404**: 510–514.

Ellis DA, Mustonen V, Rodríguez-López M, Rallis C, Malecki M, Jeffares DC, Bähler J. 2018. Uncovering Natural Longevity Alleles from Intercrossed Pools of Aging Fission Yeast Cells. *Genetics* genetics.301262.2018.

Emde A-K, Schulz MH, Weese D, Sun R, Vingron M, Kalscheuer VM, Haas SA, Reinert K. 2012. Detecting genomic indel variants with exact breakpoints in single- and paired-end sequencing data using SplazerS. *Bioinformatics* **28**: 619–27.

Feldman JL, Peterson CL. 2019. Yeast Sirtuin Family Members Maintain Transcription Homeostasis to Ensure Genome Stability. *Cell Rep*. **27**: 2978-89.

Ferreira MG, Cooper JP. 2004. Two modes of DNA double-strand break repair are reciprocally regulated through the fission yeast cell cycle. *Genes Dev*. **18**: 2249–54.

Forsburg SL. 1993. Comparison of *Schizosaccharomyces pombe* expression systems. *Nucleic Acids Res* **21**: 2955–6.

Gaillard H, Aguilera A. 2016. Transcription as a Threat to Genome Integrity. *Annu Rev Biochem* **85**: 291–317.

Gangloff S, Achaz G, Francesconi S, Villain A, Miled S, Denis C, Arcangioli B. 2017. Quiescence unveils a novel mutational force in fission yeast. *Elife* **6**: e27469.

García-Rubio M, Aguilera P, Lafuente-Barquero J, Ruiz JF, Simon MN, Geli V, Rondón AG, Aguilera A. 2018a. Yra1-bound RNA–DNA hybrids cause orientation-independent transcription– replication collisions and telomere instability. *Genes Dev*. **32**: 965-977.

García-Rubio M, Barroso SI, Aguilera A. 2018b. Detection of DNA-RNA hybrids in vivo. In *Methods in Mol Biol*. **1672**: 347-361.

Ginno PA, Lott PL, Christensen HC, Korf I, Chédin F. 2012. R-Loop Formation Is a Distinctive Characteristic of Unmethylated Human CpG Island Promoters. *Mol Cell*. **45**: 814-25.

Gorbunova V, Seluanov A. 2016. DNA double strand break repair, aging and the chromatin connection. *Mutat Res* **788**: 2–6.

Guo Y, Mahony S, Gifford DK. 2012. High Resolution Genome Wide Binding Event Finding and Motif Discovery Reveals Transcription Factor Spatial Binding Constraints ed. S. Aerts. *PLoS Comput Biol* **8**: e1002638.

Haber JE. 2016. A Life Investigating Pathways That Repair Broken Chromosomes. *Annu Rev Genet* **50**: 1–28.

Hartono SR, Malapert A, Legros P, Bernard P, Chédin F, Vanoosthuyse V. 2018. The Affinity of the S9.6 Antibody for Double-Stranded RNAs Impacts the Accurate Mapping of R-Loops in Fission Yeast. *J Mol Biol*. **430**: 272-284.

Hasan A, Cotobal C, Duncan CDS, Mata J, Penkett C. 2014. Systematic Analysis of the Role of RNA-Binding Proteins in the Regulation of RNA Stability ed. G.P. Copenhaver. *PLoS Genet* **10**: e1004684.

Hasty P, Campisi J, Hoeijmakers J, van Steeg H, Vijg J. 2003. Aging and Genome Maintenance: Lessons from the Mouse? *Science* **299**: 1355–1359.

Heyer WD. 2015. Regulation of recombination and genomic maintenance. *Cold Spring Harb Perspect Med* **5**: 1–24.

Holtgrewe M. 2010. Mason – A Read Simulator for Second Generation Sequencing Data. *Tech Rep FU Berlin*. http://publications.imp.fu-berlin.de/962/ (Accessed November 4, 2018).

Hoyos-Manchado R, Reyes-Martín F, Rallis C, Gamero-Estévez E, Rodríguez-Gómez P, Quintero-Blanco J, Bähler J, Jiménez J, Tallada VA. 2017. RNA metabolism is the primary target of formamide in vivo. *Sci Rep* **7**: 15895.

Hu W, Suo F, Du L-L. 2015. Bulk Segregant Analysis Reveals the Genetic Basis of a Natural Trait Variation in Fission Yeast. *Genome Biol Evol* **7**: 3496–510.

Hu Z, Chen K, Xia Z, Chavez M, Pal S, Seol JH, Chen CC, Li W, Tyler JK. 2014. Nucleosome loss leads to global transcriptional up-regulation and genomic instability during yeast aging. *Genes Dev* **28**: 396–408.

Huertas P, Aguilera A. 2003. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol Cell* **12**: 711–21.

Jacobs KB, Yeager M, Zhou W, Wacholder S, Wang Z, Rodriguez-Santiago B, Hutchinson A, Deng X, Liu C, Horner M-J, et al. 2012. Detectable clonal mosaicism and its relationship to aging and cancer. *Nat Genet* **44**: 651–658.

Jeffares DC, Jolly C, Hoti M, Speed D, Shaw L, Rallis C, Balloux F, Dessimoz C, Bähler J, Sedlazeck FJ. 2017. Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nat Commun* **8**: 14061.

Jeffares DC, Rallis C, Rieux A, Speed D, Převorovský M, Mourier T, Marsellach FX, Iqbal Z, Lau W, Cheng TMK, et al. 2015. The genomic and phenotypic diversity of *Schizosaccharomyces pombe*. *Nat Genet* **47**: 235–241.

Jiang Y, Wang Y, Brudno M. 2012. PRISM: Pair-read informed split-read mapping for base-pair level detection of insertion, deletion and structural variants. *Bioinformatics* **28**: 2576–2583.

Jin Q-W, McCollum D. 2003. Scw1p antagonizes the septation initiation network to regulate septum formation and cell separation in the fission yeast *Schizosaccharomyces pombe*. *Eukaryot Cell* **2**: 510–20.

Kaeberlein M. 2010. Lessons on longevity from budding yeast. *Nature* **464**: 513–519.

Karagiannis J, Oulton R, Young PG. 2002. The Scw1 RNA-binding domain protein regulates septation and cell-wall structure in fission yeast. *Genetics* **162**: 45–58.

Kennedy SR, Salk JJ, Schmitt MW, Loeb LA. 2013. Ultra-Sensitive Sequencing Reveals an Age-Related Increase in Somatic Mitochondrial Mutations That Are Inconsistent with Oxidative Damage ed. B. Van Houten. *PLoS Genet* **9**: e1003794.

Kim DU, Hayles J, Kim D, Wood V, Park HO, Won M, Yoo HS, Duhig T, Nam M, Palmer G, et al. 2010. Analysis of a genome-wide set of gene deletions in the fission yeast Schizosaccharomyces pombe. *Nat Biotechnol*. **28**: 617-623.

Kimura K, Koike A. 2015. Analysis of genomic rearrangements by using the Burrows-Wheeler transform of short-read data. *BMC Bioinformatics* **16**: S5.

Kobayashi T. 2014. Ribosomal RNA gene repeats, their stability and cellular senescence. *Proc Jpn Acad Ser B Phys Biol Sci* **90**: 119–29.

Laurie CC, Laurie CA, Rice K, Doheny KF, Zelnick LR, McHugh CP, Ling H, Hetrick KN, Pugh EW, Amos C, et al. 2012. Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nat Genet* **44**: 642–650.

Layer RM, Chiang C, Quinlan AR, Hall IM. 2014. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol* **15**: R84.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*: 1303.3997.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Li X, Manley JL. 2005. Inactivation of the SR Protein Splicing Factor ASF/SF2 Results in Genomic Instability. *Cell* **122**: 365–378.

Lombard DB, Chua KF, Mostoslavsky R, Franco S, Gostissa M, Alt FW. 2005. DNA Repair, Genome Stability, and Aging. *Cell* **120**: 497–512.

López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. 2013. The Hallmarks of Aging. *Cell* **153**: 1194–1217.

Lu KL, Nelson JO, Watase GJ, Warsinger-Pepe N, Yamashita YM. 2018. Transgenerational dynamics of rDNA copy number in Drosophila male germline stem cells. *Elife* **7**: e32421.

Maestroni L, Audry J, Matmati S, Arcangioli B, Géli V, Coulon S. 2017. Eroded telomeres are rearranged in quiescent fission yeast cells through duplications of subtelomeric sequences. *Nat Commun*. **8**: 1684.

Magrassi L, Leto K, Rossi F. 2013. Lifespan of neurons is uncoupled from organismal lifespan. *Proc Natl Acad Sci U S A* **110**: 4374–9.

Makarenko R, Denis C, Francesconi S, Gangloff S, Arcangioli B. 2020. Nitrogen starvation reveals the mitotic potential of mutants in the S/MAPK pathways. *Nat Commun*. **11**: 1973.

Mandell JG, Goodrich KJ, Bähler J, Cech TR. 2005. Expression of a RecQ helicase homolog affects progression through crisis in fission yeast lacking telomerase*. *J Biol Chem*. **280**: 5249-57.

McDowall MD, Harris MA, Lock A, Rutherford K, Staines DM, Bähler J, Kersey PJ, Oliver SG, Wood V. 2015. PomBase 2015: updates to the fission yeast database. *Nucleic Acids Res* **43**: D656-61.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**: 1297–303.

Merino MM, Rhiner C, Lopez-Gay JM, Buechel D, Hauert B, Moreno E. 2015. Elimination of

Unfit Cells Maintains Tissue Health and Prolongs Lifespan. *Cell* **160**: 461–476.

Mochida S, Yanagida M. 2006. Distinct modes of DNA damage response in S. pombe Go and vegetative cells. *Genes to Cells*. **11**: 13-27.

Murray JM, Watson AT, Carr AM. 2016. Extraction of Chromosomal DNA from Schizosaccharomyces pombe. *Cold Spring Harb Protoc* **2016**: pdb.prot090985.

Nakayama H, Nakayama K, Nakayama R, Irino N, Nakayama Y, Hanawalt PC. 1984. Isolation and genetic characterization of a thymineless death-resistant mutant of Escherichia coli K12: identification of a new mutation (recQ1) that blocks the RecF recombination pathway. *Mol Gen Genet* **195**: 474–80.

Ottaviani D, LeCain M, Sheer D. 2014. The role of microhomology in genomic structural variation. *Trends Genet* **30**: 85–94.

Quinlan AR, Hall IM, Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.

Quispe-Tintaya W, Gorbacheva T, Lee M, Makhortov S, Popov VN, Vijg J, Maslov AY. 2016. Quantitative detection of low-abundance somatic structural variants in normal cells by high-throughput sequencing. *Nat Methods* **13**: 584–6.

Rallis C, Codlin S, Bähler J. 2013. TORC1 signaling inhibition by rapamycin and caffeine affect lifespan, global gene expression, and cell proliferation of fission yeast. *Aging Cell* **12**: 563–73.

Rallis C, López-Maury L, Georgescu T, Pancaldi V, Bähler J. 2014. Systematic screen for mutants resistant to TORC1 inhibition in fission yeast reveals genes involved in cellular ageing and growth. *Biol Open* **3**: 161-71.

Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO. 2012. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**: i333–i339.

Ricchetti M, Fairhead C, Dujon B. 1999. Mitochondrial DNA repairs double-strand breaks in yeast chromosomes. *Nature* **402**: 96–100.

Roux AE, Leroux A, Alaamery MA, Hoffman CS, Chartrand P, Ferbeyre G, Rokeach LA. 2009. Pro-aging effects of glucose signaling through a G protein-coupled glucose receptor in fission yeast. ed. S.K. Dutcher. *PLoS Genet* **5**: e1000408.

Santos-Pereira JM, Aguilera A. 2015. R loops: new modulators of genome dynamics and function. *Nat Rev Genet* **16**: 583–597.

Saponaro M, Kantidakis T, Mitter R, Kelly GP, Heron M, Williams H, Söding J, Stewart A, Svejstrup JQ. 2014. RECQL5 controls transcript elongation and suppresses genome instability associated with transcription stress. *Cell* **157**: 1037–49.

Schlissel G, Krzyzanowski MK, Caudron F, Barral Y, Rine J. 2017. Aggregation of the Whi3 protein, not loss of heterochromatin, causes sterility in old yeast cells. *Science (80-)*

355: 1184-1187.

Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**: 671–675.

Shimamoto A, Kitao S, Ichikawa K, Suzuki N, Yamabe Y, Imamura O, Tokutake Y, Satoh M, Matsumoto T, Kuromitsu J, et al. 1996. A unique human gene that spans over 230 kb in the human chromosome 8p11-12 and codes multiple family proteins sharing RNA-binding motifs. *Proc Natl Acad Sci U S A*. **93**: 10913–10917.

Sindi SS, Onal S, Peng LC, Wu H-T, Raphael BJ. 2012. An integrative probabilistic model for identification of structural variation in sequencing data. *Genome Biol* **13**: R22.

Skourti-Stathaki K, Proudfoot NJ. 2014. A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. *Genes Dev* **28**: 1384–1396.

Soutoglou E, Dorn JF, Sengupta K, Jasin M, Nussenzweig A, Ried T, Danuser G, Misteli T. 2007. Positional stability of single double-strand breaks in mammalian cells. *Nat Cell Biol* **9**: 675–682.

Spivey EC, Jones SK, Rybarski JR, Saifuddin FA, Finkelstein IJ. 2017. An aging-independent replicative lifespan in a symmetrically dividing eukaryote. *Elife* **6**: e20340.

Strehler BL. 1986. Genetic instability as the primary cause of human aging. *Exp Gerontol* **21**: 283–319.

Sun XM, Bowman A, Priestman M, Bertaux F, Martinez-Segura A, Tang W, Whilding C, Dormann D, Shahrezaei V, Marguerat S. 2020. Size-dependent increase in RNA polymerase II initiation rates mediates gene expression scaling with cell size. *Curr Biol*. **30**: 1217-1230.

Suzuki S, Yasuda T, Shiraishi Y, Miyano S, Nagasaki M. 2011. ClipCrop: a tool for detecting structural variations with single-base resolution using soft-clipping information. *BMC Bioinformatics* **12**: S7.

Szilard L. 1959. On the nature of the aging process. *Proc Natl Acad Sci U S A* **45**: 30–45.

Toda T, Nakaseko Y, Niwa O, Yanagida M. 1984. Mapping of rRNA genes by integration of hybrid plasmids in Schizosaccharomyces pombe. *Curr Genet* **8**: 93–97.

Trappe K, Emde A-K, Ehrlich H-C, Reinert K. 2014. Gustaf: Detecting and correctly classifying SVs in the NGS twilight zone. *Bioinformatics* **30**: 3484–3490.

Vaidya A, Mao Z, Tian X, Spencer B, Seluanov A, Gorbunova V. 2014. Knock-in reporter mice demonstrate that DNA repair by non-homologous end joining declines with age. *PLoS Genet* **10**: e1004511.

Vijg J, Suh Y. 2013. Genome Instability and Aging. *Annu Rev Physiol* **75**: 645–668.

Villarreal DD, Lee K, Deem A, Shim EY, Malkova A, Lee SE. 2012. Microhomology Directs Diverse DNA Break Repair Pathways and Chromosomal Translocations ed. M. Lichten.

PLoS Genet **8**: e1003026.

Wahba L, Amon JD, Koshland D, Vuica-Ross M. 2011. RNase H and Multiple RNA Biogenesis Factors Cooperate to Prevent RNA:DNA Hybrids from Generating Genome Instability. *Mol Cell* **44**: 978–988.

Wahba L, Costantino L, Tan FJ, Zimmer A, Koshland D. 2016. S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. *Genes Dev*.

Wala JA, Shapira O, Li Y, Craft D, Schumacher SE, Imielinski M, Haber JE, Roberts N, Yao X, Stewart C, et al. 2017. Selective and mechanistic sources of recurrent rearrangements across the cancer genome. *bioRxiv* 187609.

Wang J, Mullighan CG, Easton J, Roberts S, Heatley SL, Ma J, Rusch MC, Chen K, Harris CC, Ding L, et al. 2011. CREST maps somatic structural variation in cancer genomes with base-pair resolution. *Nat Methods* **8**: 652–654.

Wei P-C, Chang AN, Kao J, Du Z, Meyers RM, Alt FW, Schwer B. 2016. Long Neural Genes Harbor Recurrent DNA Break Clusters in Neural Stem/Progenitor Cells. *Cell* **164**: 644–655.

White RR, Vijg J. 2016. Do DNA Double-Strand Breaks Drive Aging? *Mol Cell* **63**: 729–738.

Williams SL, Mash DC, Züchner S, Moraes CT. 2013. Somatic mtDNA mutation spectra in the aging human putamen. *PLoS Genet* **9**: e1003990.

Wood V, Gwilliam R, Rajandream M-A, Lyne M, Lyne R, Stewart A, Sgouros J, Peat N, Hayles J, Baker S, et al. 2002. The genome sequence of Schizosaccharomyces pombe. *Nature* **415**: 871–880.

Wood V, Harris MA, McDowall MD, Rutherford K, Vaughan BW, Staines DM, Aslett M, Lock A, Bähler J, Kersey PJ, et al. 2012. PomBase: a comprehensive online resource for fission yeast. *Nucleic Acids Res* **40**: D695-9.

Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. 2009. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25**: 2865–2871.

Zhao H, Zhu M, Limbo O, Russell P. 2018. RNase H eliminates R-loops that disrupt DNA replication but is nonessential for efficient DSB repair. *EMBO Rep*. **19**: e4533.

Zolan ME, Cortopassi GA, Smith CA, Hanawalt PC. 1982. Deficient repair of chemical adducts in alpha DNA of monkey cells. *Cell* **28**: 613–9.

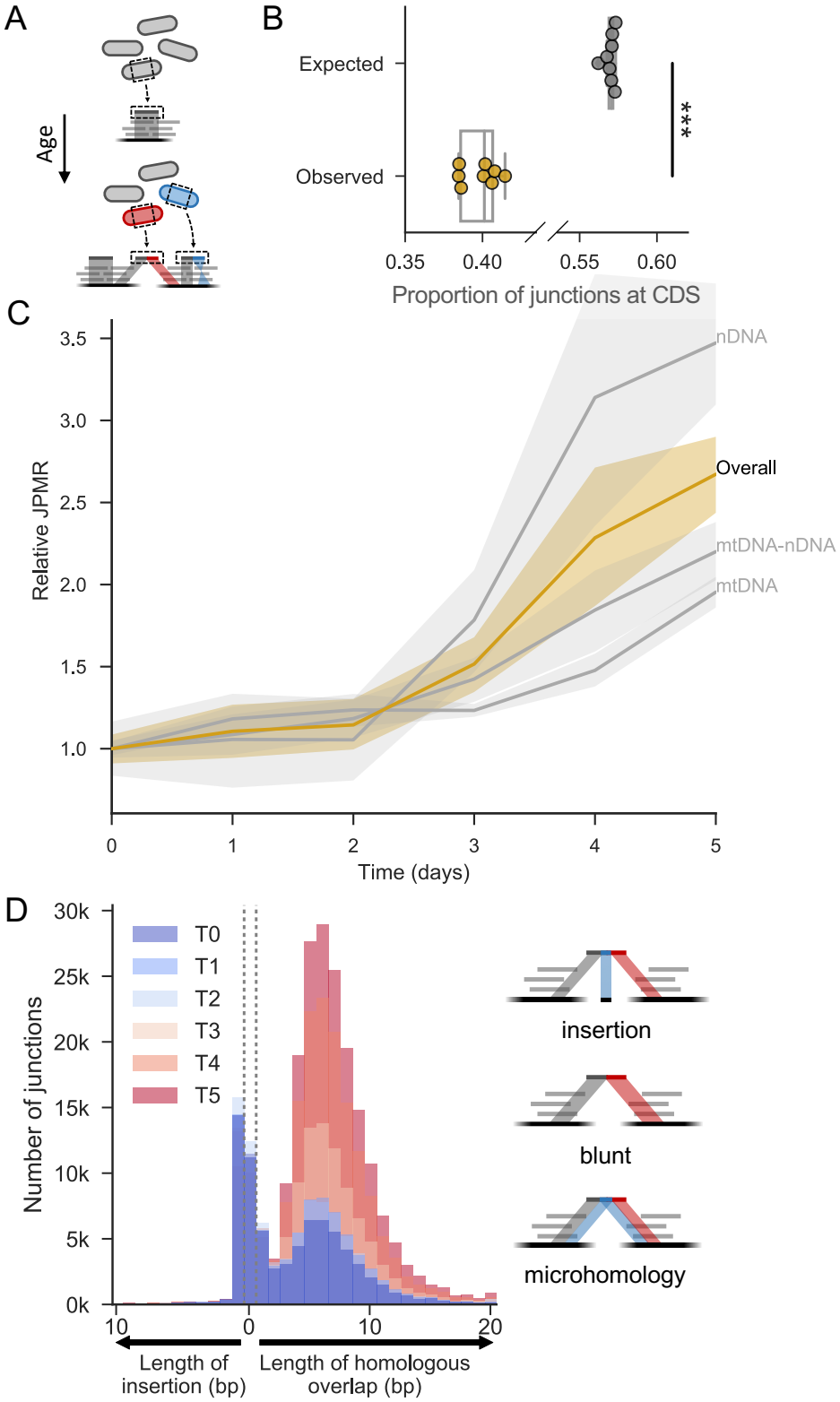

A
Age
B
Expected
Observed
***
0.35
0.40
0.55
0.60
Proportion of junctions at CDS
C
Relative JPMR
3.5
3.0
2.5
2.0
1.5
1.0
nDNA
Overall
mtDNA-nDNA
mtDNA
0
1
2
3
4
5
Time (days)
D
Number of junctions
30k
25k
20k
15k
10k
5k
0k
T0
T1
T2
T3
T4
T5
10
0
10
20
Length of insertion (bp)
Length of homologous overlap (bp)
insertion
blunt
microhomology

Figure 1.

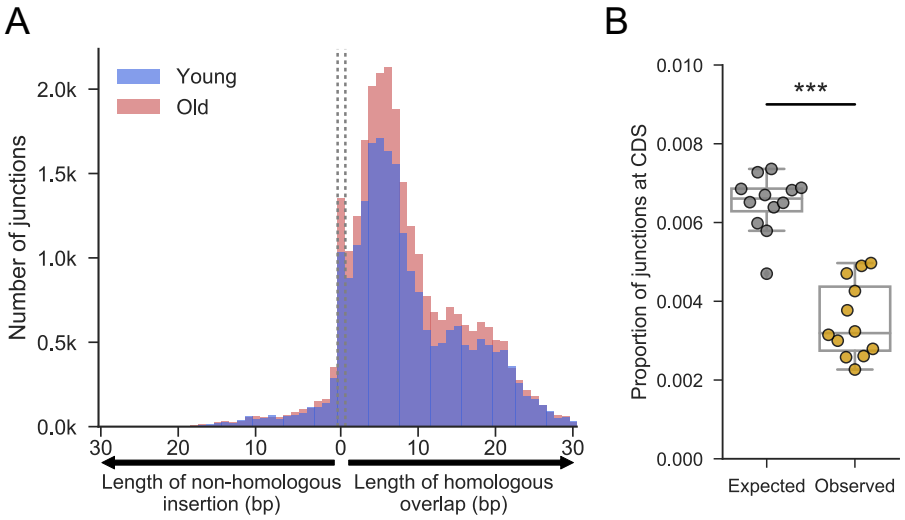

A
Young
Old
Number of junctions
2.0k
1.5k
1.0k
0.5k
0.0k
30
20
10
0
10
20
30
Length of non-homologous insertion (bp)
Length of homologous overlap (bp)
B
***
Proportion of junctions at CDS
0.010
0.008
0.006
0.004
0.002
0.000
Expected
Observed

Figure 2.

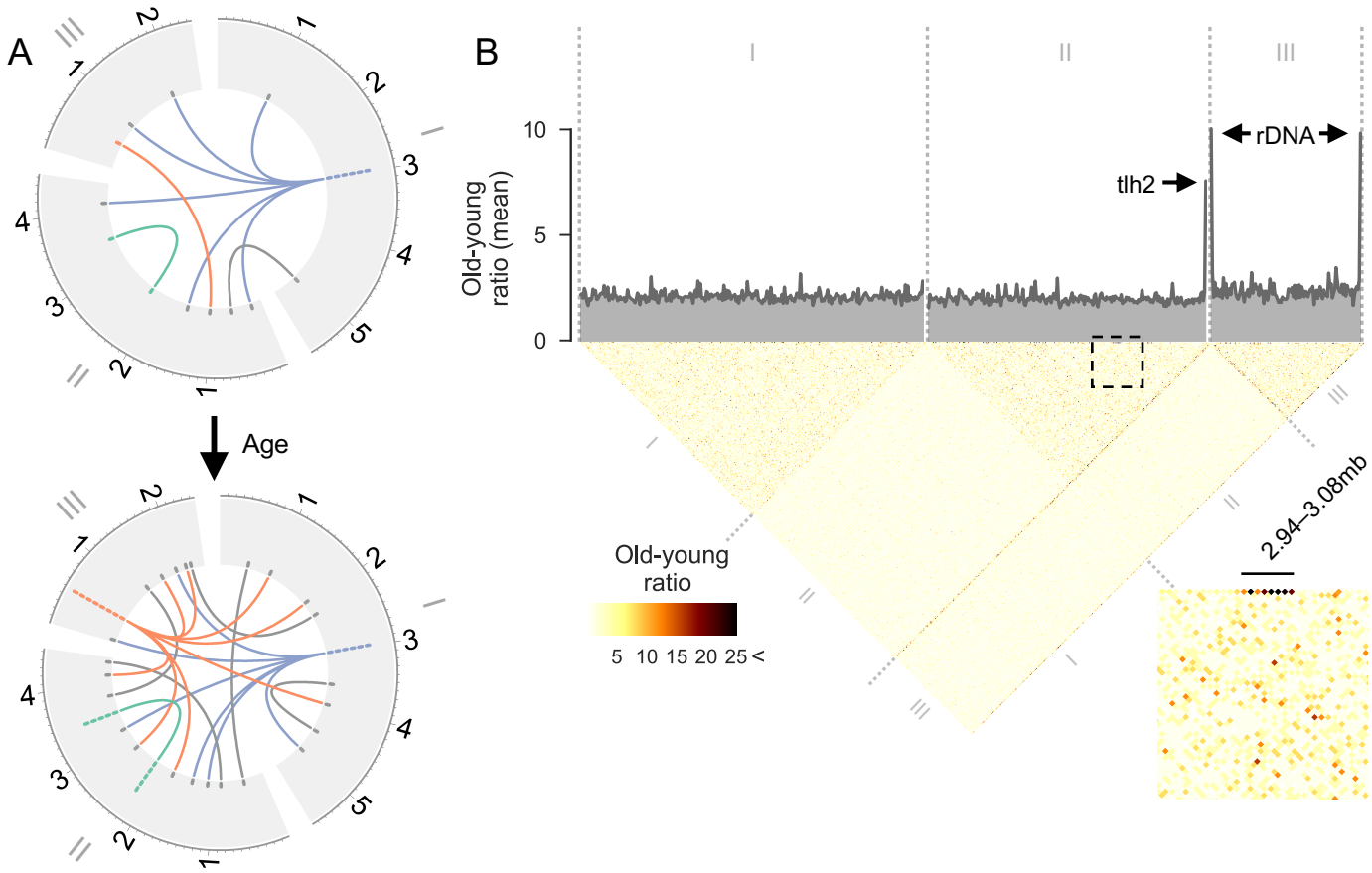
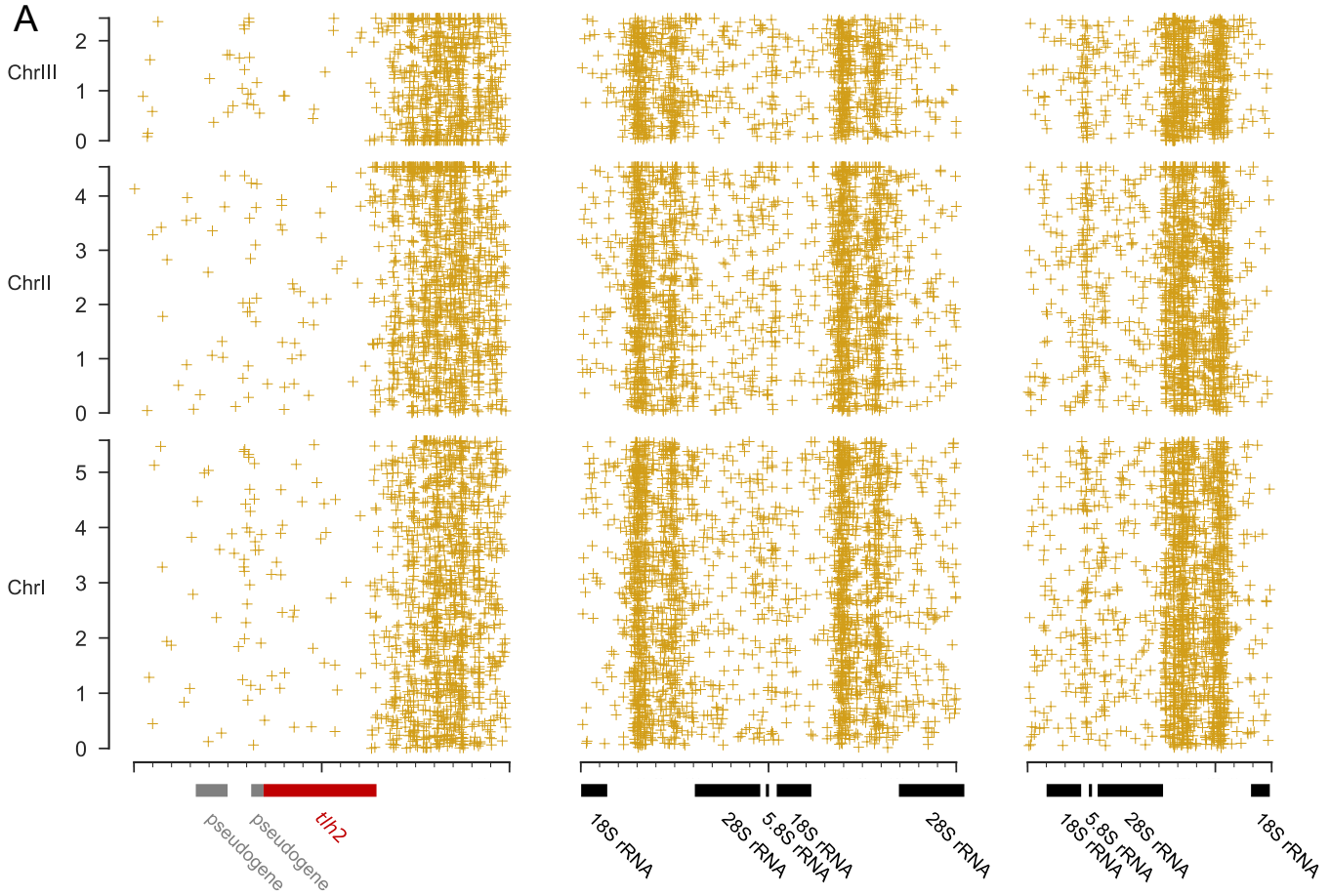
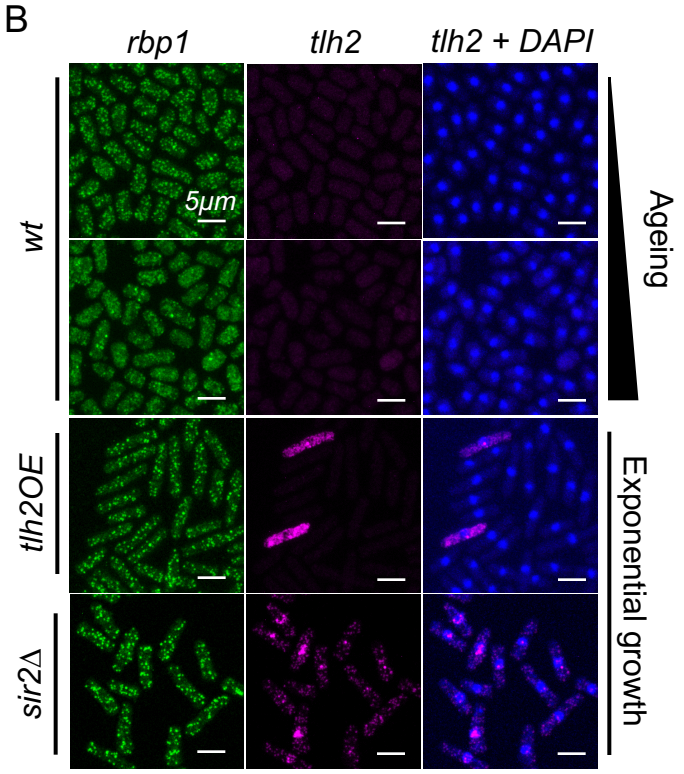

A
B
Age
Old-young ratio (mean)
10
5
0
rDNA
tlh2
Old-young ratio
5 10 15 20 25 <
2.94–3.08mb

Figure 3.


A
ChrIII
ChrII
ChrI
pseudogene
pseudogene
tlh2
18S rRNA
28S rRNA
5.8S rRNA
18S rRNA
28S rRNA
18S rRNA
5.8S rRNA
28S rRNA
18S rRNA


B
rbp1
tlh2
tlh2 + DAPI
5μm
wt
tlh2OE
sir2Δ
Ageing
Exponential growth

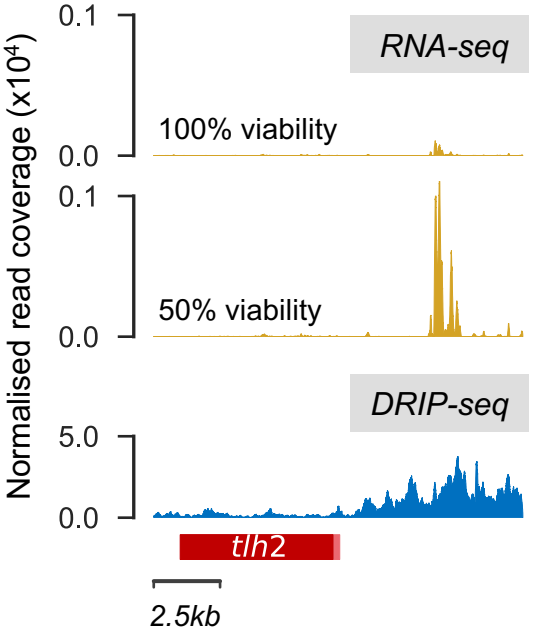

C
Normalised read coverage (x10⁴)
RNA-seq
100% viability
50% viability
DRIP-seq
tlh2
2.5kb

Figure 4.

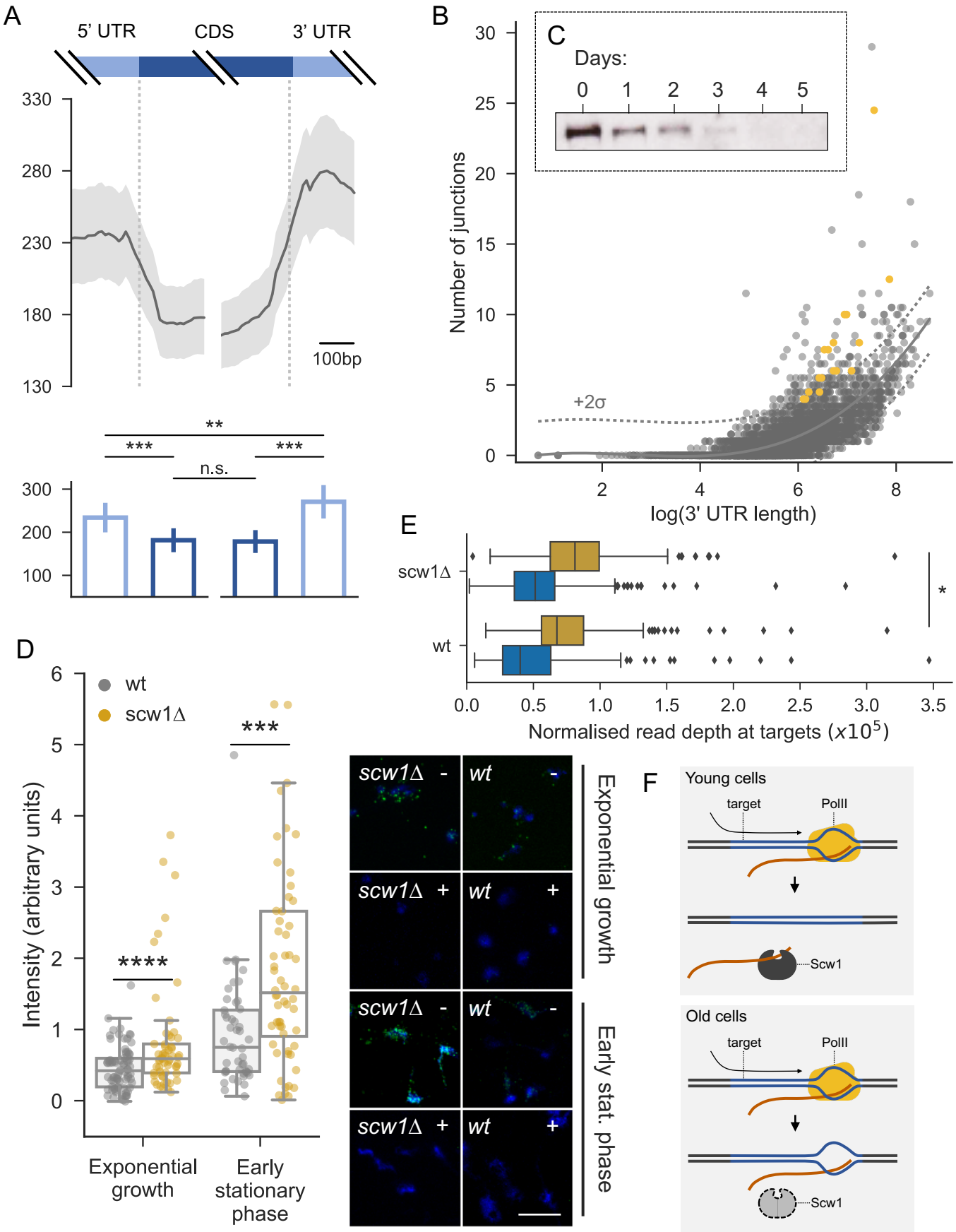

A
5' UTR
CDS
3' UTR
330
280
230
180
130
100bp
**
***
***
n.s.
300
200
100
B
C
Days:
0 1 2 3 4 5
Number of junctions
+2σ
log(3' UTR length)
E
scw1Δ
wt
*
Normalised read depth at targets (x10^5)
D
wt
scw1Δ
Intensity (arbitrary units)
****
***
Exponential growth
Early stationary phase
scw1Δ -
wt -
scw1Δ +
wt +
Exponential growth
Early stat. phase
F
Young cells
target
PolII
Scw1
Old cells
target
PolII
Scw1

Figure 5.

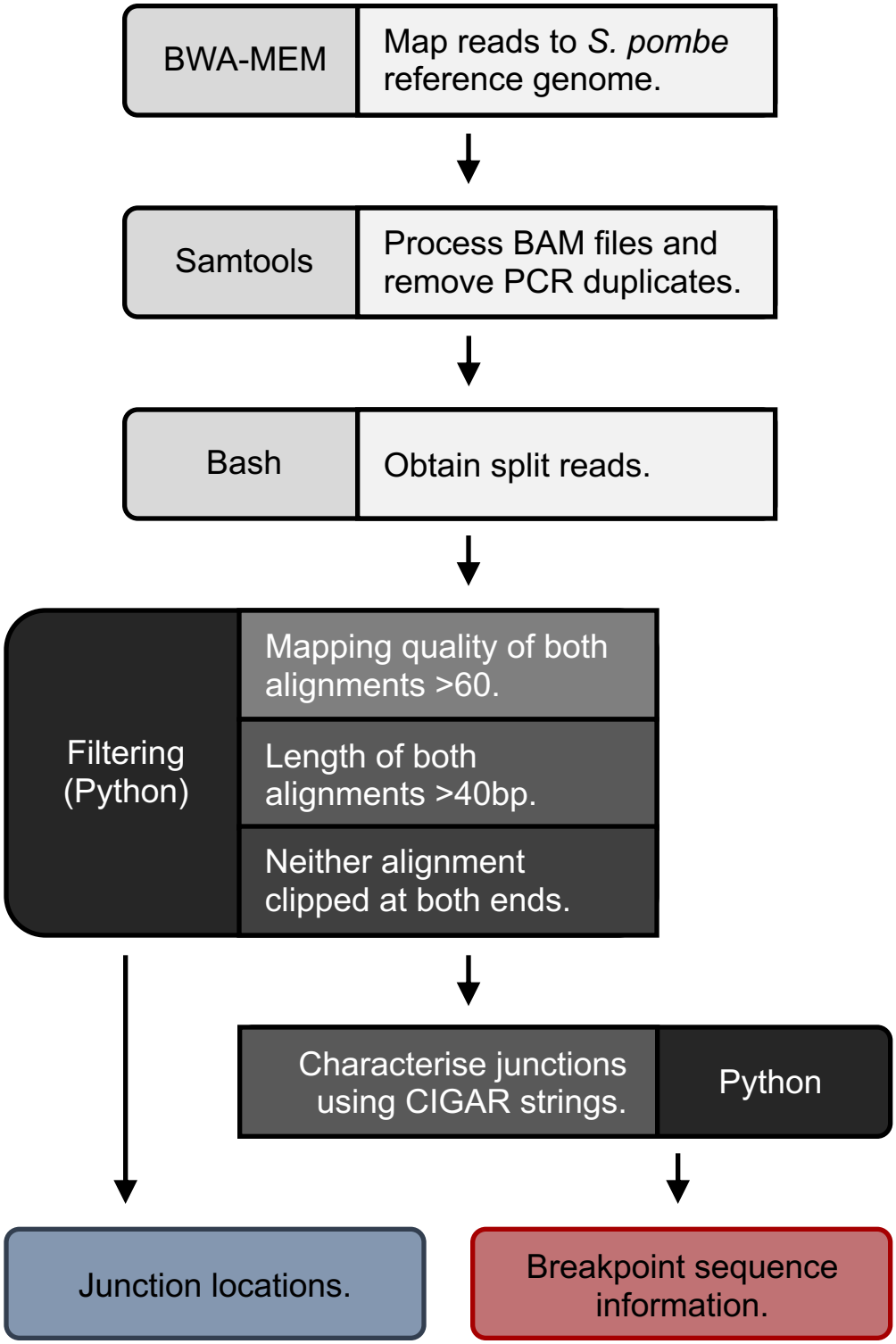

BWA-MEM
Map reads to S. pombe reference genome.
Samtools
Process BAM files and remove PCR duplicates.
Bash
Obtain split reads.
Filtering (Python)
Mapping quality of both alignments >60.
Length of both alignments >40bp.
Neither alignment clipped at both ends.
Characterise junctions using CIGAR strings.
Python
Junction locations.
Breakpoint sequence information.

Figure S1.

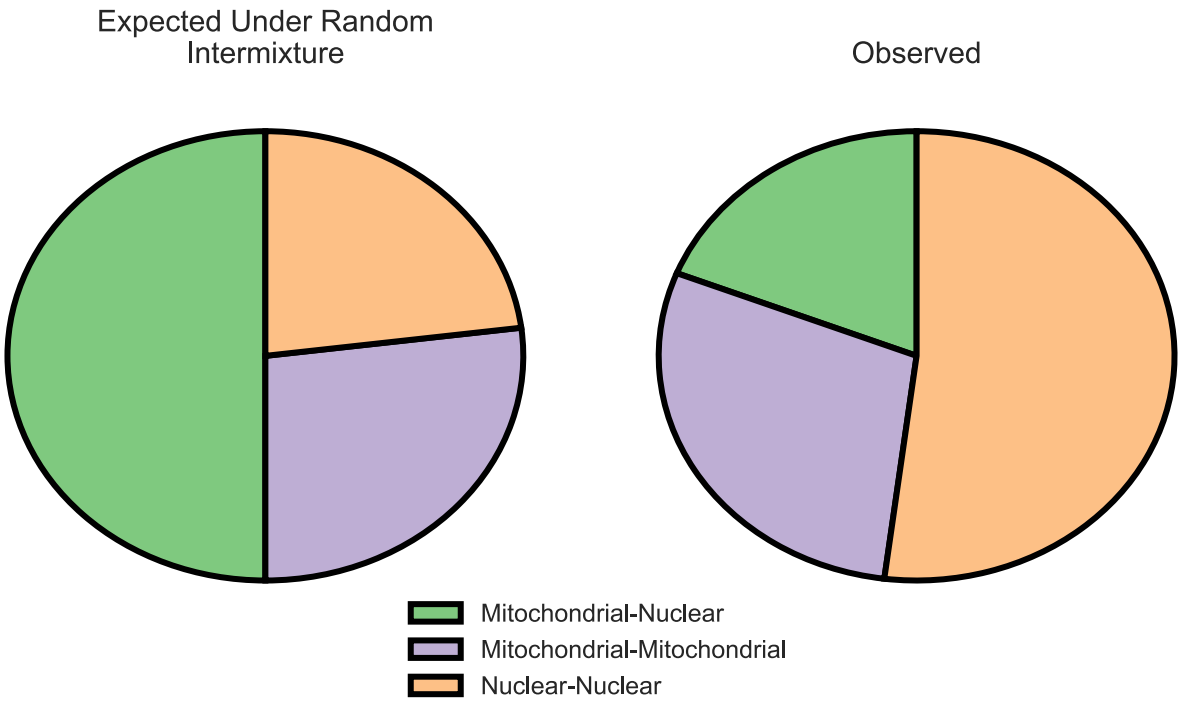

Expected Under Random Intermixture
Observed
Mitochondrial-Nuclear
Mitochondrial-Mitochondrial
Nuclear-Nuclear

Figure S2.

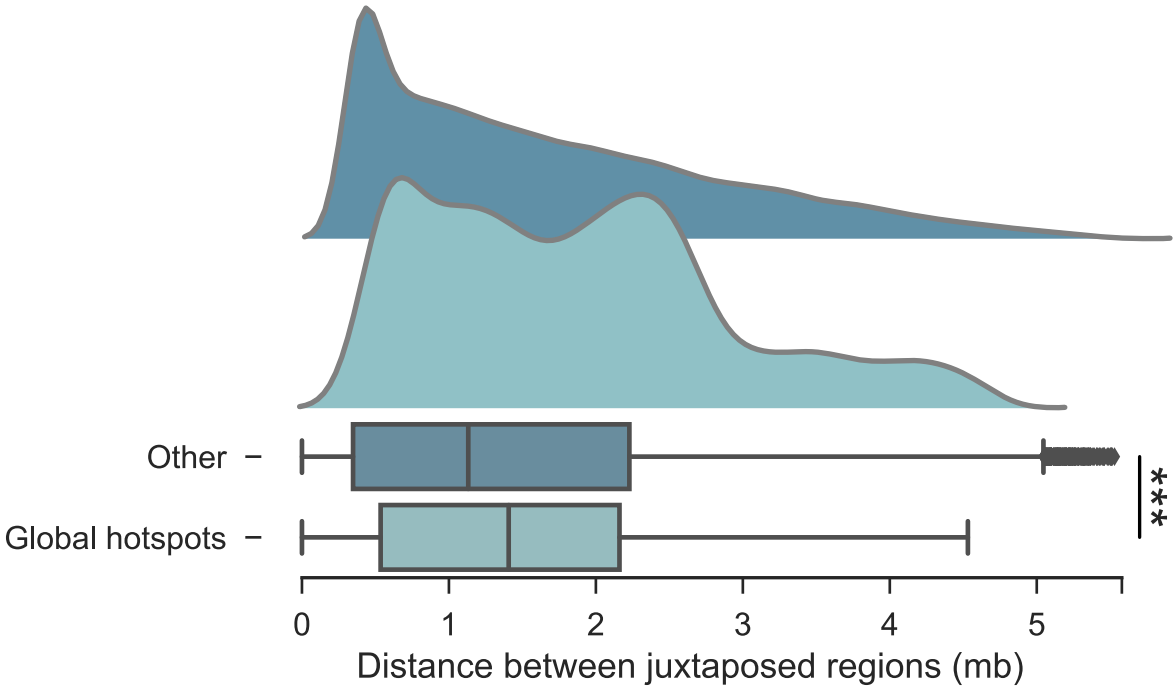
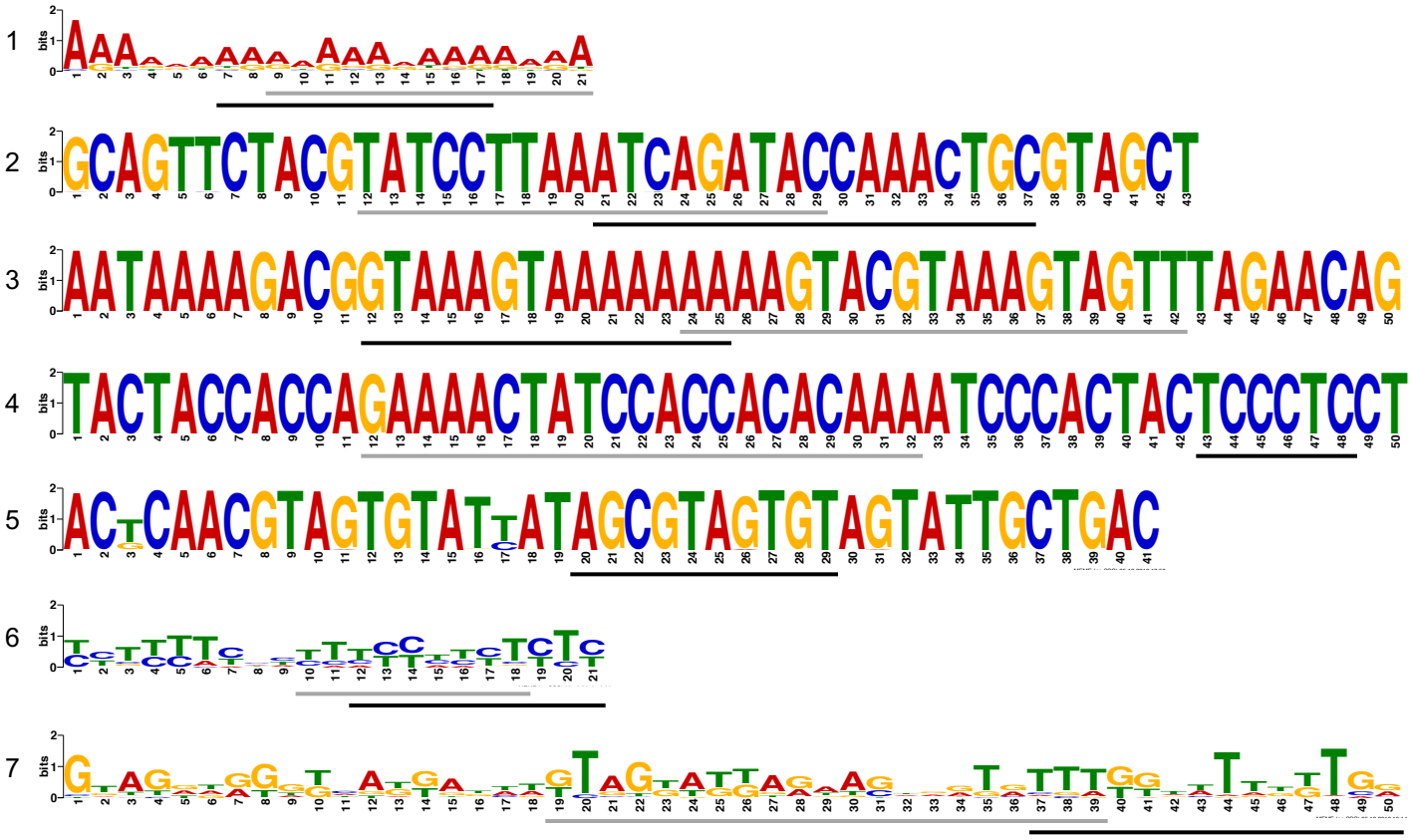

Other
Global hotspots
***
0
1
2
3
4
5
Distance between juxtaposed regions (mb)

Figure S3.

A

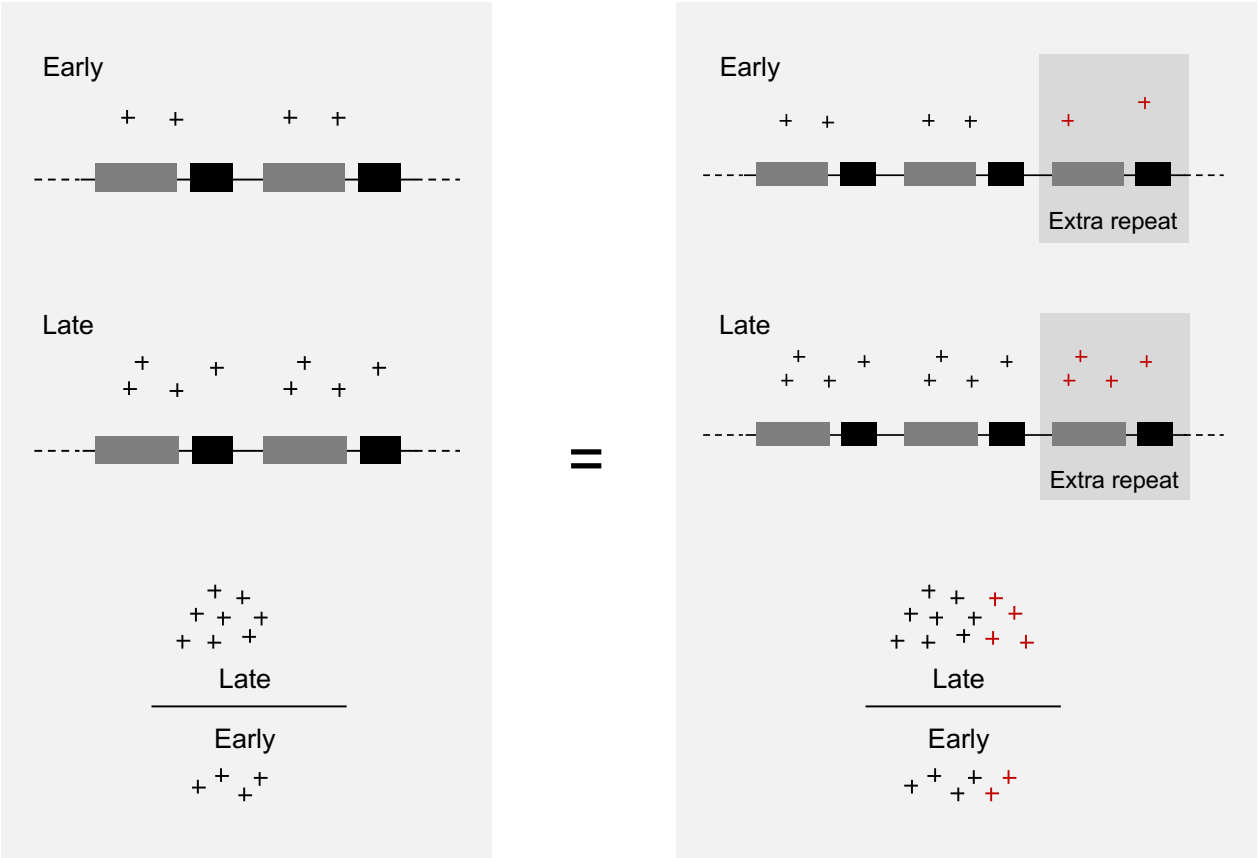


B

ID	factor	description	motif	p-value	e-value
1	ScAzf1	glucose signalling	aAAAAGAAa	<0.0001	<0.05
1	SpSte11	nitrogen starvation	TTCTTTGTTY	<0.01	0.2
2	ScOaf1	telomere silencing	atttCGGatWTAagxxxa	<0.001	0.7
3	ScFhl1	forkhead-like; ribosomal transcription	aaaDxGACGCAaWtxxgtt	<0.01	1.1
3	SpFkh2	forkhead-like; meiotic cell cycle	GTAAACAAACAAAM	<0.01	0.2
4	ScRdr1	multidrug resistance	txxgxcxtGCGGAWAtxxxxc	<0.005	0.9
4	SpRst2	glucose starvation	CCCTC	0.06	2.1
5	SpFhl1	forkhead-like	ACCCTACCCT	<0.01	0.2
6	ScAzf1	glucose signalling	aAAAAGAAa	<0.001	0.6
6	SpSte11	nitrogen starvation	TTCTTTGTTY	<0.01	0.2
7	ScSfl1	stress response; flocculation	tHxHDxataGAAGaaataWDW	<0.05	1.7
7	SpFkh2	forkhead-like; meiotic cell cycle	GTAAACAAACAAAM	<0.05	0.6

Figure S4.

A


Early
Late
Late
Early
=
Early
Extra repeat
Late
Extra repeat
Late
Early

B

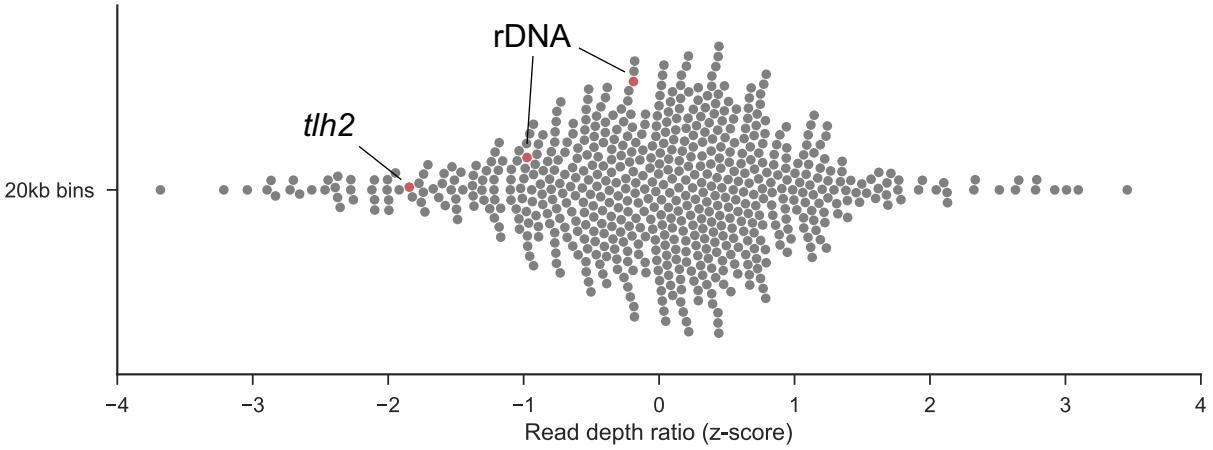

rDNA
tlh2
20kb bins
−4
−3
−2
−1
0
1
2
3
4
Read depth ratio (z-score)

Figure S5.

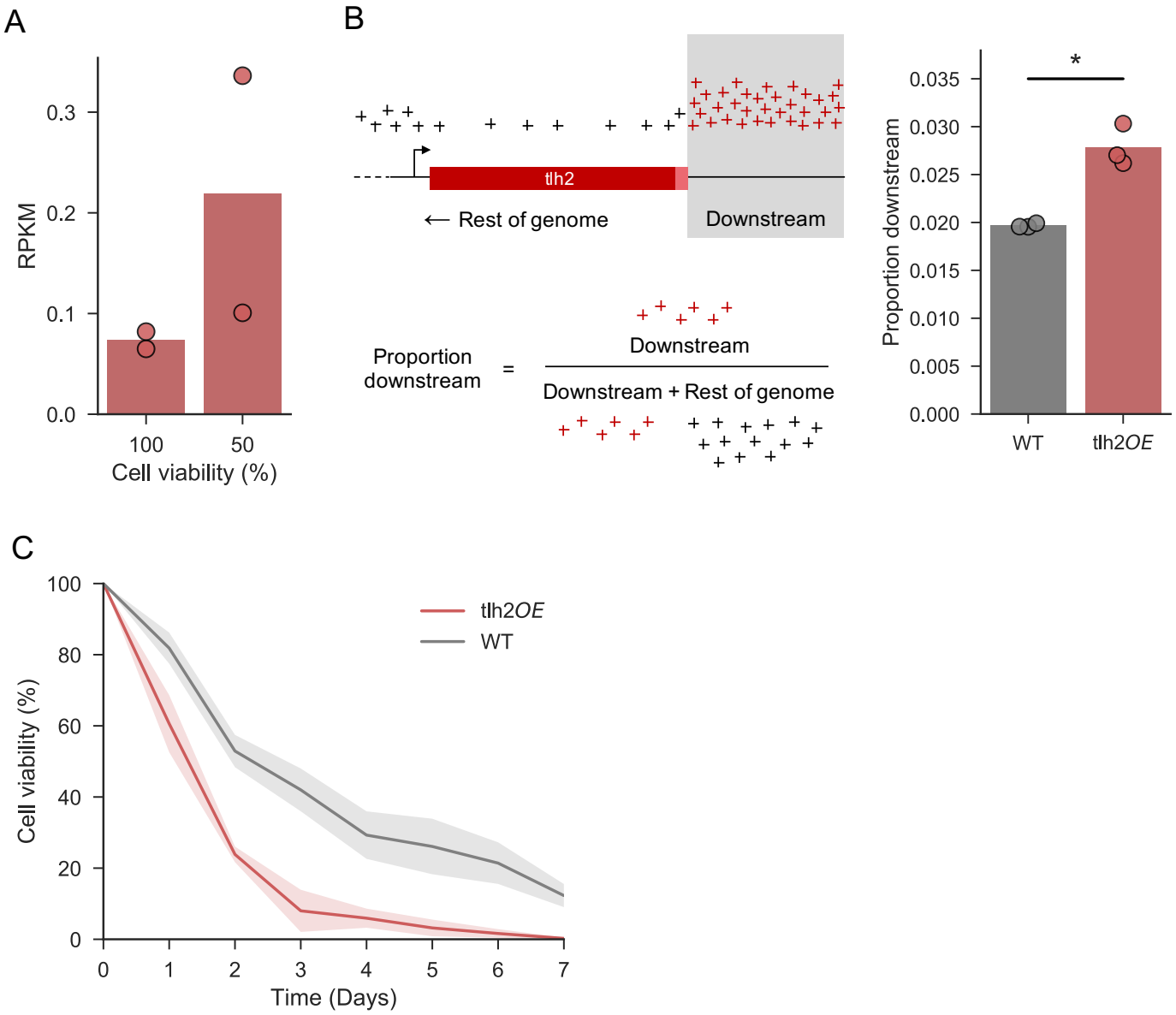

A
RPKM
0.3
0.2
0.1
0.0
100
50
Cell viability (%)
B
tlh2
← Rest of genome
Downstream
Proportion downstream
=
Downstream
Downstream + Rest of genome
Proportion downstream
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
*
WT
tlh2OE
C
Cell viability (%)
100
80
60
40
20
0
0
1
2
3
4
5
6
7
Time (Days)
tlh2OE
WT

Figure S6.

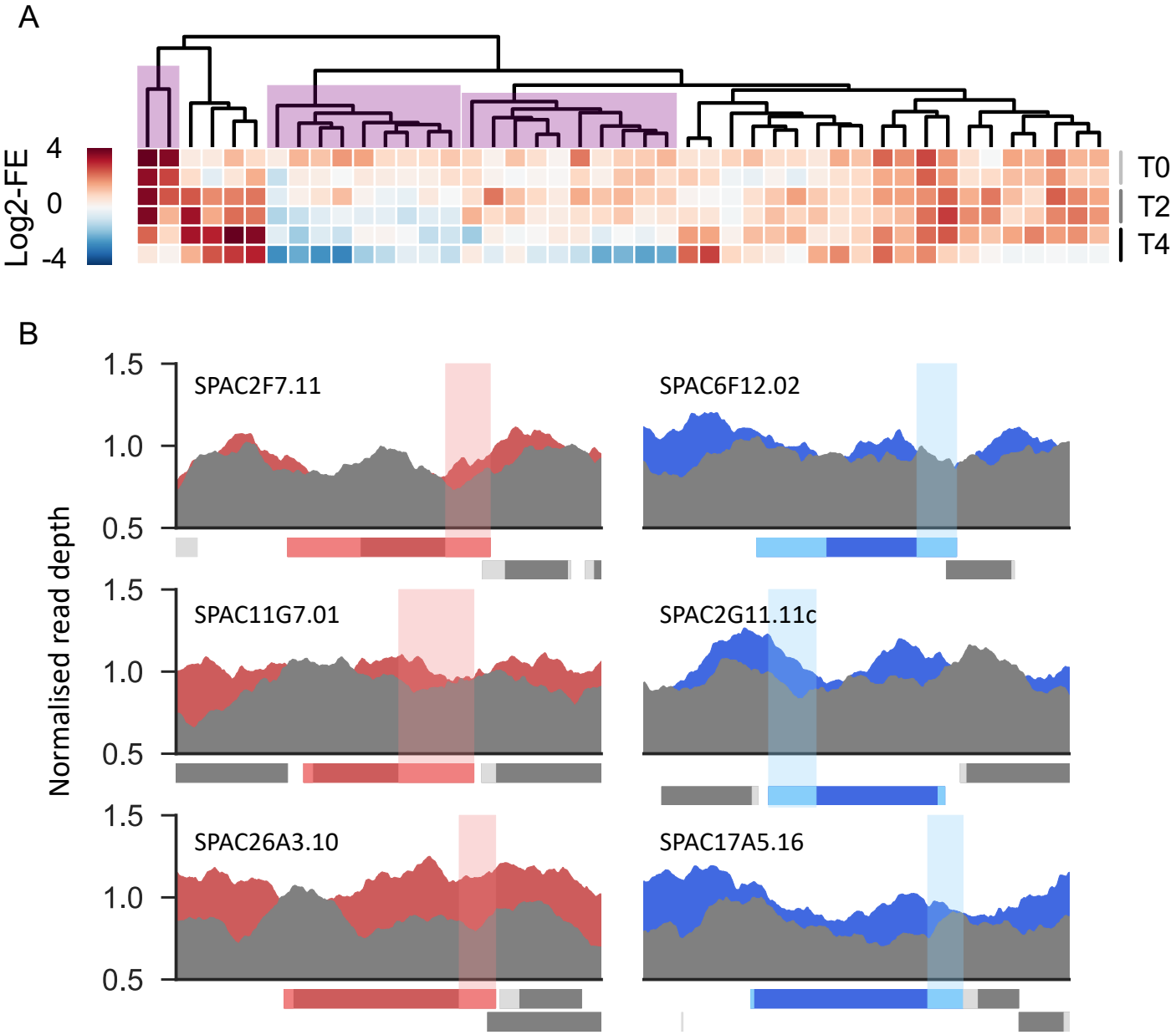
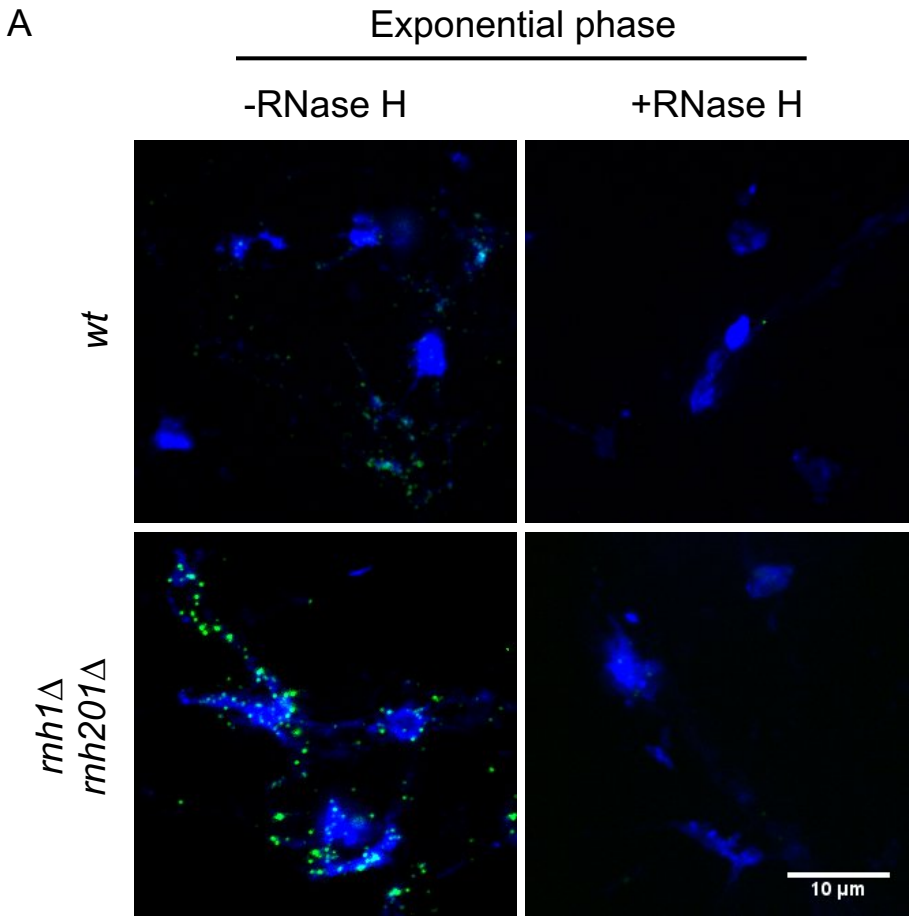

A
Log2-FE
4
0
-4
T0
T2
T4
B
Normalised read depth
1.5
1.0
0.5
SPAC2F7.11
SPAC6F12.02
SPAC11G7.01
SPAC2G11.11c
SPAC26A3.10
SPAC17A5.16

Figure S7.


A
Exponential phase
-RNase H
+RNase H
wt
rnh1Δ
rnh201Δ
10 µm

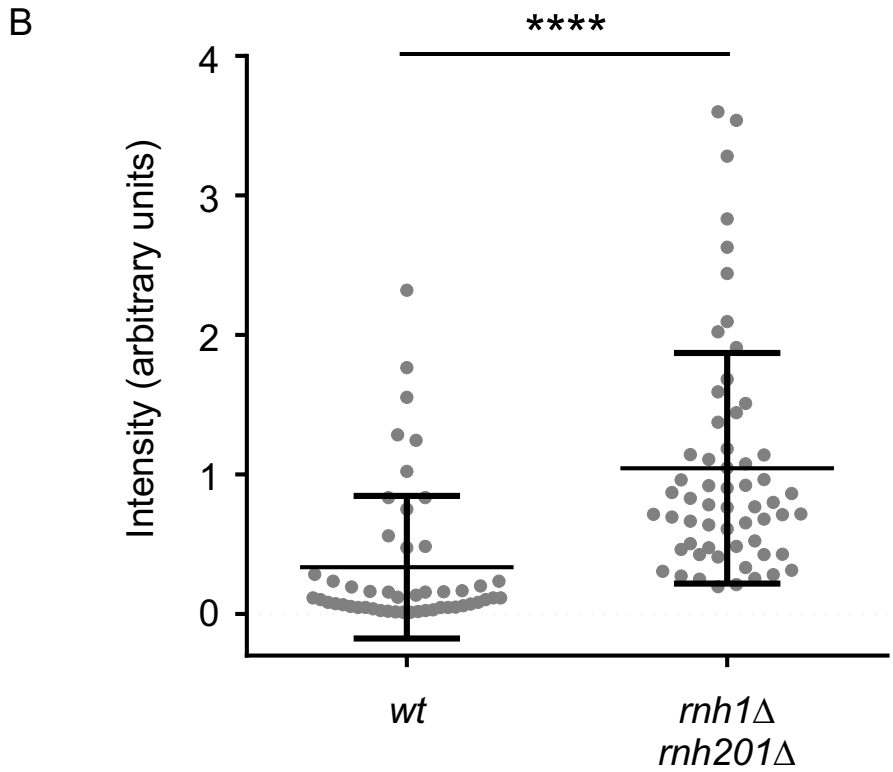

B
****
Intensity (arbitrary units)
4
3
2
1
0
wt
rnh1Δ
rnh201Δ

Figure S8.