Physiological changes during cellular ageing in fission yeast drive non-random patterns of genome rearrangements

David A. Ellis[1], Félix Reyes-Martín[2], María Rodríguez-López[1], Cristina Cotobal Martín[1],

Daniel C. Jeffares[1,3], Víctor A. Tallada[2], Jürg Bähler[1,*]

[1] Department of Genetics, Evolution & Environment and Institute of Healthy Ageing, 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] Current address: Department of Biology, University of York, York YO10 5DD, U.K.

* Corresponding author: Prof. 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

Classification: BIOLOGICAL SCIENCES: Genetics

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 markedly compromise genome structure and function. Rearrangements are challenging to detect in genomes of 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 in ageing cells have focussed on mitochondrial DNA, small genetic variants, or proliferating cells. To gain a better understanding of genome rearrangements during chronological ageing, we reduced their complexity to a single diagnostic measure – the 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 gene transcription. We present an unexpected cause of genomic instability, where age-associated reduction in an RNA-binding protein could trigger R-loop formation at target loci. This example suggests that physiological changes in processes un-related to transcription or replication can drive genome rearrangements. We also identified global hotspots for age-associated breakpoint formation, near telomeric genes and binding sites of genome regulators, linked to remote breakpoints on the same or different chromosomes. Notably, we uncovered similar signatures of genome rearrangements that accumulated in old brain cells of humans. These findings provide fresh 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-regulation proteins.

Significance Statement:

Genome instability and chromosomal rearrangements that join non-neighbouring DNA sequences have been implicated in ageing. We exploit sensitive analyses of deeply sequenced yeast genomes to uncover the rare and diverse events of chromosomal rearrangements that specifically accumulate during ageing of non-dividing cells. These rearrangements are non-randomly distributed across the genome, feature short homologous sequences near the breakpoints, and can involve interactions between different chromosomes or even between mitochondrial and nuclear DNA. Ageing-associated changes in regulatory proteins, leading to increased gene transcription or DNA-RNA interactions, can drive the non-random patterns of chromosomal rearrangements. Similar patterns of chromosomal rearrangements accumulate in non-dividing brain cells in old humans, suggesting that the mechanisms for ageing-associated rearrangements are widely conserved.

Introduction

Cellular processes like transcription and replication can trigger DNA lesions such as double-strand breaks (DSBs) (1–4). DSBs are normally repaired by homologous recombination or by non-homologous end-joining – two pathways which protect chromosomes from aberrant structural variations (5–7). A sensitive sequencing approach has revealed that DSBs occur at hotspots in mouse brain cells, in transcribed genes with neuronal functions (8). 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 (9–11). Microhomology-mediated end-joining events can link chromosomal breakpoints that are normally far apart or even on different chromosomes (12, 13). Such events lead to genome rearrangements such as inversions, duplications, translocations or deletions, which may considerably affect the function of genomes. Thus, the

3

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 (14, 15) and a decline in the efficiency and accuracy of DNA repair (15, 16). Accordingly, increased genomic instability and chromosomal rearrangements are well-known hallmarks of ageing (17–22). Impaired non-homologous end-joining in human patients and mouse models leads to accelerated ageing, and microhomology-mediated end-joining increase with age (23). However, genome re-sequencing studies during ageing have been limited to mitochondrial DNA (24, 25), small genetic variants (26), or proliferating cells (27). No systematic approaches have been applied to identify heterogeneous, rare chromosomal rearrangements in non-dividing, somatic cells (15, 28–30).

Processes affecting ageing are remarkably conserved from yeast to human, including both genetic and environmental factors (21, 31). 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* (32–35). Chronological lifespan is defined as the time a cell survives in a quiescent, non-dividing state, and is a useful model of post-mitotic ageing of somatic metazoan cells (21, 31). Here we interrogate aberrant DNA-junction sequences in genomes of non-dividing *S. pombe* cells, revealing distinct signatures and mechanistic clues for ageing-associated chromosomal rearrangements. 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 and analysed changes in the standing genetic variation as a function of age to identify longevity-associated quantitative trait loci (32). Here we report the striking new structural variation that arose in these cellular pools

during ageing. Typically, structural variant calling software requires support from multiple sequence reads (36–45). 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) (32). Such split reads represent potential breakpoint junctions of genome rearrangements, which lead to new sequence combinations (Fig. 1A). 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 more reads 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 within a chromosome, junctions joining neighbouring regions were preferred over those joining more distal regions (Fig. S3; Note S1). 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 our approach is that the breakpoint junction pairs defining a given rearrangement cannot be analysed together, and the nature 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). It is known that DSBs in nuclear DNA can be repaired with mitochondrial DNA (27, 46, 47). 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 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 (48). We conclude that genome-wide rearrangements, represented by breakpoint junctions featuring microhomologies, accumulate with the age of non-dividing cells.

Using motif discovery, we found seven long sequence motifs to be enriched at microhomology-mediated junctions (Fig. S4A). Motifs of known DNA-binding proteins from fission and budding yeast, which are shorter than the longer sequence motifs discovered, showed significant homology within these longer motifs (Fig. S4B). These proteins are involved in nutrient starvation and other stress responses or in cell-cycle control. This result implicates specific transcription factors in triggering DSBs and microhomology-mediated rearrangements.

Similar patterns of genome rearrangements accumulate in old human brain cells. Long-lived brain and other post-mitotic human cells are analogous to non-dividing yeast cells (49). 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 (24). We found a subtle increase in junctions in older brain cells (Fig. 2A), despite the low coverage and sample number in this data set (Materials and Methods). As in fission yeast (Fig. 1D), these junctions were associated with microhomology (Fig. 2B). The microhomology-associated junctions in human brain cells showed a bimodal distribution: a large

population featuring similar microhomology lengths to fission yeast (median ~8bp), and a less abundant population featuring longer microhomology (median ~16bp). Interestingly, recent work with cancer genome sequences reported a transition in the probability of junction formation at around 11bp of microhomology (48); 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. 2C). 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 (50). We conclude that similar patterns of ageing-related DNA rearrangements occur both in 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 the 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). We identified three strong global hotspots featuring numerous connections with other, typically remote sequences throughout the genome. These global hotspots were located at the right end of Chromosome II and at both ends of

Chromosome III, the latter being the sites of ribosomal DNA (rDNA) (Fig. 3B). Given that these global hotspots occur near chromosome ends, in repetitive regions (51–53), 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 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 number at rDNA repeats decreases with age, and instability in this region is linked with ageing and longevity (54–56). The decrease of rDNA copy number with age could actually be linked to junction formation in these regions. Work in primate kidney cells has demonstrated differential repair of cellular DNA sequences (57), reminiscent to the heterogeneity we observe for age-associated rearrangements. We conclude that ageing-associated rearrangements occur preferentially at either local or global hotspots.

Global hotspots are associated with transcriptional activity and DNA-binding proteins. 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 (53). 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* (58), RecQ helicases are highly conserved. Notably, mutations in two of five human RecQ helicases, BLM and WRN, lead to premature ageing syndromes (59) and another, RECQL5, alleviates transcriptional stress (60). 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 (1, 3).

Moreover, rDNA repeats can become unstable during ageing and cause cell death (54–56). RNA-seq data from ageing cells (61) showed that *tlh2* is lowly expressed in early stationary phase, but becomes more highly expressed during chronological ageing (Fig. 4B). This result is consistent with *tlh2* becoming a site of transcriptional stress in old cells. 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 markedly higher in the *tlh2* overexpression strain compared to wild-type (Fig. 4C). Thus, overexpression of *tlh2* is sufficient to trigger increased breakpoint junctions, likely reflecting rearrangements owing to transcriptional stress. Moreover, the *tlh2* overexpression strain was substantially shorter-lived than wild-type cells (Fig. 4D). These results suggest that increased transcription of *tlh2,* and associated rearrangements, affect cell survival and longevity. Alternatively, or in addition, the rearrangements might be caused by increased levels of the Tlh2 protein.

S. pombe Reb1 is a DNA-binding protein that prevents genetic instability by blocking DNA replication and RNA Polymerase I-dependent transcription at rDNA genes (62, 63). Moreover, Reb1, can mediate the interaction of distal pieces of DNA (64, 65). To test for any association of Reb1 with junction hotspots, we performed ChIP-seq with epitope-tagged Reb1 in non-dividing cells. Indeed, Reb1 bound not only to the rDNA hotspots, but also the other global hotspot downstream of *tlh2* (Fig. 4A, asterisks). Furthermore, Reb1 binding sites were significantly enriched at the other breakpoint junctions throughout the genome (Fig. 4E; Fig. S6). Although the canonical Reb1 DNA-binding motif (62) was not enriched at junctions, motifs of other factors known to mediate DNA-DNA interactions were enriched, e.g. Fkh2 (66) and Rst2 (67) (Fig. S4). Thus, Reb1 and other proteins mediating DNA interactions are associated with the global hotspots and juxtaposed junctions throughout the genome. This association may reflect these sites' need for protection from genetic instability, but could also point to a direct function of Reb1 and other DNA-binding proteins in bringing together distal genomic regions (Fig. 4F). Contrary to expectations, and the situation at local

9

hotspots, junctions between nearby regions of DNA were under-represented at global hotspots (Fig. S3). This result suggests that distinct mechanisms operate at global and local hotspots. We propose that breakpoint junctions at global hotspots may result from transcription stress and involve the action of Reb1 and/or other DNA-binding proteins that control transcription and/or chromosome organization.

Ageing-related changes in RNA-binding proteins and R-loops may promote genome rearrangements. Given the bias towards 3' gene ends 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. Note that 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 (68) showed that, of the 148 enriched genes, 17 produced transcripts that are targets of Scw1, an RNA-binding protein (RBP) (69–71), which is a significant enrichment (FDR-corrected p <0.0001). Scw1 negatively regulates its target RNAs, possibly by binding to a motif in their 3' UTRs (69). 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 suggest 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 (72), 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 loss of Scw1 function trigger rearrangements? R-loops, involving DNA:RNA

hybrids, 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 (4, 73), and they can also directly interfere with DNA repair (74). The presence of RBPs on nascent transcripts can inhibit formation of R-loops by preventing RNA annealing to the single-stranded DNA template (75–78). To test whether loss of Scw1 promotes R-loop formation, we quantified R-loops in *scw1* deletion (*scw1Δ*) and wild-type cells. 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 *scw1Δ* cells contained more R-loops than wild-type cells (Fig. 5D). We conclude that the absence of Scw1 is sufficient for increased R-loop formation, most likely at its target genes. We propose that diminished Scw1 function in old cells leads to the exposure of 3'-ends of nascent target transcripts that can re-anneal with complementary sequences in the template DNA (Fig. 5E). The resulting R-loops then trigger genome instability and associated rearrangements downstream of Scw1 target genes. Thus, age-associated expression changes 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 results provide evidence for widespread genome rearrangements during ageing of non-dividing cells, both in fission yeast and human. 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). The rearrangements that accumulate with age are characterised by microhomologies and may be triggered by transcriptional events. Global hotspots are associated with Reb1, while some local hotspots may involve ageing-related changes in RBP function and associated R-loop formation. Our finding that the age-associated decline in a RNA regulatory protein can trigger breakpoint formation at target genes highlights 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 (32). Briefly, an inter-crossed population fission yeast derived from the parental strains Y0036 and DY8531 (79–81) 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 (69). 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 (82) to introduce an *nmt1* promoter (83) 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 resuspended in EMM without glucose. Reb1-TAP cells were used for the Reb1 ChIP-seq. All transgenic strains were confirmed by PCR.

DNA extraction, library preparation and sequencing. Genomic DNA was extracted using a standard phenol chloroform procedure (84). After precipitation, DNA pellets were resuspended 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 (85), reads were aligned to the *S. pombe* reference genome (accessed May 2015 (52)) using default parameters in BWA-MEM (86, 87) (v0.7.12). Bam files were sorted, and PCR duplicates removed using Samtools (88) (v0.1.19). Split reads were then obtained using a simple shell script (integrating Samtools (88), 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 (89). 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 (24) 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 (24). 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 (90)) and output them as separate vcf files (the file format required by Mason (91) – 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 (92) and Mason (91)). 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

13

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 time points considered simultaneously) was obtained using Samtools (88) (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) (88). 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 (90) to collect the 100bp surrounding sequence. These 100bp sequences were compiled into a multi-sequence fasta file and submitted to MEME (93) for motif discovery. Only motifs with e-values lower

than 0.05 were considered. Using TomTom (93), 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. S7B.

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 (94) (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 or Reb1 binding sites, bed files for the locations of each feature were obtained. Reb1 binding sites were acquired from a single reb1 ChIP-seq repeat of a late stationary phase culture grown in EMM; the coordinates of PomBase-annotated (95, 96) 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 (94) (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 (68).

Chromosomal spreads. R-loop immunostaining and fluorescence quantification was performed as

described (97), 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 (98).

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 (69),

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 (34). 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 time points 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 (99). For experiments using

Scw1-TAP, ChIP-seq samples were collected at the same time points as for the RIP-chip experiments

(see above). Reb1-TAP and wild-type cells were grown in EMM and fixed upon entry into early

stationary phase (OD$_{600}$ ~8). A single repeat was performed for this experiment. ChIP-seq libraries

17

were sequenced using an Illumina MiSeq instrument. For Reb1 binding site localization, Bowtie2 (Langmead, B. 2012) with default settings was used to map the reads to the *S. pombe* genome build ASM294v2, downloaded from Pombase (http://www.pombase.org/) (95, 96). SAMtools (86) and BED tools (94) were used for sequence manipulation. Peak calling was performed using GEM (100). The ChIP-seq data are available in the European Nucleotide Archive under the study accession PRJEB30570.

Acknowledgements

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 (52)).

(C) Number of breakpoint Junctions Per Mapped Read (JPMR) passing filter (Materials and Methods) as a function of cellular age. 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 time points in blue through to late time points 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.

(A) Number of breakpoint Junctions Per Mapped Read (JPMR) passing filtering (Materials and Methods) as a function of donor age. Samples were classed as young (blue) and old (red) for analysis in B.

19

(B) 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).

(C) 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 (101) 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. Global hotspots are associated with transcriptional activity and Reb1-bound DNA.

(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. Asterisks mark centre of peaks called from Reb1 binding data.

(B) 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).

(C) 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).

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

(E) Proportion of junctions expected next to Reb1 binding sites (based on simulated data) *vs* proportion that are observed (peaks called by ChIP-seq; Materials and Methods; Fig. S6). One-sided two sample Mann-Whitney U to test whether observed is greater than expected (U=64, p <0.001).

(F) Model for how global hotspots might trigger rearrangements. The *tlh2* gene is lowly transcribed in young cells, while in old cells (or a subset of old cells) *tlh2* becomes more highly transcribed, leading to collisions between RNA Polymerase II and distal DNA that has been brought to the region by Reb1.

21

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 αS9.6 antibody(102) (yellow), with (+) or without (-) RNase H treatment (which removes R-loops).

Scale bar: 10µm.

(E) 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 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. Genome-wide distribution of all Reb1 binding sites. Rug plot showing the location of the 41 strongest Reb1 binding sites called from ChIP-seq data of early stationary phase cells. Chromosomal location is shown on the x axis. This set of binding sites was used for Fig. 4A & 4E.

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 (published in reference (69)) 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.

24

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 (48).

25

References

1. Gaillard H, Aguilera A (2016) Transcription as a Threat to Genome Integrity. *Annu Rev Biochem* 85(1):291–317.

2. Heyer WD (2015) Regulation of recombination and genomic maintenance. *Cold Spring Harb Perspect Med* 5(8):1–24.

3. Aguilera A, García-Muse T (2013) Causes of Genome Instability. *Annu Rev Genet*. doi:10.1146/annurev-genet-111212-133232.

4. 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(13):1384–1396.

5. Soutoglou E, et al. (2007) Positional stability of single double-strand breaks in mammalian cells. *Nat Cell Biol* 9(6):675–682.

6. Difilippantonio MJ, et al. (2000) DNA repair protein Ku80 suppresses chromosomal aberrations and malignant transformation. *Nature* 404(6777):510–514.

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

8. Wei P-C, et al. (2016) Long Neural Genes Harbor Recurrent DNA Break Clusters in Neural Stem/Progenitor Cells. *Cell* 164(4):644–655.

9. Ottaviani D, LeCain M, Sheer D (2014) The role of microhomology in genomic structural variation. *Trends Genet* 30(3):85–94.

10. Decottignies A (2007) Microhomology-Mediated End Joining in Fission Yeast Is Repressed by Pku70 and Relies on Genes Involved in Homologous Recombination. *Genetics* 176(3):1403–1415.

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

12. Bétermier M, Bertrand P, Lopez BS, Hendrickson E, Ramsden D (2014) Is Non-Homologous End-Joining Really an Inherently Error-Prone Process? *PLoS Genet* 10(1):e1004086.

13. Villarreal DD, et al. (2012) Microhomology Directs Diverse DNA Break Repair Pathways and Chromosomal Translocations. *PLoS Genet* 8(11):e1003026.

14. Lombard DB, et al. (2005) DNA Repair, Genome Stability, and Aging. *Cell* 120(4):497–512.

15. White RR, Vijg J (2016) Do DNA Double-Strand Breaks Drive Aging? *Mol Cell* 63(5):729–738.

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

17. Szilard L (1959) On the nature of the aging process. *Proc Natl Acad Sci U S A* 45(1):30–45.

18. Curtis HJ (1963) Biological Mechanisms Underlying the Aging Process. *Science (80-)* 141(3582).

Available at: http://science.sciencemag.org/content/141/3582/686 [Accessed June 14, 2018].

19. Strehler BL (1986) Genetic instability as the primary cause of human aging. *Exp Gerontol* 21(4–5):283–319.

20. Hasty P, Campisi J, Hoeijmakers J, van Steeg H, Vijg J (2003) Aging and Genome Maintenance: Lessons from the Mouse? *Science (80-)* 299(5611):1355–1359.

21. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G (2013) The Hallmarks of Aging. *Cell* 153(6):1194–1217.

22. Vijg J, Suh Y (2013) Genome Instability and Aging. *Annu Rev Physiol* 75(1):645–668.

23. Vaidya A, et al. (2014) Knock-in reporter mice demonstrate that DNA repair by non-homologous end joining declines with age. *PLoS Genet* 10(7):e1004511.

24. Williams SL, Mash DC, Züchner S, Moraes CT (2013) Somatic mtDNA mutation spectra in the aging human putamen. *PLoS Genet* 9(12):e1003990.

25. 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. *PLoS Genet* 9(9):e1003794.

26. Gangloff S, et al. (2017) Quiescence unveils a novel mutational force in fission yeast. *Elife* 6. doi:10.7554/eLife.27469.

27. Hu Z, et al. (2014) Nucleosome loss leads to global transcriptional up-regulation and genomic instability during yeast aging. *Genes Dev* 28(4):396–408.

28. Quispe-Tintaya W, et al. (2016) Quantitative detection of low-abundance somatic structural variants in normal cells by high-throughput sequencing. *Nat Methods* 13(7):584–6.

29. Laurie CC, et al. (2012) Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nat Genet* 44(6):642–650.

30. Jacobs KB, et al. (2012) Detectable clonal mosaicism and its relationship to aging and cancer. *Nat Genet* 44(6):651–658.

31. Kaeberlein M (2010) Lessons on longevity from budding yeast. *Nature* 464(7288):513–519.

32. Ellis DA, et al. (2018) Uncovering Natural Longevity Alleles from Intercrossed Pools of Aging Fission Yeast Cells. *Genetics*:genetics.301262.2018.

33. Roux AE, et al. (2009) Pro-aging effects of glucose signaling through a G protein-coupled glucose receptor in fission yeast. *PLoS Genet* 5(3):e1000408.

34. 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(4):563–73.

35. 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(2). Available at: http://bio.biologists.org/content/3/2/161 [Accessed November 21, 2017].

36. 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(21):2865–2871.

37. Wang J, et al. (2011) CREST maps somatic structural variation in cancer genomes with base-pair resolution. *Nat Methods* 8(8):652–654.

38. Emde A-K, et al. (2012) Detecting genomic indel variants with exact breakpoints in single- and paired-end sequencing data using SplazerS. *Bioinformatics* 28(5):619–27.

39. 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(3):R22.

40. Rausch T, et al. (2012) DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* 28(18):i333–i339.

41. 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(20):2576–2583.

42. Trappe K, Emde A-K, Ehrlich H-C, Reinert K (2014) Gustaf: Detecting and correctly classifying SVs in the NGS twilight zone. *Bioinformatics* 30(24):3484–3490.

43. Layer RM, Chiang C, Quinlan AR, Hall IM (2014) LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol* 15(6):R84.

44. Kimura K, Koike A (2015) Analysis of genomic rearrangements by using the Burrows-Wheeler transform of short-read data. *BMC Bioinformatics* 16(Suppl 18):S5.

45. Jeffares DC, et al. (2017) Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nat Commun* 8:14061.

46. Ricchetti M, Fairhead C, Dujon B (1999) Mitochondrial DNA repairs double-strand breaks in yeast chromosomes. *Nature* 402(6757):96–100.

47. Decottignies A (2005) Capture of extranuclear DNA at fission yeast double-strand breaks. *Genetics* 171(4):1535–48.

48. Wala JA, et al. (2017) Selective and mechanistic sources of recurrent rearrangements across the cancer genome. *bioRxiv*:187609.

49. Magrassi L, Leto K, Rossi F (2013) Lifespan of neurons is uncoupled from organismal lifespan. *Proc Natl Acad Sci U S A* 110(11):4374–9.

50. Merino MM, et al. (2015) Elimination of Unfit Cells Maintains Tissue Health and Prolongs

28

Lifespan. *Cell* 160(3):461–476.

51. Toda T, Nakaseko Y, Niwa O, Yanagida M (1984) Mapping of rRNA genes by integration of hybrid plasmids in Schizosaccharomyces pombe. *Curr Genet* 8(2):93–97.

52. Wood V, et al. (2002) The genome sequence of Schizosaccharomyces pombe. *Nature* 415(6874):871–880.

53. Mandell JG, Goodrich KJ, Bähler ¶ J, Cech TR (2004) Expression of a RecQ Helicase Homolog Affects Progression through Crisis in Fission Yeast Lacking Telomerase*. *J Biol Chem*. doi:10.1074/jbc.M412756200.

54. Spivey EC, Jones SK, Rybarski JR, Saifuddin FA, Finkelstein IJ (2017) An aging-independent replicative lifespan in a symmetrically dividing eukaryote. *Elife* 6. doi:10.7554/eLife.20340.

55. Kobayashi T (2014) Ribosomal RNA gene repeats, their stability and cellular senescence. *Proc Jpn Acad Ser B Phys Biol Sci* 90(4):119–29.

56. 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. doi:10.7554/eLife.32421.

57. Zolan ME, Cortopassi GA, Smith CA, Hanawalt PC (1982) Deficient repair of chemical adducts in alpha DNA of monkey cells. *Cell* 28(3):613–9.

58. Nakayama H, et al. (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(3):474–80.

59. Bohr VA (2008) Rising from the RecQ-age: the role of human RecQ helicases in genome maintenance. *Trends Biochem Sci* 33(12):609–20.

60. Saponaro M, et al. (2014) RECQL5 controls transcript elongation and suppresses genome instability associated with transcription stress. *Cell* 157(5):1037–49.

61. Atkinson SR, et al. (2018) Long noncoding RNA repertoire and targeting by nuclear exosome, cytoplasmic exonuclease, and RNAi in fission yeast. *RNA* 24(9):1195–1213.

62. Jaiswal R, et al. (2016) Functional architecture of the Reb1-Ter complex of Schizosaccharomyces pombe. *Proc Natl Acad Sci U S A*. doi:10.1073/pnas.1525465113.

63. Zhao A, Guo A, Liu Z, Pape L (1997) Molecular cloning and analysis of Schizosaccharomyces pombe Reb1p: sequence-specific recognition of two sites in the far upstream rDNA intergenic spacer. *Nucleic Acids Res* 25(4):904–10.

64. Singh SK, Sabatinos S, Forsburg S, Bastia D (2010) Regulation of Replication Termination by Reb1 Protein-Mediated Action at a Distance. *Cell*. doi:10.1016/j.cell.2010.08.013.

65. Jakociunas T, et al. (2013) Subnuclear relocalization and silencing of a chromosomal region by

an ectopic ribosomal DNA repeat. *Proc Natl Acad Sci U S A* 110(47):E4465-73.

66. Ostrow AZ, et al. (2017) Conserved forkhead dimerization motif controls DNA replication timing and spatial organization of chromosomes in S. cerevisiae. *Proc Natl Acad Sci U S A* 114(12):E2411–E2419.

67. Asada R, et al. (2017) Recruitment and delivery of the fission yeast Rst2 transcription factor via a local genome structure counteracts repression by Tup1-family corepressors. *Nucleic Acids Res* 45(16):9361–9371.

68. Bitton DA, et al. (2015) AnGeLi: A Tool for the Analysis of Gene Lists from Fission Yeast. *Front Genet* 6:330.

69. 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. *PLoS Genet* 10(11):e1004684.

70. 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(3):510–20.

71. Karagiannis J, Oulton R, Young PG (2002) The Scw1 RNA-binding domain protein regulates septation and cell-wall structure in fission yeast. *Genetics* 162(1):45–58.

72. 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(6330). Available at: http://science.sciencemag.org/content/355/6330/1184/tab-pdf [Accessed July 23, 2018].

73. Santos-Pereira JM, Aguilera A (2015) R loops: new modulators of genome dynamics and function. *Nat Rev Genet* 16(10):583–597.

74. Amon JD, Koshland D (2016) RNase H enables efficient repair of R-loop induced DNA damage. *Elife* 5. doi:10.7554/eLife.20533.

75. Huertas P, Aguilera A (2003) Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol Cell* 12(3):711–21.

76. Aguilera A (2002) The connection between transcription and genomic instability. *EMBO J* 21(3):195–201.

77. Li X, Manley JL (2005) Inactivation of the SR Protein Splicing Factor ASF/SF2 Results in Genomic Instability. *Cell* 122(3):365–378.

78. 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(6):978–988.

79. 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(12):3496–510.

80. Brown WRA, et al. (2011) A Geographically Diverse Collection of Schizosaccharomyces pombe Isolates Shows Limited Phenotypic Variation but Extensive Karyotypic Diversity. *G3 (Bethesda)* 1(7):615–26.

81. Jeffares DC, et al. (2015) The genomic and phenotypic diversity of Schizosaccharomyces pombe. *Nat Genet* 47(3):235–241.

82. Bähler J, et al. (1998) Heterologous modules for efficient and versatile PCR-based gene targeting inSchizosaccharomyces pombe. *Yeast* 14(10):943–951.

83. Forsburg SL (1993) Comparison of Schizosaccharomyces pombe expression systems. *Nucleic Acids Res* 21(12):2955–6.

84. Murray JM, Watson AT, Carr AM (2016) Extraction of Chromosomal DNA from Schizosaccharomyces pombe. *Cold Spring Harb Protoc* 2016(5):pdb.prot090985.

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

86. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–1760.

87. Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*. Available at: http://arxiv.org/abs/1303.3997 [Accessed September 16, 2017].

88. Li H, et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.

89. 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(Suppl 14):S7.

90. Cock PJA, et al. (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25(11):1422–1423.

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

92. McKenna A, et al. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20(9):1297–303.

93. Bailey TL, et al. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* 37(Web Server):W202–W208.

94. Quinlan AR, Hall IM, Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841–842.

95. Wood V, et al. (2012) PomBase: a comprehensive online resource for fission yeast. *Nucleic Acids Res* 40(Database issue):D695-9.

96. McDowall MD, et al. (2015) PomBase 2015: updates to the fission yeast database. *Nucleic Acids Res* 43(Database issue):D656-61.

97. Hoyos-Manchado R, et al. (2017) RNA metabolism is the primary target of formamide in vivo. *Sci Rep* 7(1):15895.

98. Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 9(7):671–675.

99. Cotobal C, et al. (2015) Role of Ccr4-Not complex in heterochromatin formation at meiotic genes and subtelomeres in fission yeast. *Epigenetics Chromatin* 8:28.

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

101. Consortium TEP (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489(7414):57–74.

102. Boguslawski SJ, et al. (1986) Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids. *J Immunol Methods* 89(1):123–30.
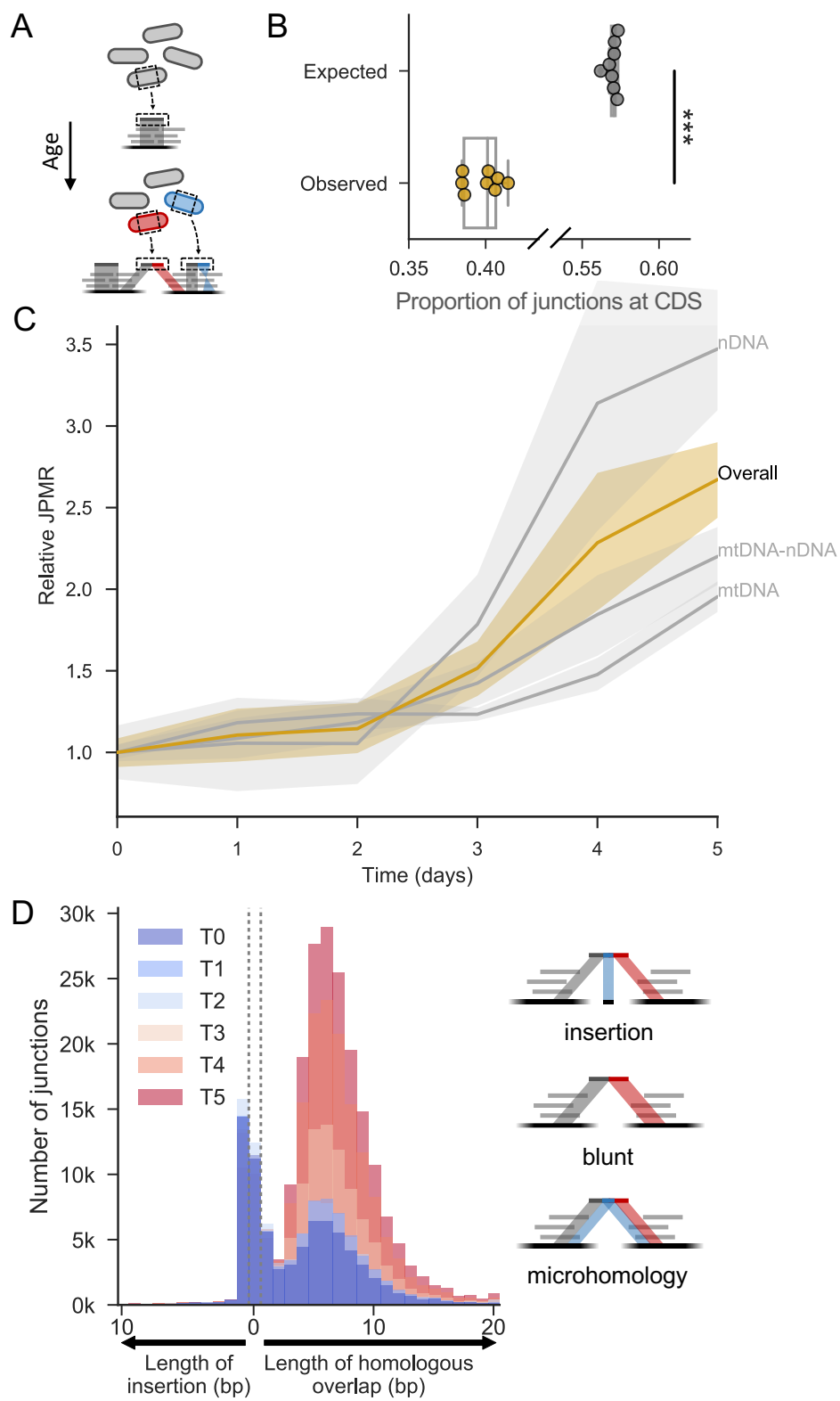
Figure 1.
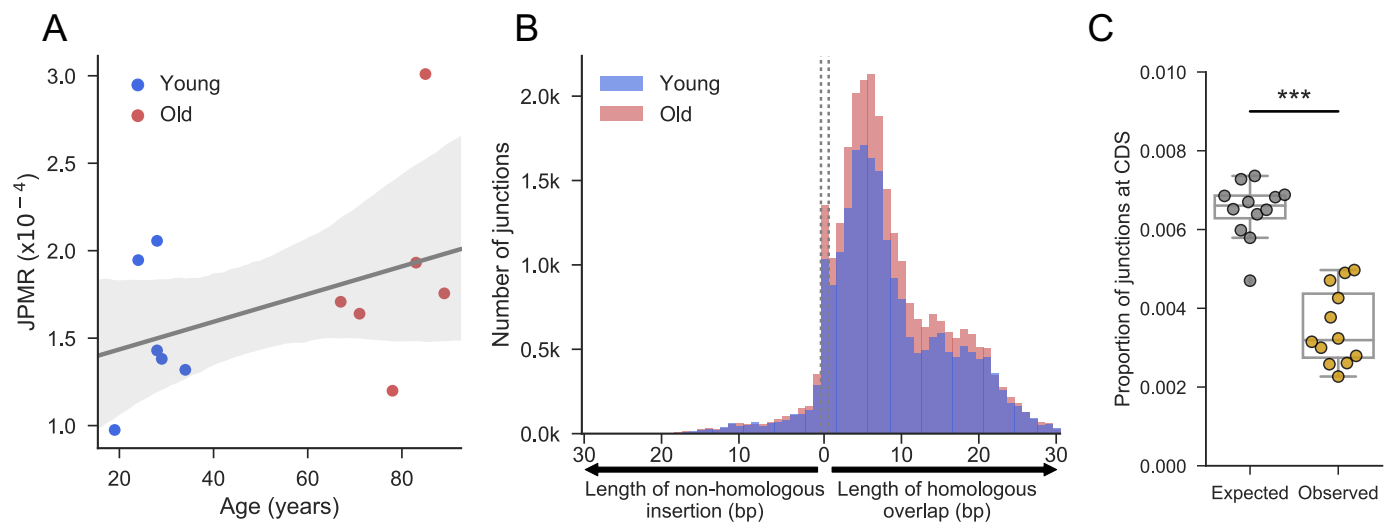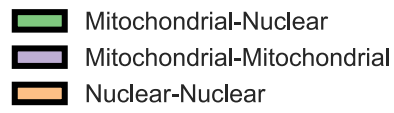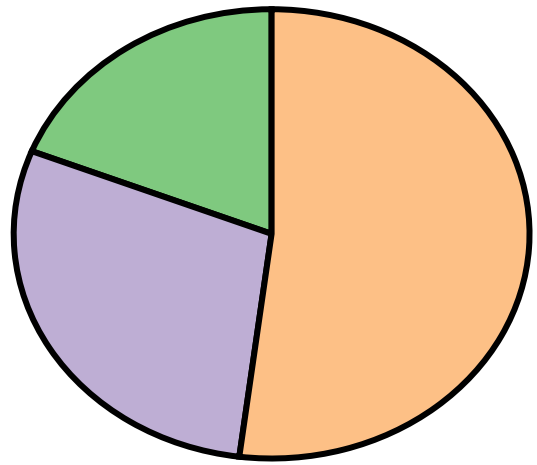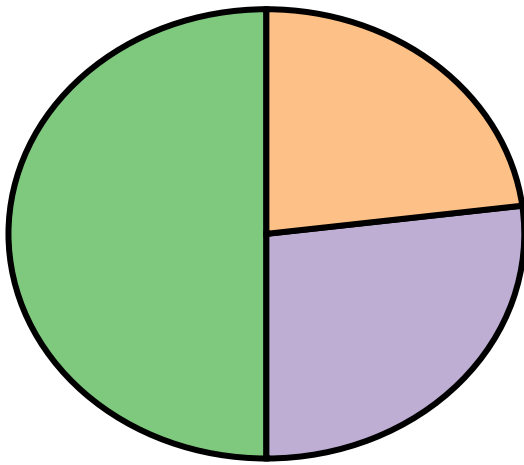
Figure 2.
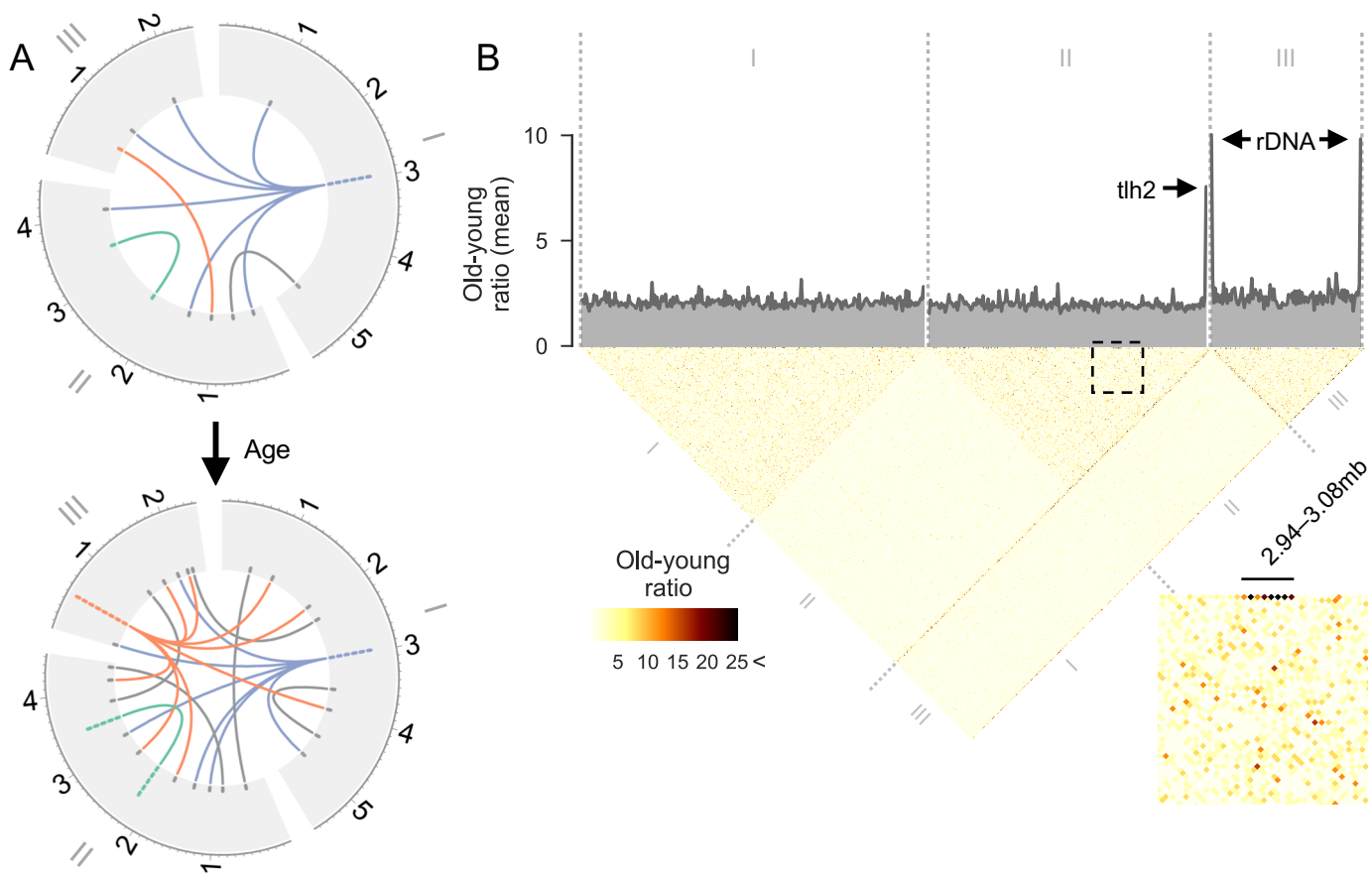
Figure 3.
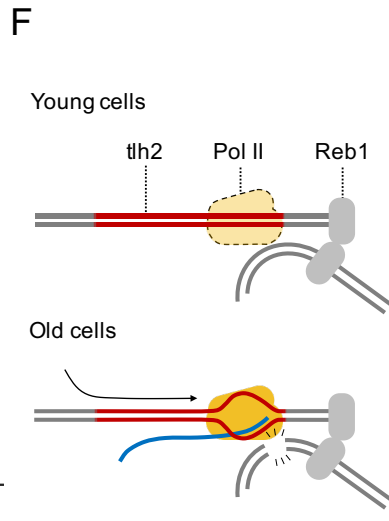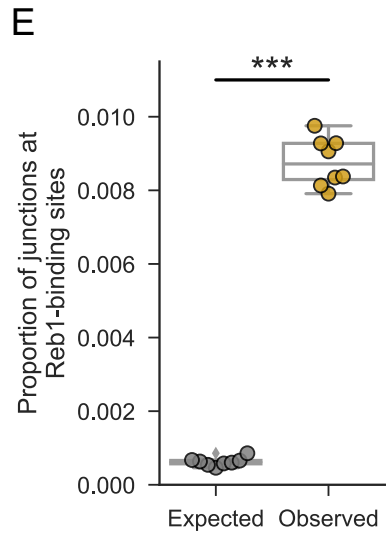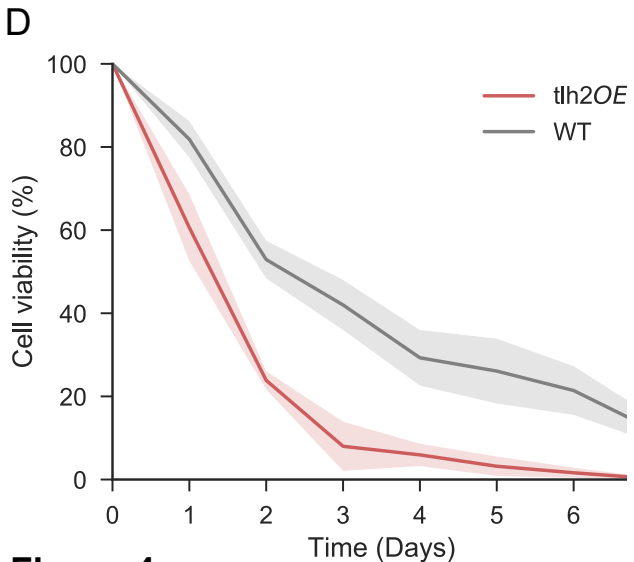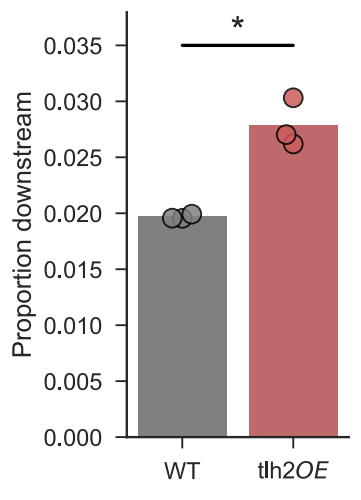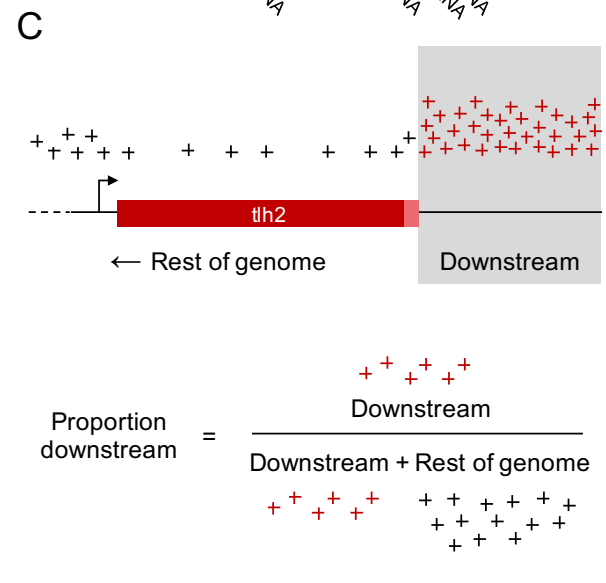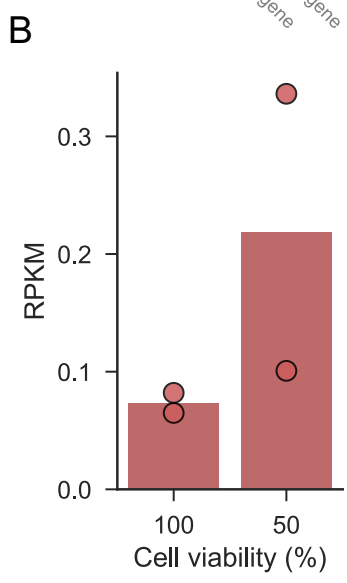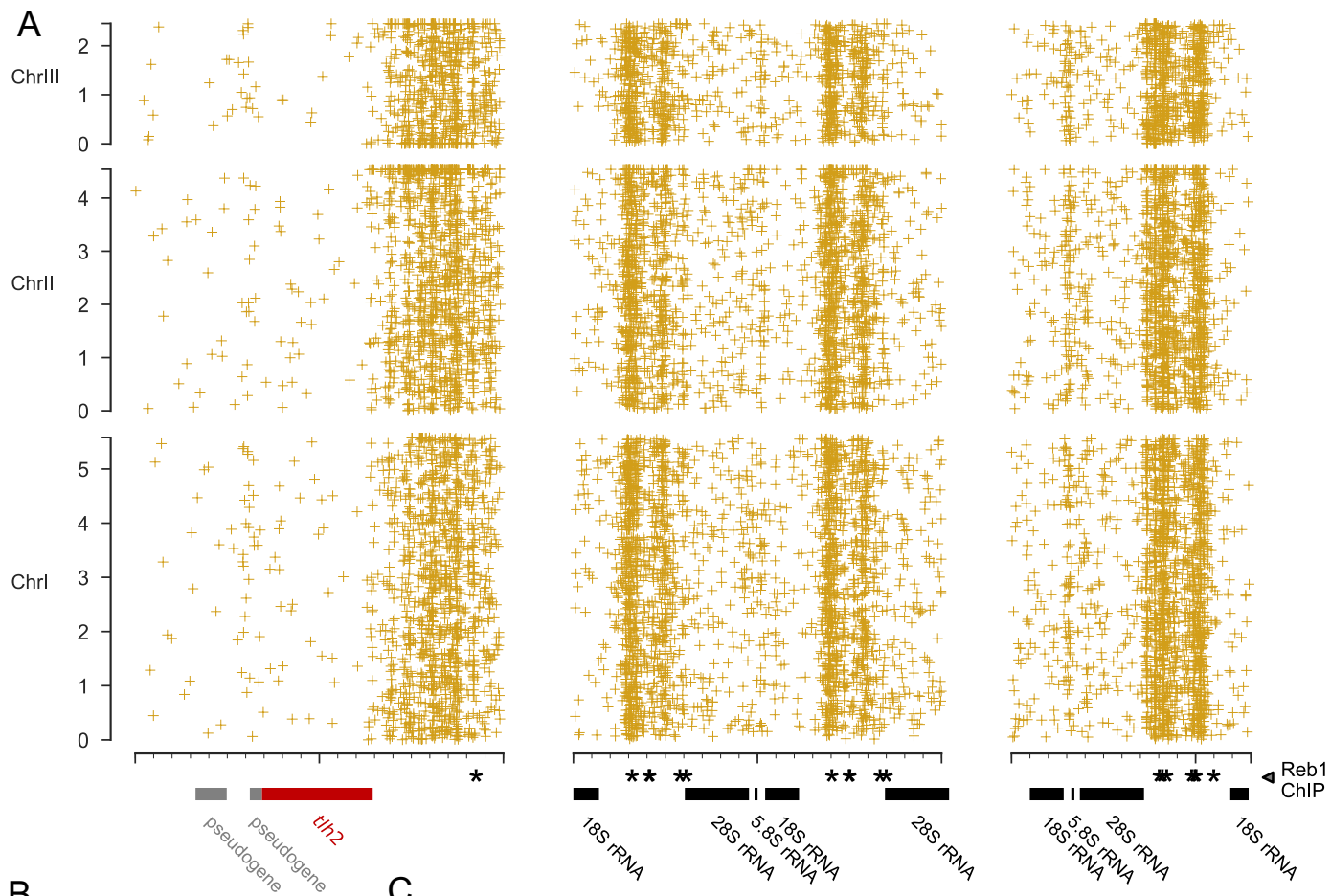
Figure 4.
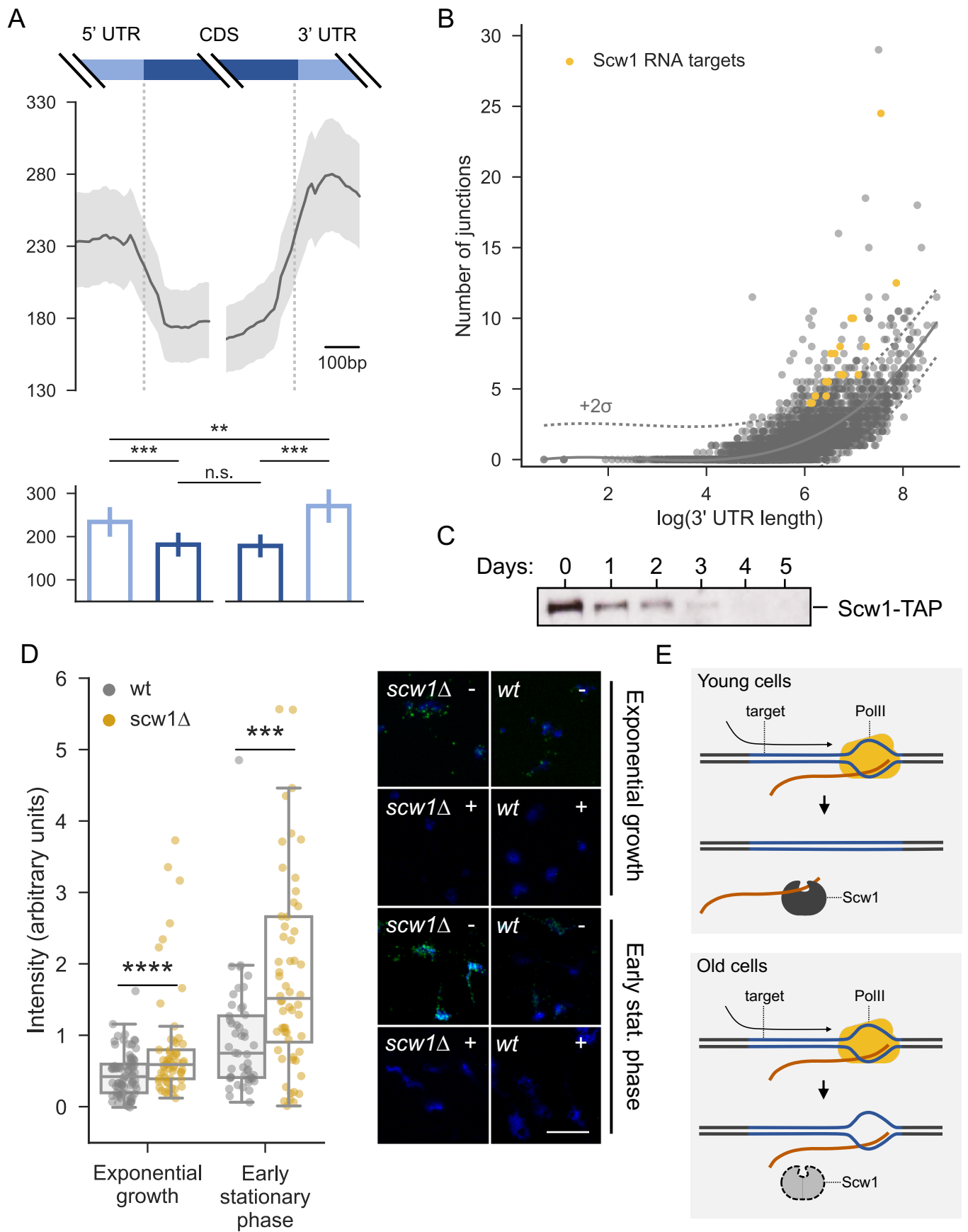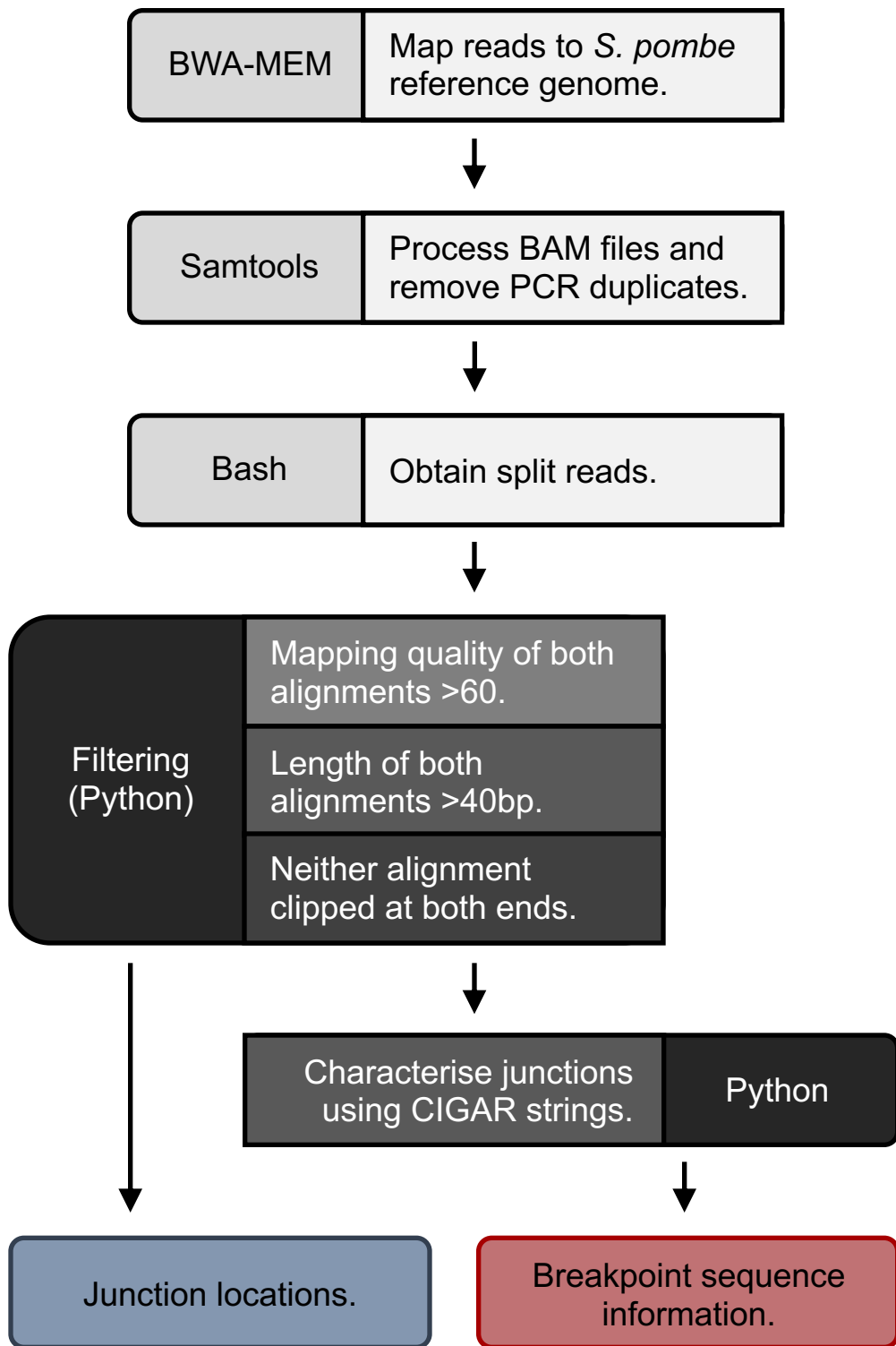
Figure 5.

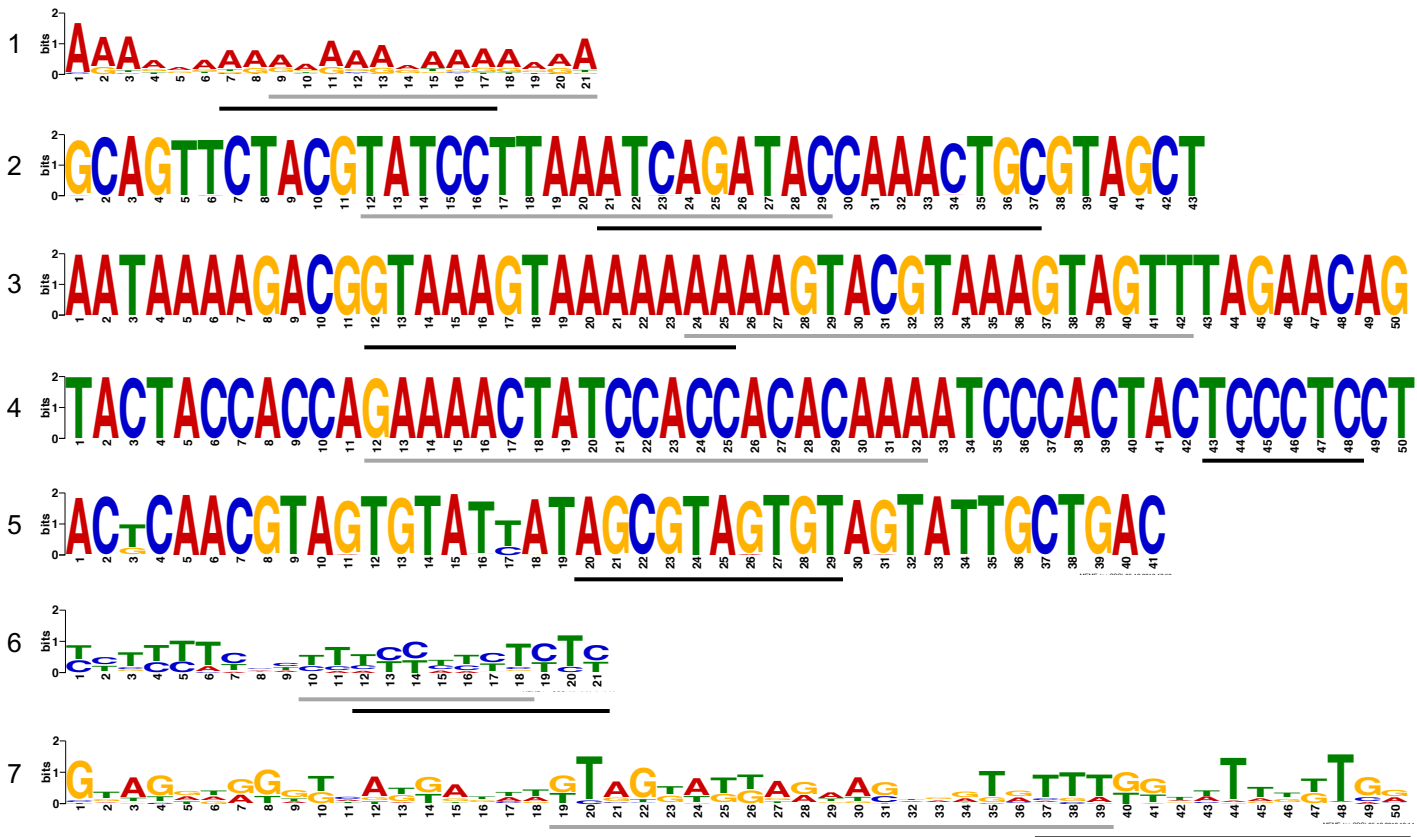**Supplementary Figure 1.**

Expected Under Random Intermixture

Observed

Legend:
- Mitochondrial-Nuclear
- Mitochondrial-Mitochondrial
- Nuclear-Nuclear

Supplementary Figure 2.
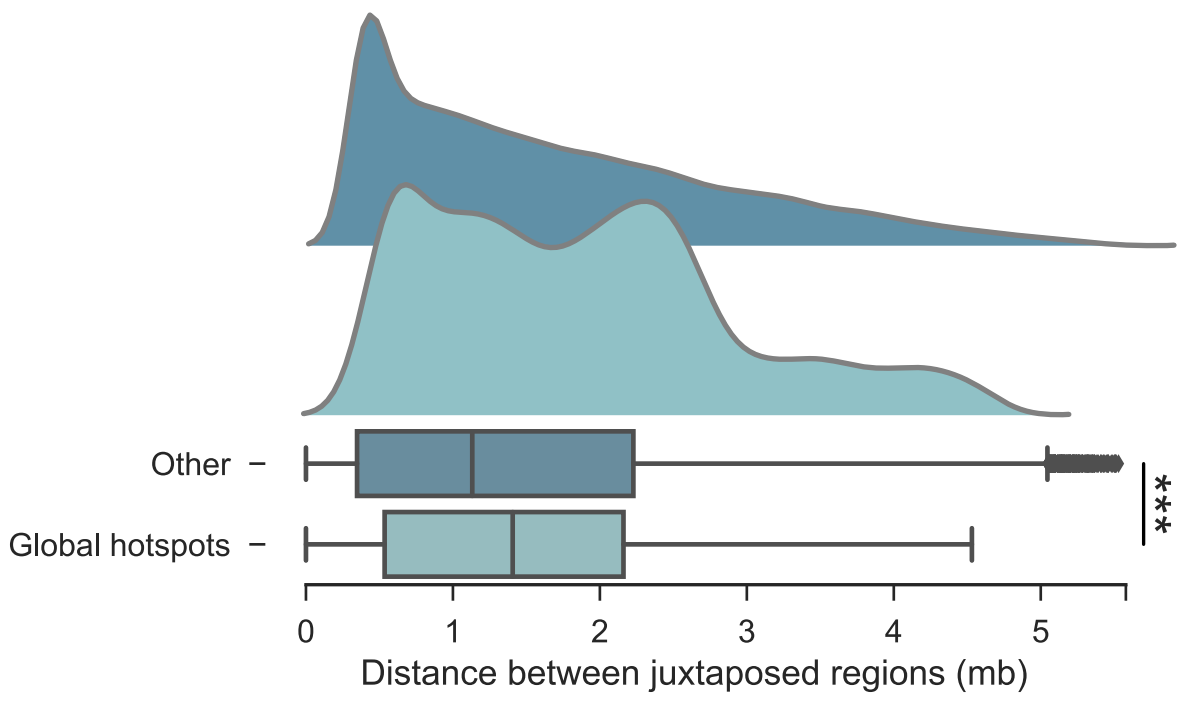
Supplementary Figure 3.

A

B

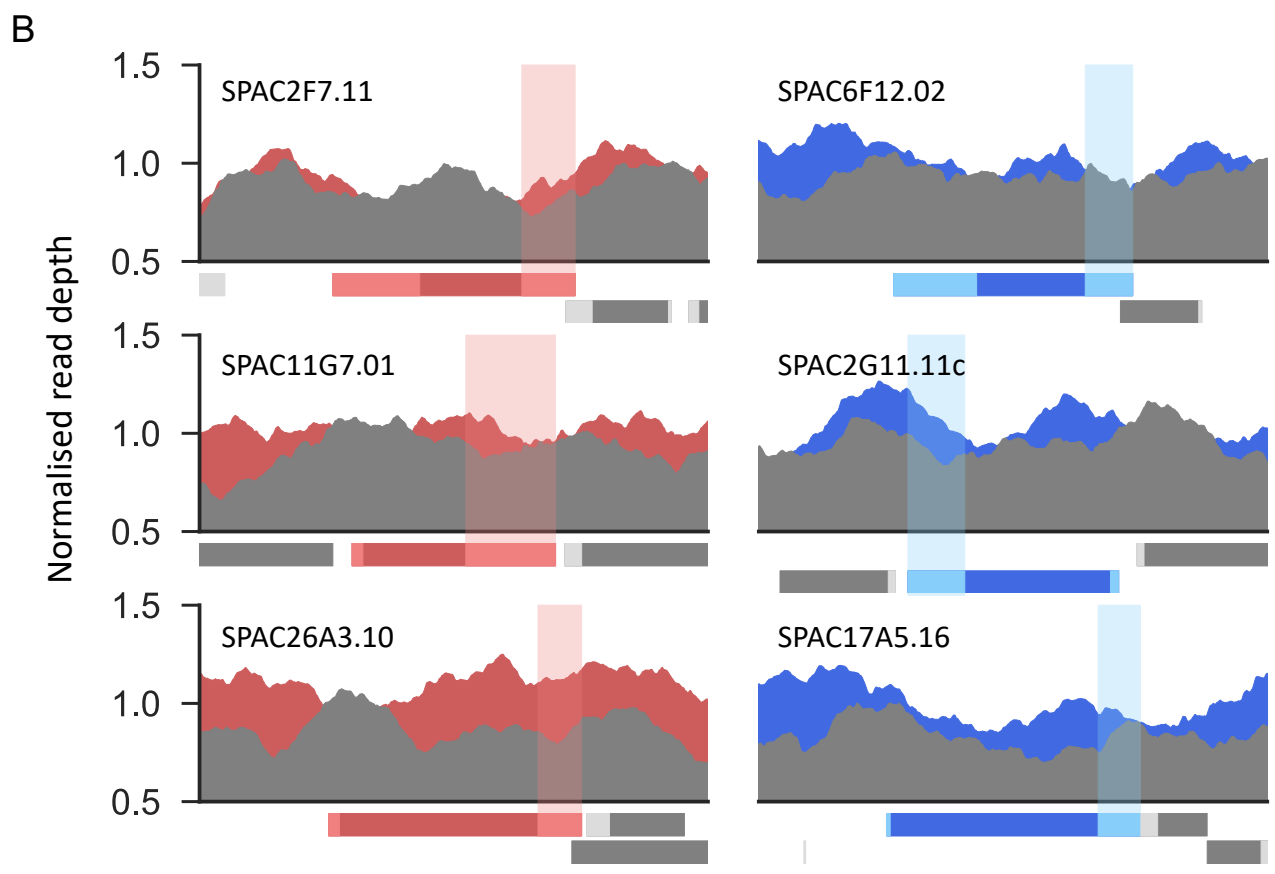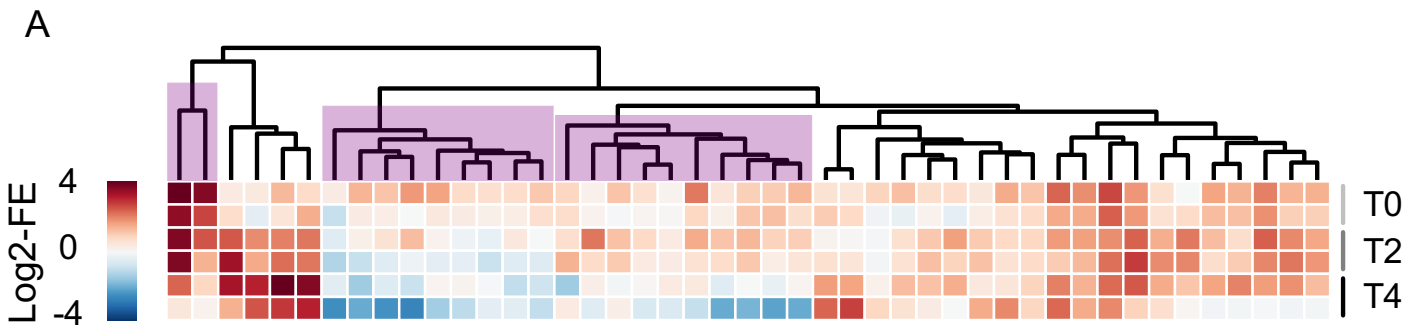| ID	factor	description	motif	p-value	e-value
1	ScAzf1	glucose signalling	aAAAAGAAa	<0.0001	<0.05
1	SpSte11	nitrogen starvation	TTCTTTGTTY	<0.01	0.2
2	ScOaf1	telomere silencing	atttCGGatWTAagxxxa	<0.001	0.7
3	ScFhl1	forkhead-like; ribosomal transcription	aaaDxGACGCAaWtxxgtt	<0.01	1.1
3	SpFkh2	forkhead-like; 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

Supplementary Figure 4.

A
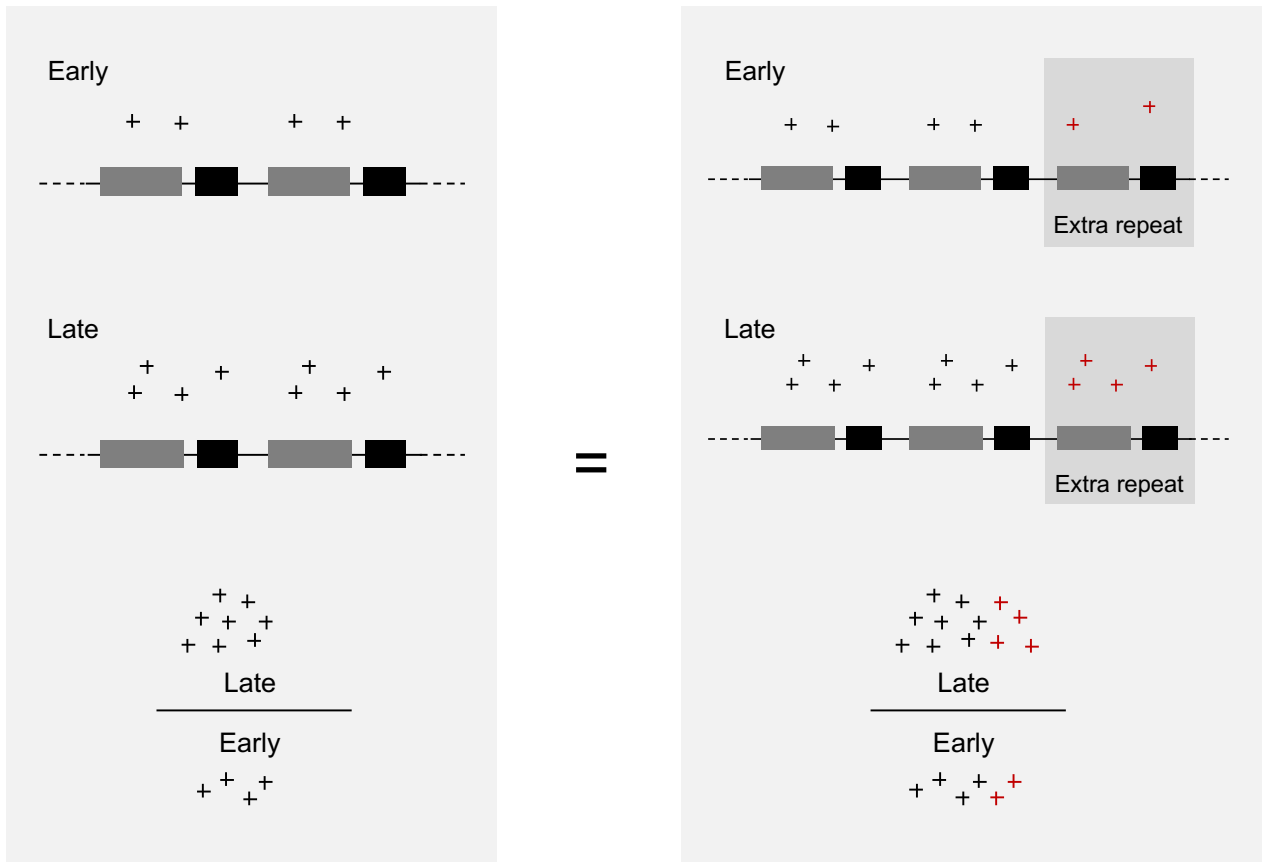
B
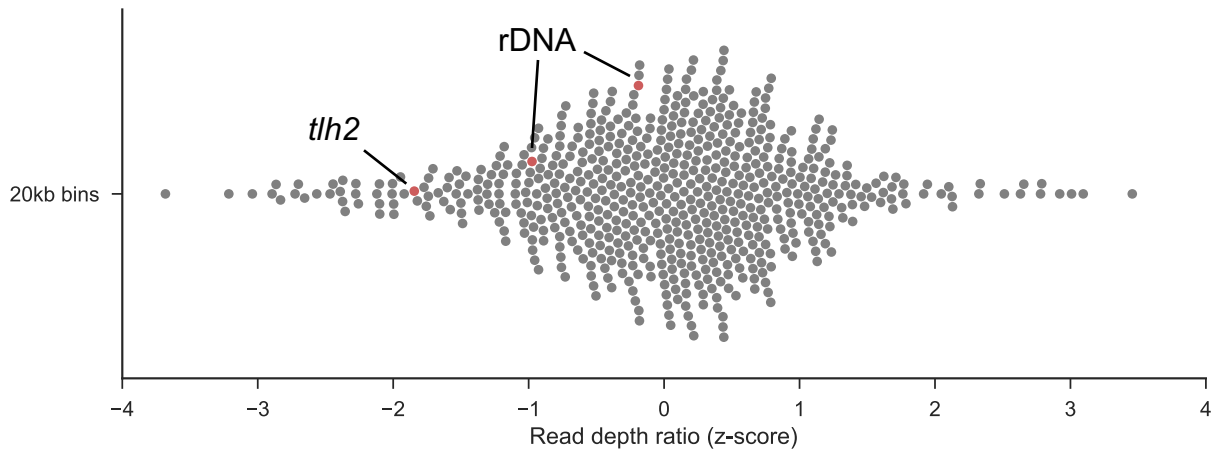
Supplementary Figure 5.
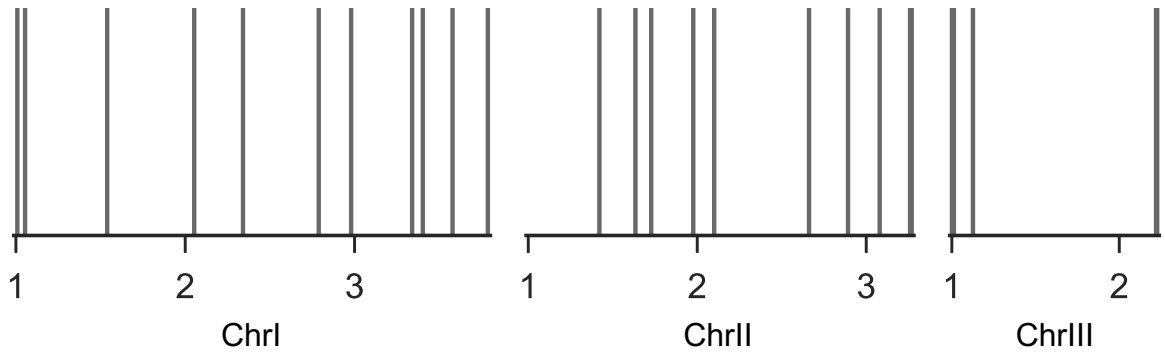
Supplementary Figure 6.

Supplementary Figure 7.