

Molecular Underpinnings and Environmental Drivers of Spontaneous Loss of Heterozygosity in *Drosophila* Intestinal Stem Cells

Lara Al zouabi¹, Marine Stefanutti¹, Nick Riddiford¹, Natalia Rubanova^{1,2}, Mylène Bohec³, Nicolas Servant², and Allison Bardin¹

¹ Institut Curie, PSL Research University, CNRS UMR 3215, INSERM U934, Stem Cells and Tissue Homeostasis Group, 75248 Paris, France

² Bioinformatics, Biostatistics, Epidemiology and Computational Systems Unit, Institut Curie, PSL Research University, INSERM U900, Paris 75005, France.

³ ICGex Next-Generation Sequencing platform, Institut Curie, PSL Research University, 75005 Paris, France.

Abstract

The genome stability of adult stem cells is of particular importance as these cells maintain long-term self-renewal capacity and can contribute extensively to adult tissues.

During development and aging, genome mutation leading to loss of heterozygosity (LOH) can uncover recessive phenotypes and be propagated within tissue compartments. This phenomenon occurs in normal human tissues, and is prevalent in pathological genetic conditions and cancers. While previous studies in yeast have defined distinct DNA repair mechanisms that can promote LOH, the predominant pathways underlying LOH in complex somatic tissues of multicellular organisms are not well understood. In addition, how environmental triggers such as pathogenic bacterial infection may impact LOH is unclear. Here, we investigate the mechanisms giving rise to LOH in adult intestinal stem cells in *Drosophila*. Our data indicate that infection with the enteric pathogenic bacteria, *Erwinia carotovora carotovora* 15 but not *Pseudomonas entomophila* increases LOH frequency. Using whole-genome sequencing of somatic LOH events, we demonstrate that they arise primarily via mitotic recombination. Molecular features of recombination sites and genetic evidence argue against formation via break-induced replication and instead support cross-over events arising from double Holliday junction-based repair. This study provides a mechanistic understanding of mitotic recombination in stem cells *in vivo*, an important mediator of LOH.

Introduction

Studies over the past 10 years have brought to light the fact that healthy, adult tissues are composed of a patchwork of diverse genomes arising from somatic mutation of stem and progenitor cells [1]. A major focus of this body of work has been on the acquisition of somatic point mutations, which are easily detected by short-read sequencing-based approaches [2-4]. Less attention, however, has focused on loss of heterozygosity (LOH) of alleles, more challenging to detect in bulk sequencing. LOH is frequently due to copy number changes, removing one allele, but can also be copy neutral (cnLOH), alternatively referred to as uniparental disomy where 2 identical alleles are present. While the mechanisms underlying development of cnLOH are not fully understood, chromosome mis-segregation would lead to events affecting whole chromosomes. In contrast, repair of DNA damage mediated by recombination machinery would lead to cnLOH events affecting only chromosome arms or segments of chromosomes.

cnLOH, also known as “mitotic recombination”, was originally identified in classic studies from *Drosophila* ([5] and reviewed in [6]), though it also has important beneficial and detrimental consequences on a wide-variety of human pathologies. For example, it has been shown to underlie the spontaneous cure of skin diseases such as Ichthyosis and Epidermolysis Bullosa [7-10] and hematopoietic pathologies like Diamond-Blackfan anemia [11-14]. In these instances, a heterozygous dominant disease-causing allele is reverted to a wild-type allele upon recombination between homologous chromosomes followed by cell division (reviewed in [15]). Mitotic recombination plays a substantial role in both sporadic and familial cancers, particularly problematic for individuals who have germline mutations in *adenomatous polyposis coli* (*APC*), *retinoblastoma* (*Rb*) and *neurofibromatosis 1* (*NF1*) [16-20]. Furthermore, cnLOH has been proposed to represent an important vulnerability in cancers [21]. A better understanding of the process of mitotic recombination and its impact on healthy tissues and disease pathologies is merited.

Mechanistic insight of the molecular events of mitotic recombination has come largely from studies using budding and fission yeast [22-25]. In response to a DNA break, homologous

recombination can proceed either through the classic double-strand break repair pathway using a double Holliday Junction intermediate (dHJ) [26], via synthesis-dependent strand annealing, or by Break-Induced Replication (BIR) pathway (reviewed in [25, 27, 28]). However, only the classic dHJ and BIR pathways lead to large stretches of LOH along chromosome arms. In the classic dHJ pathway, DNA break repair proceeds via dHJ intermediates which are often processed by dissolution, creating non-crossover products [29, 30]. Alternatively dHJs are resolved via cleavage by endonucleases, which can promote crossover leading to LOH of heterozygous alleles [29]. Break-induced replication, on the other hand, involves error-prone synthesis of large portions of chromosomes resulting in LOH of heterozygous alleles within the copied region [25, 28]. Studies in mammalian cell culture and in *Drosophila* germline cells have provided additional insight into molecular mechanisms of mitotic recombination and indicated a large degree of conservation of enzymes and processes [31-37]. In particular, *Drosophila* studies revealed that mitotic recombination is normally suppressed by the activity of DNA pol theta-mediated end joining [34]. Nevertheless, many questions are outstanding. The roles of repair by the classic dHJ pathway versus BIR in somatic tissues of Metazoa, for example, is currently unknown.

The *Drosophila* adult midgut has become a powerful model system for understanding healthy tissue dynamics as well as cancer initiation, tumor progression and aging (reviewed in [38]). The intestinal epithelium is composed of differentiated enterocytes (ECs) and enteroendocrine cells (EEs) that are replaced by the asymmetric divisions of intestinal stem cells (ISCs), which are the primary dividing cell type in the midgut [39, 40]. While tissue turnover is slow in unchallenged intestines, ISC proliferation can be rapidly induced by damaging agents such as pathogenic bacteria [41-43], and other damaging agents [44, 45]. The microbiome also impacts the progression of genetically-induced tumors [46, 47]. Additionally, induced tumor models in the *Drosophila* intestine have defined cell-autonomous and non-cell-autonomous signals from the surrounding cellular niche that promote tumor growth and survival signals [48-57]. Thus, the fly gut has provided insight into tumor progression using induced tumor models.

Previously, we demonstrated that the intestine is prone to spontaneous tumor initiation through somatic mutation of ISCs involving gene deletion, chromosome rearrangement, and transposon insertion [58-60]. Furthermore, our findings suggested that a separate process, likely depending on recombination machinery, can promote LOH [58]. Here, we take advantage of the fly intestine model to address the underlying molecular mechanisms and drivers of spontaneously arising LOH in adult stem cells.

Results

Spontaneous loss of heterozygosity increases with age

In order to systematically study mechanisms and frequencies of LOH, we chose a null allele of *Suppressor of Hairless* (*Su(H)*) on Chromosome (Chr) 2L (*Su(H)^{Δ47}*, “*Su(H)-*”) as a marker gene. It encodes a transcriptional factor of the Notch pathway and its loss-of-function leads to a readily detectable tumor phenotype phenocopying that of *Notch* with large clones composed of neoplastic ISCs and EE cells [61]. The midguts of a majority of aged *Su(H)-/+* female flies presented an overall wild-type appearance, composed of large polyploid enterocytes (ECs) with interspersed enteroendocrine (EE cells) and diploid progenitor cells (ISCs and EBs; **Figure 1A-1C**). Less frequently, aged *Su(H)-/+* midguts were detected with patches of tissue with a *Su(H)* loss-of-function phenotype comprised of an accumulation of Delta (DI) positive ISCs and Prospero+ (Pros+) EEs were apparent (**Figure 1D, 1E**). These data suggest that, as we previously demonstrated for other Notch pathway components [58], the spontaneous inactivation of the wild-type allele of *Su(H)* occurs during aging representing LOH events.

The frequency of spontaneously arising LOH clones was found to increase with age. While only 0.9% guts had mutant LOH clones in 1-week old *Su(H)-/+* flies, by 3 weeks of age this rose to 10.8%, and further increased to 73% of midguts by 6 weeks of age (**Figure 1F**). 6-week old wild-type *w¹¹¹⁸* females (+/+) had interspersed ISCs and EEs and lacked aberrant clusters of these cells seen in LOH conditions (**Figure 1A, 1F**).

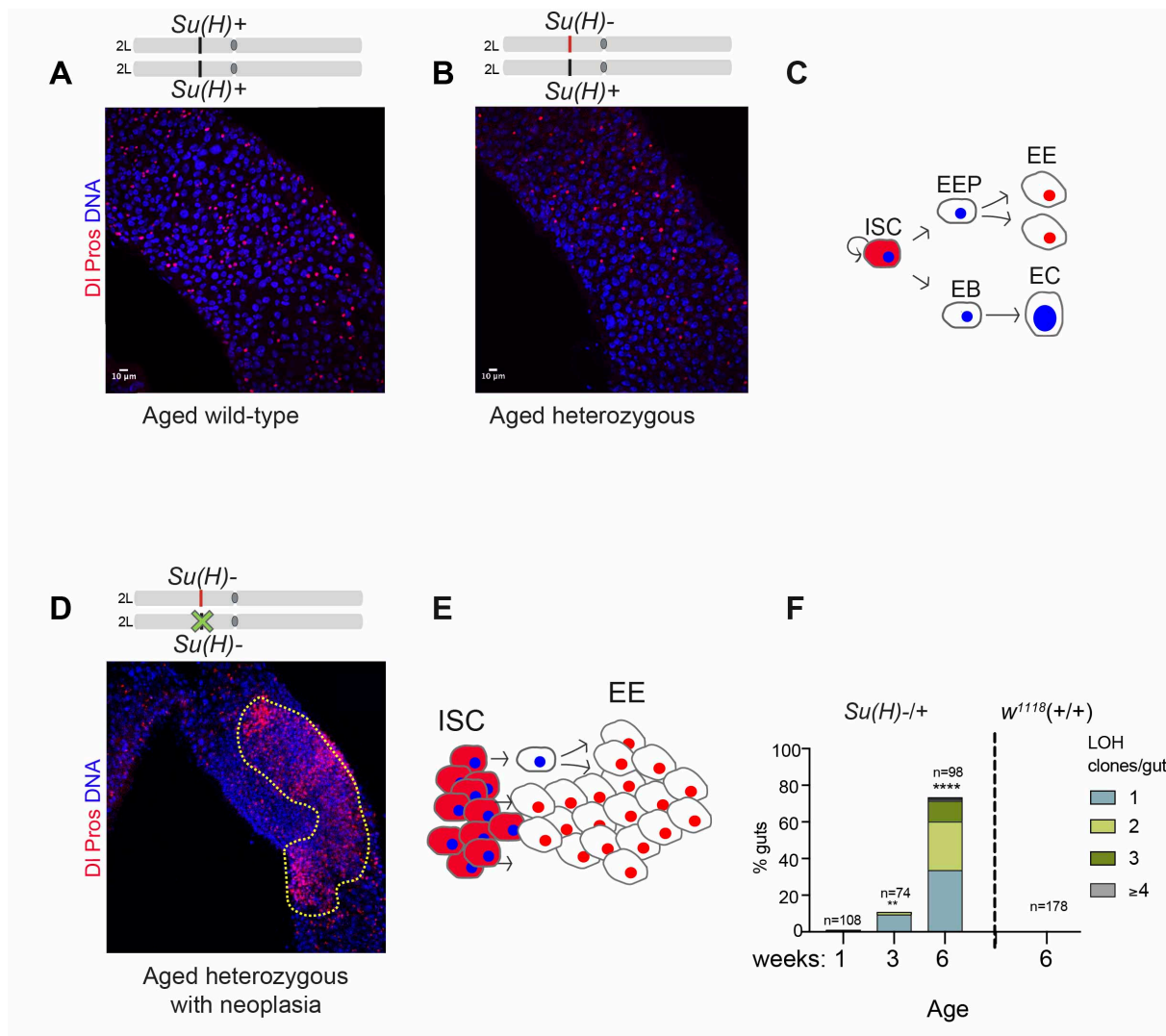


Figure 1: Aging ISCs acquire spontaneous LOH events

(A) A wild-type aged gut at 6 weeks. Polyploid enterocytes (ECs), identified by the large nuclear size (DAPI, in BLUE), are the primary cell type in the gut and are interspersed with diploid enteroendocrine cells (EEs, Pros, nuclear RED), and ISCs (cytoplasmic puncta of Delta [DI] staining RED). (B) A heterozygous aged gut at 6 weeks. A majority of tissue in the intestines of *Su(H)-/+* flies were like that of wild-type flies. (C) ISC lineage: ISCs divide to self-renew and produce EB progenitors that directly differentiate into ECs, representing ~90% of cells in the tissue. Less frequently, around 10% of ISCs, divide to self-renew and produce EE progenitors, which divide once to make 2 EE cells. (D) An example of a heterozygous *Su(H)-/+* midgut at 6 weeks of age with a neoplastic clone showing loss-of-function phenotype (outlined in yellow), composed of an excess of DI+ ISCs and Pros+ EEs (E). (F) Frequency of LOH clones in *Su(H)-/+* midguts at 1, 3 and 6 weeks of ages. The 1 week time point was used to calculate statistical significance. No midgut showed a neoplastic LOH clone in wild-type intestines.

** $p < 0.01$; **** $p < 0.0001$ (Fisher's exact test, two-tailed).

Infection with the pathogenic enteric bacteria *Ecc15* increases loss of heterozygosity

The gut is an organ that responds rapidly to changes in the environment, triggering stem cell proliferation in response to epithelial cell death [62]. To test the role of external environmental factors in driving LOH, we first wanted to validate that environmental alteration occurring in adult life could indeed affect LOH frequencies in the midgut. X-ray irradiation is known to induce chromosomal breaks that lead to LOH in developing larvae [5]. Young 1-week old *Su(H)*^{-/+} flies were irradiated with 40 Gy ionizing radiation (IR) and compared with unirradiated flies at 3- and 6-weeks post-IR (**Figure 2A-2C**). While non-IR treated flies contained 10.8% of guts with spontaneously arising LOH events at 3 weeks, IR-treated flies had a significant increase in LOH events, with 78.4% of midguts containing at least one LOH clones at 3 weeks (**Figure 2B**). In addition, most IR-treated midguts had multiple events (61.9%). At 6 weeks, almost 100% of IR-treated flies had very large LOH clones, often occupying almost the entire midgut, strongly suggesting clone fusion (**Figure 2C**). Importantly, these data illustrate that changes to ISCs during adult life can increase LOH frequency in the midgut.

We then wanted to assess whether natural pathogens may impact somatic mutation via LOH in the midgut. The pathogenic bacterial strain *Erwinia carotovora carotovora* 15 (*Ecc15*) can infect *Drosophila melanogaster*, cause epithelial damage, and promote ISC proliferation [42]. As previously reported, *Ecc15* treatment led to an increase in mitotic cells after 24 hours of treatment compared to untreated controls (**Figure 2D**). Punctual exposure of *Su(H)*^{-/+} flies to *Ecc15* during 24 hours, once per week for 3 weeks, was performed with midguts assessed at 5 weeks of age. *Ecc15* treatment led to an increase in frequency and number of LOH events compared to controls: 40.6% upon *Ecc15* treatment had at least 1 LOH event compared to 30.3% of guts in the untreated control (**Figure 2E**).

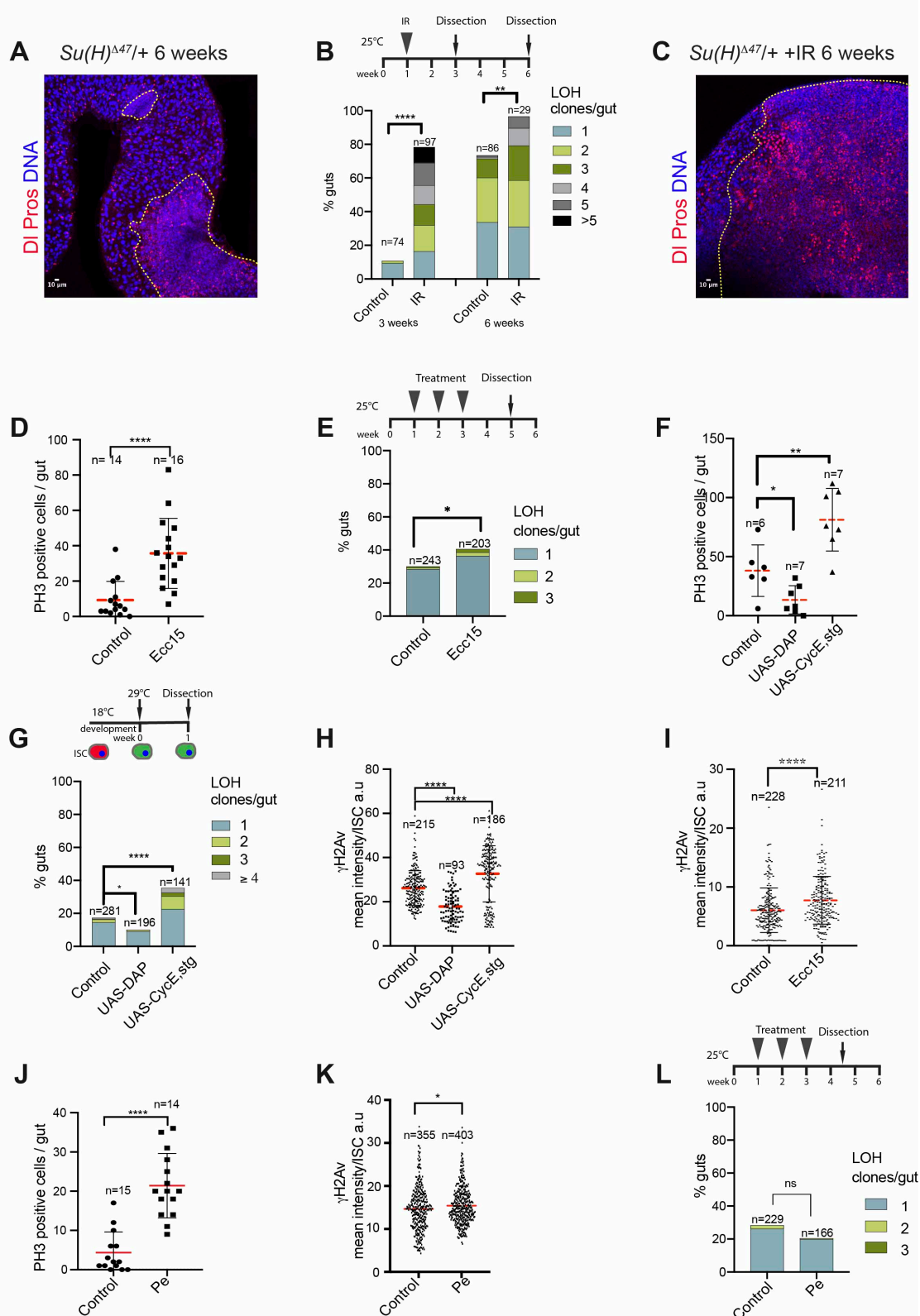


Figure 2: *Ecc15* promotes DNA damage and loss of heterozygosity

(A) Example of 2 LOH events in aged *Su(H)*^{-/+} gut. DI (ISCs, cytoplasmic RED), Pros (EEs, nuclear RED), DAPI (BLUE). (B) A significant increase in LOH frequency was found in *Su(H)*^{-/+} flies irradiated at 1-week of age with 40 Gray ionizing radiation (IR), compared to unirradiated flies at 3 and 6 weeks post-IR. Note, control data are also used in Figure 1D. (C) An example of a large LOH clone in an aged IR-treated *Su(H)*^{-/+} gut outlined in yellow DI (ISCs, cytoplasmic RED), Pros (EEs, nuclear RED), DAPI (BLUE). (D) Infection of 3-5 day old flies with pathogenic bacteria *Ecc15* stimulates ISC proliferation 24h after treatment as assessed by Phospho-Histone H3 (PH3), (t-test with Welch's correction). (E) Frequency of LOH clones in *Ecc15* treated flies compared with control. Flies were treated for 24 hours in weeks 1, 2 and 3 and were dissected in week 5, providing time for recovery to minimize overall toxicity and avoid lifespan reduction (Fisher's exact test, two-tailed). (F) PH3 quantification of control guts or those overexpressing the cell cycle inhibitor *DAP* or cell cycle drivers *CycE* and *stg* expressed in stem cells using *DI-GAL4* combined with *tub-GAL80ts* for 7 days (t-test with Welch's correction). (G) Frequency of LOH in midguts of controls and those overexpressing the cell cycle inhibitor *DAP* or cell cycle drivers *CycE* and *stg* expressed in stem cells *DI-GAL4* combined with *tub-GAL80ts* at 7 days (Fisher's exact test, two-tailed). (H) Comparison of γ H2Av mean intensity (arbitrary units) in ISCs overexpressing *DAP* (n=93, N=5) or cell cycle drivers *CycE* and *stg* in ISCs (n=186, N=6) compared with control (n=215, N=8). 'n' is the total number of ISCs quantified, 'N' is the number of guts assayed. (I) Infection of 3-5 day old flies with pathogenic bacteria *Ecc15* promoted an increase in γ H2Av 24h after treatment (n=211, N=6) compared with control (n=228, N=6; t-test with Welch's correction). (J) 3-5 day old flies dissected immediately after 24 hours of *Pe* treatment to assess a proliferative response (t-test with Welch's correction). (K) Comparison of γ H2Av mean intensity in ISCs in *Pe* treated (n=403, N=5) and control flies (n=355, N=8) after 24h of treatment of 3-5 days old flies (t-test with Welch's correction). (L) Frequency of LOH clones in *Pe* treated flies compared with control (Fisher's exact test, two-tailed).

*p < 0.05; **p < 0.01, ****p < 0.0001; not significant (n.s.).

We suspected that the effect of *Ecc15* on LOH might be due to the enhanced

140 number of cells divisions, thereby increasing the likelihood of replicative DNA damage.

To test whether increased cell division could alter DNA damage and LOH frequency,

we modified cell division rates in adult ISCs (*DI*^{GAL4} combined with *tub-GAL80ts*), using

either: 1.) the overexpression of the cyclin-dependent kinase inhibitor *Dacapo* (*Dap*), known to block the G1/S transition and thus reduce cell division [63], or 2.) the

145 combined overexpression of *Cyclin E* and *string*, previously shown to increase ISC

proliferation [64]. The frequency of proliferation, number of LOH events, and quantity of γ H2Av,

a mark of DNA damage, were significantly decreased upon *Dap* expression and increased

upon combined overexpression of *CycE* and *Stg* (Figure 2F-H). Consistent with proliferation

driving DNA damage, *Ecc15* treatment was also found to increase γ H2Av levels in ISCs (Figure 2I). These data suggest that *Ecc15* treatment could potentially impact LOH due to its effect on stem cell proliferation, which may be directly linked to the observed effects on DNA damage.

We then asked whether another pathogenic bacteria resulted in a similar effect on LOH. *Pseudomonas entomophila* (*Pe*) treatment stimulated ISC proliferation as previously reported (Figure 2J, [43]), and was found to increase a DNA damage, γ H2Av (Figure 2K), though to a lesser extent than *Ecc15*. However, treatment with *Pe* did not significantly alter LOH frequency (Figure 2L). Together, our findings demonstrate that environmental changes such as gut enteropathogens can influence the frequency of LOH events, but that they are not all equal in their capacity to induce DNA damage or LOH.

Whole genome sequencing reveals that LOH arises through mitotic recombination

We then wanted to determine the molecular mechanism underlying LOH. While our previous study hinted towards mitotic homologous recombination as a mechanism [58] as LOH frequency was diminished by balancer chromosomes known to suppress recombination, other mechanisms could not definitively be excluded. For example, chromosome loss or deletion might lead to LOH. In addition, LOH occurring via cross-over could not be distinguished from that driven by BIR in our previous study [58]. Finally, a recent study suggested that an unusual chromosome segregation mechanism, “amitosis”, can lead to LOH in the *Drosophila* midgut under conditions of starvation stress [65]. In the proposed mechanism of amitosis, enteroblast progenitor cells of the gut, after 1 round of endoreplication while they are 4n ploidy, are thought to undergo a reductive cell division leading to the segregation of 2 chromosomes originating from the same parent into one daughter cell [65].

Therefore, to differentiate between these mechanisms, whole genome sequencing (WGS) of *Su(H)* LOH clones was performed to determine the molecular nature of inactivation events (Figure 3A-C). LOH clones from 5-week-old flies were identified with a GFP reporter of EE cells, which aberrantly accumulate in the tumors as GFP+ clusters (*Su(H)*-/-; *Pros*^{V1Gal4/}

UAS-nlsGFP; **Figure 3B**). Genomic DNA was then isolated from the “tumor” mutant clone and “normal” head and Illumina paired-end (150bp) sequencing was performed on 28 female and 6 male intestinal neoplastic clones along with the head sample from the same fly as a control of the *Su(H)*^{-/+}; *Pros*^{V1Gal4}/*UAS-nlsGFP* genetic background (**Supplementary Table 1**, **Figure 3C**). Four female samples were excluded from further analysis due to low sequencing coverage after mapping to the *Drosophila* genome and removing of duplicate reads (red samples in **Supplementary Table 1**). Importantly, the parental genotypes carried a large number of single nucleotide polymorphisms (SNPs), used in the bioinformatic analysis detailed below.

To distinguish between different mechanisms of LOH, whole-genome sequencing (WGS) data of 24 female and 6 male neoplasia samples tumor/normal pairs was analyzed including copy number changes, structural variants, somatic point mutations and changes to the zygosity of parental SNPs (**Figure 3D**). In the 24 female samples, no point mutations in *Su(H)* or whole chromosomal copy number change occurred Chr 2 on which *Su(H)* is located, ruling out point mutation and aneuploidy as causes of LOH for these samples (**Supplementary Figure 1**). One sample showed a large structural variant (deletion) spanning 1.38MB of *Su(H)* (sample F20; **Supplementary Figure 2**). In one sample (F28), no LOH or mutation could be found. However, in the majority of samples (22/24 samples), changes in SNP heterozygosity were detected as a shift in zygosity from a variant allele frequency (VAF) of ~0.5 to 0.75 or above or 0.25 and below on Chr 2L (**Figure 4A, 4B, Supplementary Figure 1**). As LOH samples have some contaminating wild-type adjacent cells, this number does not go to 1 or 0. The positions of LOH invariably arose at locations between the *Su(H)* locus and the centromere 7.5Mb away and extended throughout the chromosome to the telomere, thereby resulting in LOH of the *Su(H)* locus (**Figure 4A, 4B**; for all 22 samples see **Supplementary Figure 1**). As SNP changes affected only a portion and not the entire chromosome arm, the data supported a mitotic recombination-based

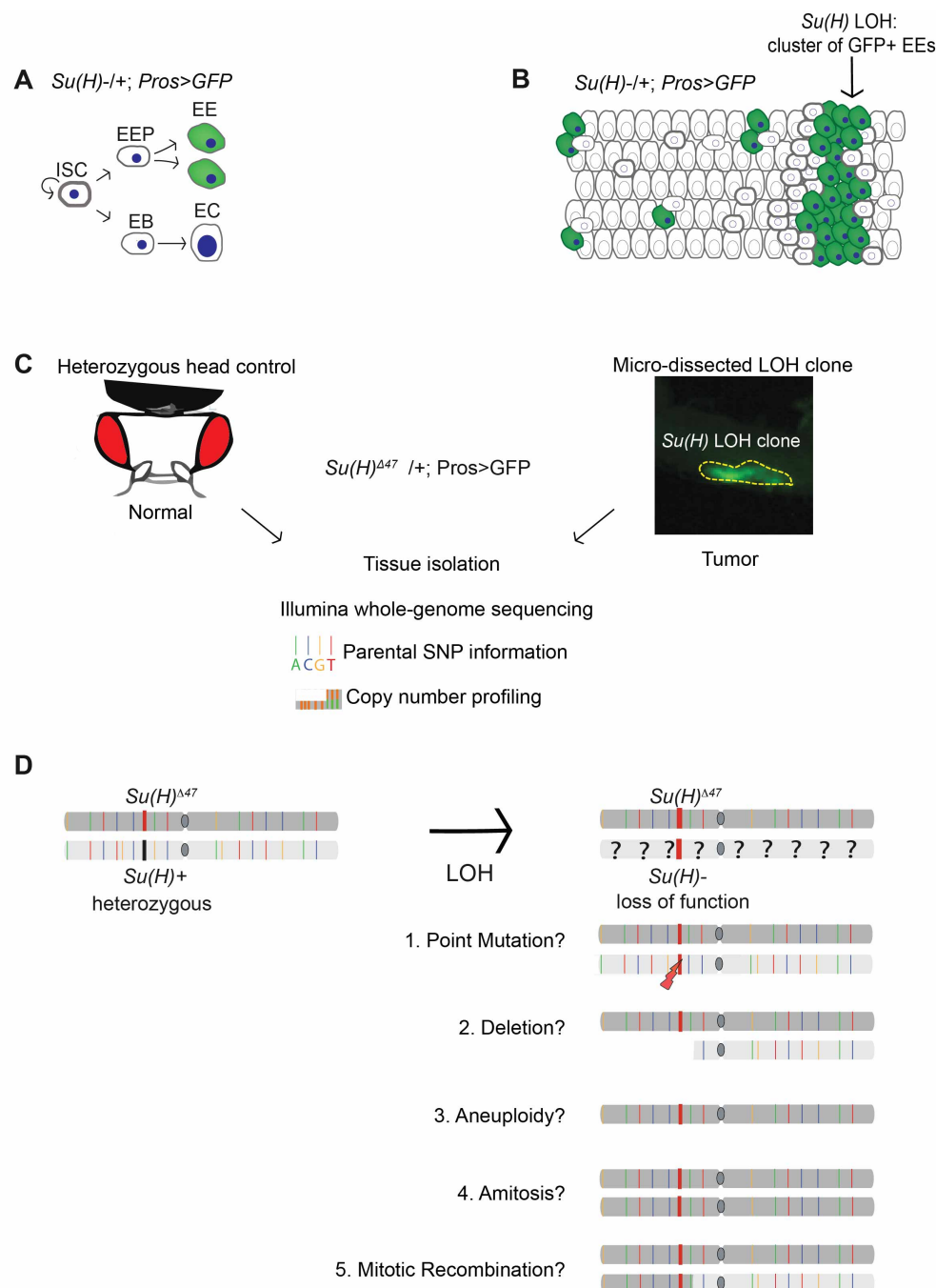


Figure 3: Sequencing setup and analysis predictions

(A) *ProsGal4* UAS-GFP transgenes drive GFP specifically in EEs allowing for isolation of large green LOH clones composed of ~1000-5000 cells. (B) Schematic of a *Su(H)* LOH clone recognized by the accumulation (marked by an arrow) of GFP+ EEs. (C) The GFP+ LOH clones were microdissected along with the head of the same fly for Illumina paired-end sequencing. Of note, the head was used as a control because it contains primarily diploid cells like the neoplastic LOH (composed of ISCs and EEs) as opposed to adjacent gut tissue, which has complex polyploid genomes due to the presence ECs. Germline single nucleotide polymorphisms (SNPs), somatic point mutations (SNVs), and copy number profiling was assessed. (D) SNP, SNV, and copy number profiling distinguish between 5 potential mechanisms by which LOH can arise.

mechanism and not amitosis, which would affect the entire chromosome. Further validation of this was obtained taking advantage of the chromosomal position of a UAS-GFP located on Chr 2L (28F3-28F5), more telomeric than the *Su(H)* locus on Chr 2L (35B), whose expression was lost in LOH clones (**Supplementary Figure 1B-C'**).

210 Additionally, evidence for LOH via mitotic recombination in 1 of the males samples was also found. A shift in allele frequency of SNPs on Chr 2, resulted in LOH of a large portion of the chromosome arm, including the *Su(H)* locus (**Supplementary Figure 3A-A'**). The remaining 5 male neoplastic samples had inactivating deletions of the *Notch* locus located on Chr X, hemizygous in males (**Supplementary Figure 3B-B'**), consistent with events that we
215 previously described in wild-type male flies leading to neoplasia [58, 60]. Therefore, in total, mitotic recombination was detected in 23 samples (22 females and 1 male).

These data provide evidence that the primary means of LOH leading to spontaneous neoplasia formation in intestinal stem cells is mitotic recombination. Importantly, our data rule out a major contribution of other mechanisms of LOH including (1) point mutations, (2)
220 deletions, (3) chromosome loss, and (4) amitosis. Furthermore, the long track of LOH extending to the telomere rules out repair mechanisms that result in short tracks of LOH that do not lead to crossover, including gap repair, single-strand annealing, and synthesis-dependent strand annealing [27].

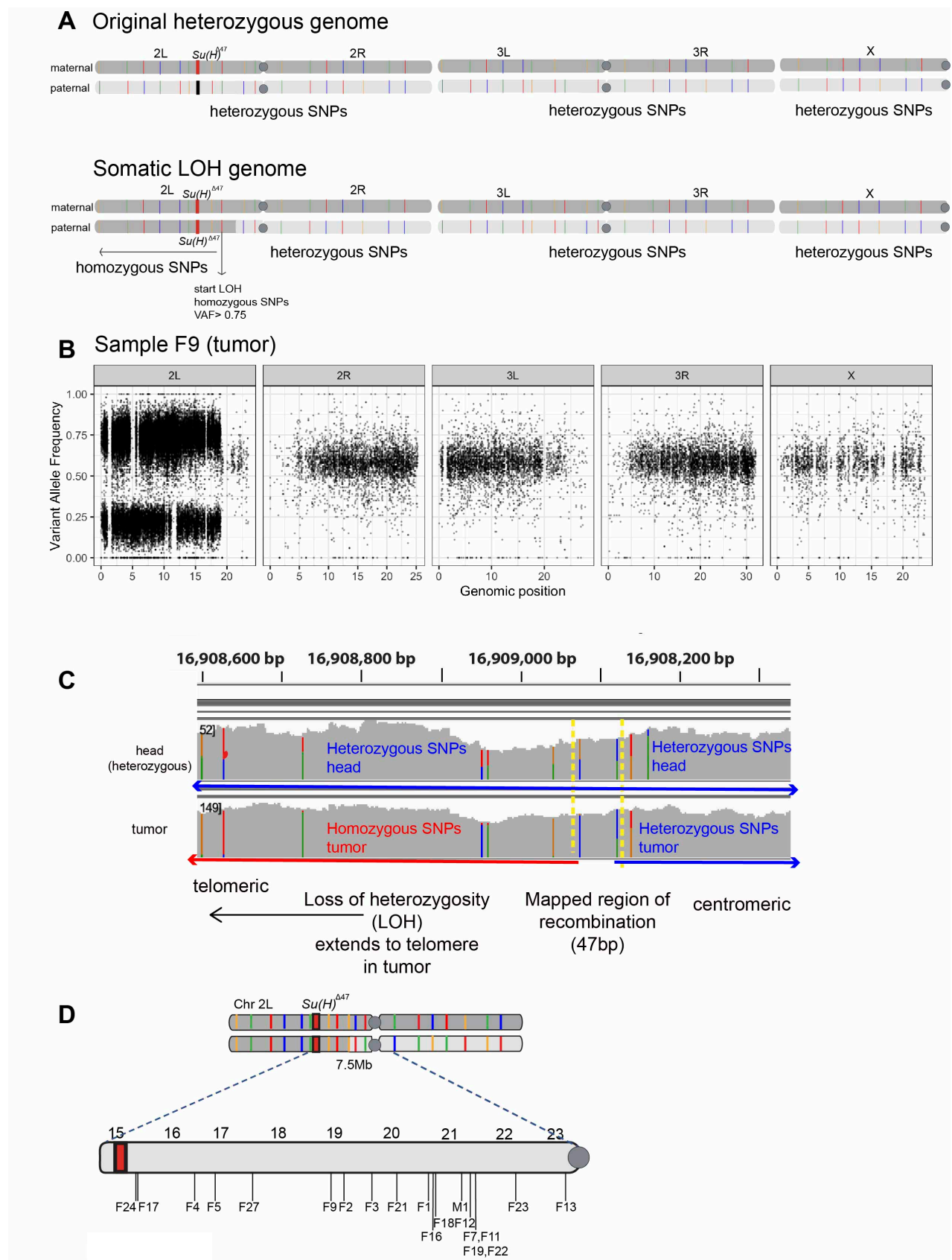


Figure 4: Sequencing *Su(H)* LOH clones reveals loss of heterozygosity primarily arises though mitotic recombination

(A) A schematic corresponding to the Variant Allele Frequency (VAF) plot below. (B) A representative VAF plot of a female sample (F9). In this LOH sample, chromosomes 2R, 3L, 3R and X had heterozygous SNPs, represented at ~0.5 VAF. In contrast, chromosome 2L has undergone an LOH event, with SNPs becoming homozygous (>0.75%) shown in the left panel. In a scenario where the tumor purity is 100%, the VAFs in LOH regions should be 1 (100%) or 0, however given the manual nature of the tumor micro-dissections, contaminating non-tumor cells make up part of the sequenced tumor resulting in deviations from this. (C) IGV view of mapping of recombination sites in LOH neoplasias compared to heads. Both head and tumor (LOH) samples show heterozygous SNPs on the right side, seen by coloured SNPs each at roughly 50% of the sequencing depth. In contrast, on the left side, these SNPs now become homozygous in the tumor, where they are still heterozygous in the head. Between the yellow lines is the region where recombination took place, mapped within 47bp the last heterozygous SNP and the first homozygous SNP in tumor. (D) All of the mapped recombination sites occurred between *Su(H)* locus (red) and the centromere in female samples (F1-F27) and the male sample M1. Numbers on the chromosome arm correspond to chromosome coordinates.

The Histone Locus Cluster is enriched for sites of mitotic recombination

We next wanted to understand whether there are underlying DNA sequence features that contribute to the DNA damaging event driving mitotic recombination. We were able to map the sites where the recombination occurred in 20 out of the 23 samples, detected as the chromosomal region (sequence) between the first homozygous SNP (from the centromere) and the position of the last heterozygous SNP (from the centromere) in the LOH “tumor” sample (see Methods and **Figure 4C, 4D, Supplementary Table 2**). Recombination sites were mapped from a resolution of 47 bp to 111kb and had a median size of 1865bp (**Supplementary Table 2**). Three samples had low purity of the neoplastic clone precluding mapping of the region of recombination (see Methods). The position of recombination was then compared to various sequence features. There was no significant overlap with mapped R loops [66] and predicted sequences for form non-B-form DNA such as G-quadruplexes, cruciform DNA, though did find an a significant enrichment for short inverted repeats (SIRS, **Supplementary Figure 4**).

Interestingly, 4/20 samples had a recombination site that arose within the *Histone Locus Cluster*, a region of 110 kb within the 7.5 Mb region between the centromere and the *Su(H)* locus (**Figure 5A**). These are only 4/20 samples and, thus, should be interpreted with caution. However, the occurrence of recombination sites within this small region (110 kb) was

significantly enriched to what was expected using simulated data (**Figure 5B**). The *Histone Locus Cluster* is an array of 100 copies in tandem each containing 5 Histone genes (H2A, H2B, H3, H4 and H1) (**Figure 5A**), which has features that could contribute to replication problems. First, it has tandem repeats which may cause problems for the replication fork. Secondly, the *Histone Locus Cluster* is also only highly transcribed during S phase when Histones are incorporating into newly synthesized DNA [67], which could possibly make it more prone to replication fork collisions with the transcription machinery and lead to DNA damage. Further investigation of the *Histone Locus Cluster* could provide important insight into genomic features potentially driving mitotic recombination from the homologous chromosome.

Rad51 promotes loss of heterozygosity

Our data above indicate that mitotic recombination plays a major role in LOH in intestinal stem cells. As recombination results from DNA double-strand break repair (DSBR) that utilizes the Rad51 protein to invade a donor DNA molecule (**Figure 6A**) [27, 33], we predicted an involvement of the *Drosophila* Rad51 protein SpnA. We hypothesized that in absence of Spindle-A (SpnA), damaged stem cells would be unable to repair DNA and would consequently die, reducing the number of LOH clones detected. Consistent with this, LOH frequency decreased from 32.6% in *Su(H)*-/+ controls to 7.5% in *Su(H)*-/+; *SpnA*⁰⁹³/*SpnA*⁰⁵⁷ null mutants (**Figure 6B**). No obvious effect on stem cell number occurred in *SpnA*⁰⁹³/*SpnA*⁰⁵⁷ mutants, ruling out the trivial explanation that the ISC population is reduced (**Supplementary Figure 5A, B**). Similar to *SpnA* mutants, there was a reduction in LOH clones upon knockdown of *Rad51* (*SpnA*) specifically in adult stem cells (*DI-Gal4*, *tub-Gal80ts*; **Supplementary Figure 5C**, see Methods for information on scoring). We conclude that LOH arises in ISCs largely from a Rad51-dependent repair mechanism driving mitotic recombination.

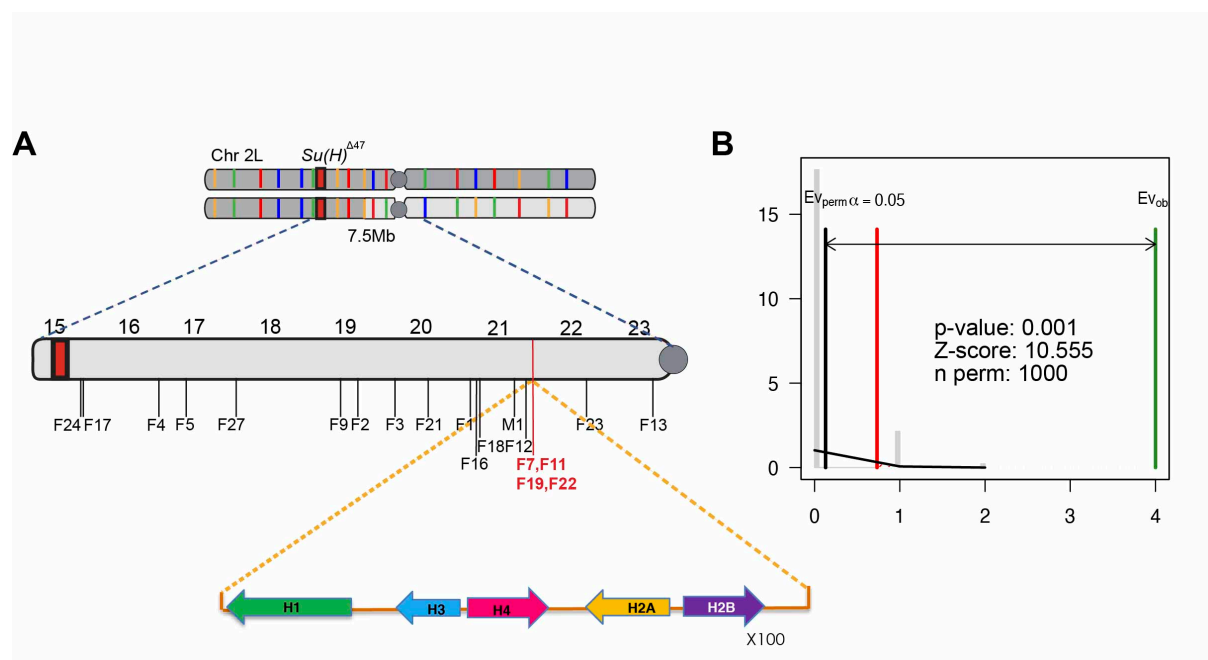


Figure 5: The Histone Locus-Cluster: a putative genomic feature driving mitotic recombination

(A) 4 samples (F7, F11, F19 and F22 shown in red) had recombination sites that occurred within the Histone Locus Cluster. Due to the repeat nature, the exact positions could not be mapped within this ~110kb region. Below the chromosome arm is a schematic of the Histone Locus-Cluster; an array of 100 copies of 5 Histone genes. (B) Permutation test carried out using RegioneR, $p < 0.001$. Test association between mapped recombination sites and the Histone Locus-Cluster. Shuffles: 1000 constricted to the region between $Su(H)$ and the centromere (where recombination can be detected).

Genetic requirements and molecular signatures support double Holliday junction

275 resolution leading to crossing-over

Two distinct mechanisms of mitotic recombination could explain the long tracts of LOH that were detected in intestinal stem cells: (1) Crossover, resulting from a classic double-strand break repair involving a dHJ structure, whose resolution would lead to reciprocal exchange of segments of the homologous chromosomes (**Figure 6C**). (2) BIR, in which an error-prone polymerase would copy material directly from the homologous chromosome

280

(**Figure 6D**). While both mechanisms could lead to long stretches of LOH spanning the

chromosome arm and depend on *Rad51*, they differ in their genetic requirements. For this reason, we first decided to test roles of DNA repair proteins specific to each mechanism.

The resolution of dHJ DNA structures relies on resolvases such as Gen1 and Mus81 [34, 68]. While *mus81^{nhe1}* mutant flies did not show consistent differences between mutant and control (**Supplementary Figure 6A**), the loss-of-function of Gen1 resolvase, encoded by *gen*, on the other hand, showed a significant decrease in LOH events reducing LOH from 59.3% in controls to 42.16% in *Gen1* mutants (**Figure 6E**). These data suggest that *Gen1* is important to promote LOH likely through facilitating DNA repair via dHJ resolution that results in crossover.

BIR relies on the DNA helicase Pif1, important to unwind the double-strand DNA thereby allowing to the copying DNA from the homologous donor chromosome [69-72] (**Figure 6D**). *Pif1* null mutant flies are not viable [70]. Therefore, *Pif1* RNAi was expressed specifically in adult ISCs though it did not significantly alter LOH frequency in *Su(H)*^{-/+} flies (**Supplementary Figure 6B**), suggesting that Pif1 may not play an integral role in the repair giving rise to the LOH events, though inefficient knockdown cannot be excluded. We, therefore, also examined the genomic features of the chromosomes with LOH.

Taking advantage of the high density of parental SNPs, we could assess the regions surrounding the recombination site for evidence for potential molecular features associated with dHJ resolution or BIR. 16/20 samples showed a simple shift from heterozygous to homozygous, and appear to be homozygous throughout the chromosome arm (**Figure 4C**, **Supplementary Table 2**). In contrast, 4/20 samples had evidence for intervening tracts where SNPs would alternate from heterozygous to homozygous and then homozygous to heterozygous again, followed by heterozygous to homozygous and crossover (**Supplementary Figure 7A-C**). For example, sample F1 had an initial shift from heterozygous to homozygous SNPs (at position Chr 2L: 20701878-20702379), then 1.5kb more distally, a region of heterozygosity was again detected, clearly supported by 2 informative SNPs, followed by again a shift from heterozygous to homozygous throughout the chromosome arm (**Supplementary Figure 7A**). Similar shifts from heterozygous to homozygous and

homozygous to heterozygous could be detected on 3 other samples (**Supplementary Figure 7B, 7C**, and data not shown for sample F13 where more than 3 shifts were detected suggesting a complex event). These types of tracts could result from predicted outcomes of resolution of dHJ structures, mismatch repair of heteroduplex DNA arising from dHJ structures, or from template-switching previously shown to occur in 20% of BIR events [25, 73].

In order to further distinguish between mitotic recombination driven by dHJ resolution versus by BIR, we next assessed the accumulation of somatic point mutations, a hallmark of BIR but not dHJ processing ([74] and reviewed in [28]). Somatic point mutation (single-nucleotide variant; SNV) calling was performed on the genomic samples and none showed evidence of increased mutations near the recombination sites or throughout the arm of Chr 2L (**Figure 6F, Supplementary Figure 6C**). Therefore, these data argue against BIR as a primary mechanism driving LOH in ISCs.

Altogether, these data support an important role of LOH events being driven by cross-over arising from dHJ resolution, and not BIR. While this mechanism allows repair of double-strand breaks in stem cells promoting their survival, it leads to LOH of the tumor suppressor *Su(H)* and drives neoplasia initiation.

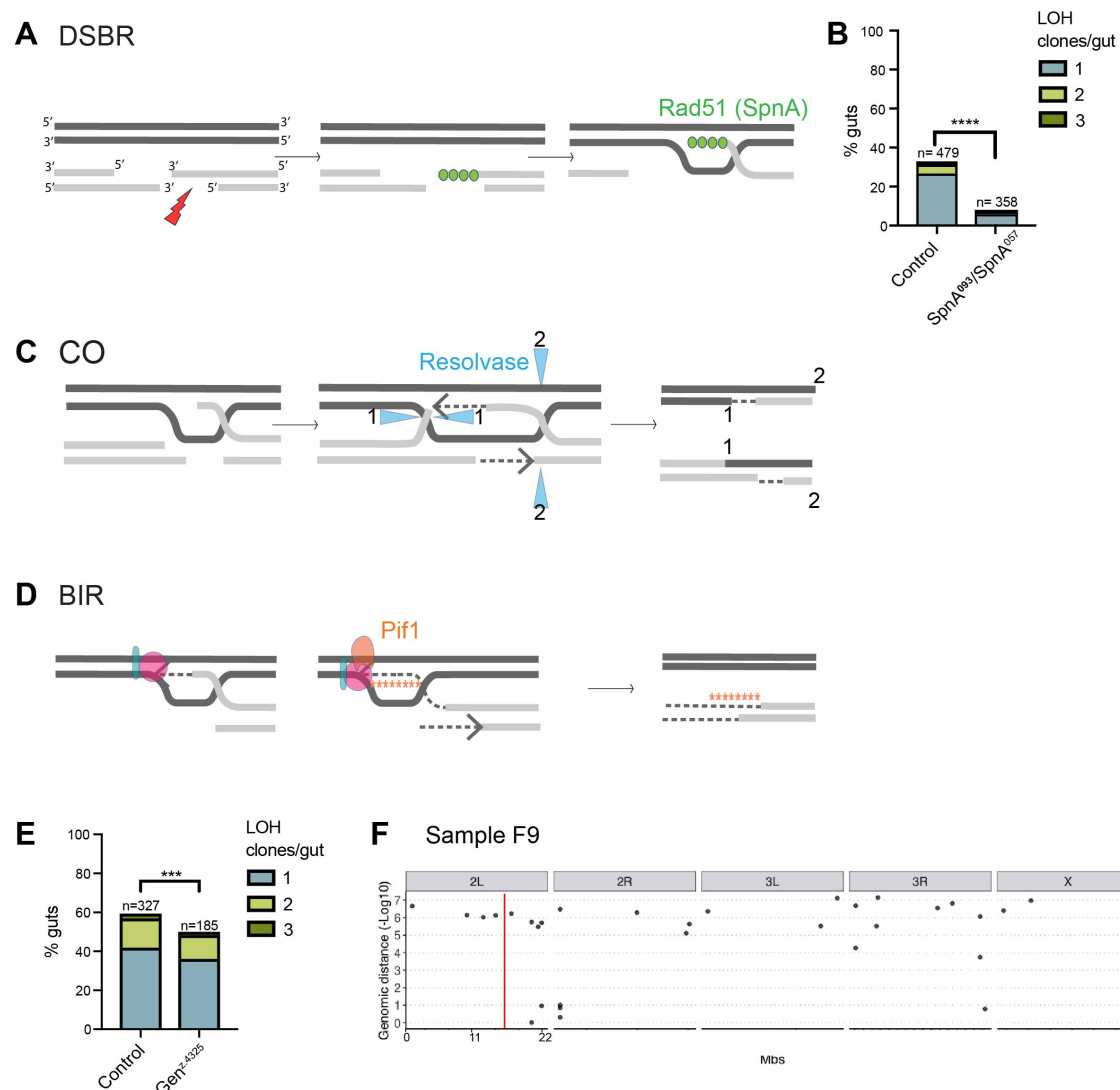


Figure 6: LOH largely arises from a Rad-51 dependent mitotic recombination consistent with double Holliday junction resolution

(A) Double-strand break repair (DSBR) is initiated by 5' end resection and Rad51(SpnA) coating of a 3' end. (B) Frequency of LOH events in control compared to SpnA null mutant. Fisher's exact test p<0.0001. (C) A crossover (CO) mechanism is mediated by a dHJ and relies on resolvases. Cleavage at sites 1 and 2 result in CO occurring between homologous chromosomes. Depending on the DNA strand, the resulting repaired chromosome can either contain one long LOH tract, or intervening regions of gene conversion nearby the initial breakpoint. (D) Break-induced replication (BIR) relies on a Pif1 helicase to unwind the template that is used for repair. Repair of the broken chromosome occurs by leading strand copying of the template, and subsequent lagging strand synthesis. It is prone to mutagenesis (stars). The resulting chromosomes are shown, where one is unaltered and one has been repaired by the de novo synthesis. (E) Frequency of LOH events in control compared to gen null mutant. Fisher's exact test p<0.0001. (F) Representative rainfall plot showing no mutation pileup by the mapped recombination site in sample F9. The red line indicates the site of mitotic recombination.

Discussion

Our findings reveal essential intrinsic and extrinsic factors acting on DNA damage and repair in adult stem cells influencing LOH, an important mechanism of tumor-suppressor inactivation. Using whole-genome sequencing and genetic approaches, our data suggest that a major contributor to genome alteration in *Drosophila* ISCs involves mitotic recombination-driven LOH that primarily occurs through a classic HR mechanism depending on dHJ resolution.

We find that differences in the intestinal microbiota can have distinct impacts on the mutation rate of a tissue. Infection by *Ecc15*, but not *Pe*, altered LOH frequency. *Ecc15* and *Pe* both induce damage to the intestinal epithelium and promote ISC proliferation [42, 43]. Interestingly, we find here that gut enteric pathogenic bacteria can also increase marks of DNA damage. In our experiments, *Ecc15* led to a higher amount of DNA damage than *Pe*, which could explain why *Ecc15* increased LOH frequency but not *Pe*. We also found that under the doses of bacteria that we used, *Ecc15* led to a more pronounced increase in proliferation than *Pe* as assayed 24 hours after infection. Of note, the effects of *Ecc15* on DNA damage and LOH can be explained by the increased amount of replication-associated DNA damage as expression of *CycE* and *Stg* in ISCs can also increase LOH frequency and DNA damage. It is possible, however, that other differences between these bacteria or their interaction with the midgut epithelium could also underlie the effect on LOH. Indeed, these bacteria induce overlapping as well as distinct changes in gene expression in the midgut [75]. In addition, at high doses, *P.e.* can inhibit the translation machinery and block ISC proliferation. It is possible that roles in translation or subtle differences in stem cell viability may preclude detections of changes in LOH frequency in *Pe* conditions. In humans, inflammatory bowel disease and colorectal cancer have been linked to alteration of the microbiome [76-78]. In particular, the NC101 strain of *Escherichia coli* causes DNA damage [79], and creates a specific mutational signature [80]. Determining how additional pathogenic bacteria and other changes to environmental conditions affect LOH events may provide strategies for cancer reduction or prevention.

Using high-resolution SNP-based mapping of recombination regions, our study provides direct evidence of mitotic recombination as an underlying mechanism occurring in adult somatic stem cells. Our genetic experiments support these molecular findings as we found reduced LOH in mutant contexts of *Rad51* (*spnA*), and *gen*. Our recent studies illustrated that, in addition to genome alteration by recombination described here, point mutations, deletions, complex structural variants, and transposable element mobility all alter the somatic genome of ISCs and could, in theory, contribute to LOH [59, 60]. Indeed, we did find one example of LOH driven by a large 1.38 Mb deletion event. However, our data strongly support mitotic recombination being the primary mechanism by which LOH is generated in ISCs. Importantly, we found no evidence supporting a BIR mechanism as no increase in point mutations was found at LOH regions for the 23 samples that we assessed as would be predicted from previous work [74]. While our study could only assess those events inducing LOH, it is likely that a much higher number of DNA double strand breaks are produced but repaired using the sister chromosome, which would be undetectable in our assay. Our findings that dHJ resolution is an important contributor to LOH in somatic tissues suggest that targeting this pathway may be beneficial in patients with germline mutations in recessive tumor suppressor genes or dominant disease genes.

One important question is whether the sites at which recombination arises are simply sites of DNA damage or whether certain underlying genomic features specifically promote crossover events. Previous studies mapping positions of crossover, for the most part, have been performed using low-resolution techniques that precluded fine mapping of DNA sequences. To our knowledge, our study is the first high-resolution mapping in a Metazoan somatic tissue. In a study of cell lines derived from colorectal cancer driven by LOH of APC, sites of mitotic recombination were resolved to low resolution (~4.5 Mb) and the authors suggest that the sites were non-random and possibly associated with low copy repeats (LCRs) [16]. Previous findings mapped mitotic recombination within the *Drosophila* male germline at low resolution using marker genes concluded that sites were non-random and provided indirect evidence that they may be linked to problems arising during replication [33]. A more

precise mapping has been achieved in yeast [81-84]. These studies found a mitotic recombination hotspot near sites of replication termination, inverted repeats of Ty transposable elements, and G-rich quadruplex sequences [81-84]. While our sample size (n=23) is limited, we did not detect enrichment at mapped crossover points of these sequences. However, we did find a significant enrichment over what would be predicted by chance at SIRS and the *Histone Locus-Cluster*. Our findings of enrichment of SIRS at recombination sites are consistent with previous reports of their ability to stall replication forks and their enrichment at translocation break-points in human cancers [85]. The *Histone Locus-Cluster* encodes a ~100 tandem repeats of 5 histone genes (*Histone 1*, *2A*, *2B*, *3A* and *3B*). Why this region has an elevated number of recombination sites than would be predicted by chance is not currently clear, however, we speculate that it is related to the unique features of these genomic locus. One peculiarity of these genes (excluding *Histone H1*) is their exclusive expression at high levels specifically during S phase [67], at which time collisions with DNA polymerase could occur. In addition, this genomic region is also unique in that it forms a nuclear subdomain, the Histone-locus body, where specific *Histone 3'* RNA processing occurs, as these are unique in lacking canonical poly A tails. Consistent with DNA damage at the *Histone Locus Cluster*, the *HIST1H* cluster in human B-cells is found to acquire marks of γ H2AX [86]. The potential of other genes highly expressed during S phase to be hotspots of DNA damage merits further investigation.

Our study underscores the notion that adult organs are genetic collages, where somatic mutations in adult progenitor cells drive divergent genomes within a common tissue. Given the number of events that we detect here for a marker at one genomic location on 1 out of the 5 major chromosome arms, we can estimate that ~40 LOH events occur per gut for distal genes near the telomeres. As there are ~1000 ISCs per midgut, we can estimate that 1 in 25 ISCs has an LOH event, raising the likelihood that this mechanism of somatic genetic diversity could alter tissue dynamics during the ageing process. Aside from arising somatically in Metazoa, mitotic recombination also occurs in plants as well as diatoms, where it is proposed to contribute to adaptive evolution of clonal populations [87, 88]. Future studies will

likely reveal both common as well as species-specific molecular and environmental mediators of mitotic recombination.

Methods

Drosophila stocks

The following fly stocks and alleles were used in this study: From the Bloomington stock center, P(UAS-GFP.nls) (BL 4776). The following stocks were generous gifts: *SpnA*⁰⁹³, *SpnA*⁰⁵⁷, *w*¹¹¹⁸ (M. McVey), *Gen1Z*⁴³²⁵, *Mus81*^{nhe1} (J. Sekelsky), *ProsV1Gal4* (J. de Navascués), *Su(H)*⁴⁴⁷, (F. Schweisguth), *O-fut1*^{4R6} (K. Irvine), *DI*^{Gal4} (S. Hou), *N*^{55e11} (F. Schweisguth), *UAS-DAP*, *UAS-CycE*, *Stg*, *tubGAL80ts* ; *DI-GAL4* (B. Edgar). From the Vienna *Drosophila* Resource Centre (VDRC): *Rad51(spna)* RNAi (VDRC 13362), *Mus81* RNAi (VDRC 33688), *Pif1* RNAi;-(VDRC 34533). For all experiments of LOH, the second chromosomes assayed for LOH were identical between the controls and experimental samples.

Drosophila husbandry

For standard aging of experiments in **Figures 1, 2A-E, 2I-L, 6B, 6E, Supplementary Figures 5A-B, 6A**, female flies were maintained at 25°C on a standard medium composition. For aging experiments, adult progeny were collected at 25°C over 3-4 days. Females were aged with males in the same cage (plastic cages 1 cm diameter, 942 ml). 400-600 flies/cage. Freshly yeasted food was provided in petri dishes every 1-2 days. Flies were transferred every 7 days without CO₂ anaesthesia to clean cages. Dead flies were scored upon each food change to assess survival rates. Female flies were dissected at 5-6 weeks unless otherwise indicated.

For experiments in **Figures 2F-H, Supplementary Figures 1B-C', 5C, 6B** the genetic background *Su(H)^{Δ47}/UAS-GFP; Df^{Gal4} tub-GAL80ts/+* was used to drive stem cell specific expression in the ISC, due to the *Df^{Gal4}* driver. Crosses were maintained at 18°C on standard medium, flies were also maintained at 18°C during development and metamorphosis. Newly eclosed flies were collected over 5-7 days. Flies were maintained at 29°C thereafter. The shift to 29°C induced ISC specific UAS-driven transgene as the Gal80 repressor is temperature sensitive. Flies were dissected after 1 week for *UAS-DAP*, *UAS-CycE*, *stg* or after 2 weeks at 29°C for *Rad51RNAi*, *Mus81 RNAi*, *Pif1 RNAi*.

Bacteria Treatments

Adult *Su(H)^{Δ47}/+* flies were treated for 24 hours on filter paper soaked with *Ecc15* / *Pe* or control solution (see below) covering sugar agar (1.5%) plates. *Ecc15*/ *Pe* treatment: a 1:1 mix of OD 200 *Ecc15* culture and 5% sucrose. Control: a 1:1 mix of LB and 5% sucrose. Treatment was repeated once per week for 3 weeks, followed by a 1-week-recovery before dissection at 5 weeks. Proliferation response was assayed by phospho-histone 3 staining 24 hours after treatment in young 3-5 day-old flies.

X-ray induction

1 week old flies were placed in the X-ray generator CIXD and exposed to 40 Gray at the RadeXp facility at Institut Curie.

Immunofluorescence

Midgut fixation and immunofluorescence staining were performed as described previously described in [61]. Adult female midguts were dissected in PBS and then fixed at room temperature (RT) for 2 hours in 4% paraformaldehyde. Guts were trimmed and incubated in PBS 50% glycerol for 30 minutes before equilibration in PBS 0.1% Triton X-100 (PBT) to clean the lumen. Fixed and cleaned guts were then washed in PBT for at least 30

min before addition of primary antibodies (overnight at 4°C or 3-5 hours at RT). After at least 30 min wash, secondary antibodies were incubated 3-5 hours before DAPI staining (1mg/ml) and mounted in 4% N-propyl-galate, 80% glycerol.

The following primary antibodies were used: mouse anti-Delta extra-cellular domain (1/1000; DSHB C594.9b 10X concentrate), mouse anti-Pros (1/1000; DSHB MR1A-c 10X concentrate); chicken anti-GFP (1/2000, Invitrogen #A10262), rabbit anti-PH3, (1:1000; Millipore #06-570 lot 3746384), mouse anti-γH2Av (1:100; DSHB #UNC93-5.2.1 10X concentrate).

Imaging was performed using Zeiss LSM900 and LSM780 confocal microscopes and epifluorescence widefield microscope at the Curie Institute imaging facility with serial optical sections taken at 1 μm intervals (512X512 or 1024X1024) using 20X or 40X oil objectives through the whole-mounted posterior midguts.

Quantification

Quantification was carried out blind. Results quantified are pooled data from either 3 biological replicates in **Figures 2L, 6E** and **Supplementary Figure 5C, 6A, 6B** or pooled data of 4 biological replicates in **Figures 2E, 2G, 6B**.

LOH scoring

LOH events in females heterozygous for *Su(H)* were scored as clusters of at least 20 Delta and/or Prospero positive diploid cells.

Scoring of *Su(H)* LOH clones in *DI-Gal4* background using UAS-GFP

Experiments in **Figures 2G, Supplementary Figures 1B-C', 5C, 6B** the genetic background *Su(H)^{Δ47}/UAS-GFP; DI^{Gal4}/+* was used to drive stem cell specific expression in the ISC, due to the *DI^{Gal4}* driver. However, as *DI^{Gal4}* is a loss of function allele of *DI*, also a *Notch* signaling component, two types of neoplastic LOH clone could arise: those where *Su(H)*

is inactivated and those where *DI* undergoes LOH. We distinguished between these two possibilities by taking advantage of a UAS-GFP transgene located on Ch2L more distal on the chromosome arm to *Su(H)*. Therefore, LOH through recombination of *Su(H)* results in neoplastic clones that are GFP negative. These events were scored. In contrast, LOH resulting from other events including mitotic recombination of 3R leading to DI LOH, are GFP+. These events were scored, but not counted in the analysis. Statistical analysis for LOH clones was carried in out in Prism using Fisher's exact test (two-tailed) was performed and significant values were reported as: * $p < 0.05$, *** $p < 0.001$; **** $p < 0.0001$, ns=not significant.

PH3 quantification

PH3 positive cells in the midgut were counted using the epifluorescence microscope.

γ H2Av quantification

Images were acquired on the LSM900 confocal microscope. A maximum Z-projection was generated for all images on Image J (FIJI version 1.0). Only nuclear γ H2Av intensity in the ISCs was measured.

Sample collection for whole-genome sequencing

Su(H)^{Δ47/+}; ProsV1Gal4; UAS-nlsGFP were used to visually identify midguts containing LOH neoplasias. The region of the LOH neoplasia was manually microdissected. An estimate of 50-80% purity of the LOH neoplasia can be achieved. These neoplastic LOH tumors were dissected together with the fly head. Genomic DNA was isolated using QIAamp DNA MicroKit (Qiagen) according to the manufacturer's instructions.

Library preparation: Library preparation was performed with the Nextera XT kit and KAPA Hyper Plus ROCHE kit by the NGS facility of the Institut Curie. Samples were sequenced on one full flow cell (1600M clusters) on the NovaSeq in a paired-end 150bp mode.

Nextera XT Sample Prep Kit (Illumina) was used to prepare DNA sequencing libraries from 0.5 ng of genomics DNA. A step of enzymatic tagmentation with Nextera transposome was done to fragment DNA and add adapter sequences (Unique Dual Indexing strategy). A final amplification of the library was then performed with 12 cycles. After qPCR quantification (KAPA library quantification Kit, Roche), sequencing were performed on the Illumina NovaSeq 6000 using 2 x 150 cycles to get ~40 M paired-end reads per sample.

KAPA Hyper Plus Kits (Roche) was used to prepare DNA sequencing libraries from 1.3 ng of genomics DNA. A first step of enzymatic fragmentation of 20 minutes at 37°C was done. An end-repair and A-tailing step on dsDNA fragments have produced end-repaired dsDNA fragments. A ligation overnight with the adapters (Unique Dual Indexing strategy) and a final amplification of the library was then performed with 15 cycles. After qPCR quantification (KAPA library quantification Kit, Roche), sequencing were performed on the Illumina NovaSeq 6000 using 2 x 150 cycles to get ~40 M paired-end reads per sample.

Bioinformatics

Analysis of whole-genome sequencing data

Sequencing analysis of LOH samples was performed using custom scripts implemented in the Nextflow (Di Tommaso, 2017) pipeline nf-lohcator <https://github.com/nriddiford/nf-lohcator>.

Briefly, adapter sequences were removed using trimmomatic 0.39 [91]. Trimmed reads were aligned to release 6.12 of the *Drosophila melanogaster* reference genome (FlyBase) using bwa-mem version 0.7.17 (Li H, 2013, arXiv:1303.3997v2 [q-bio.GN]). Fastqc (Andrews, D. (2010)) FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>, bamstats (<https://biopet.github.io/bamstats/1.0.1/index.html>) and multiqc [92] were used to provide quality control files.

Alignment files were sorted and indexed using SAMtools [93] followed by the generation of pileup output summarizing the base calls of aligned reads to the reference sequence using mpileup command in SAMtools for each sample. VarScan2 version 2.4

(<https://github.com/dkoboldt/varscan/releases>) was then run on pileups generated. “Lohcator”, a custom python script <https://github.com/nriddiford/nf-lohcator/blob/master/bin/lohicator.py> developed in the lab was then ran to identify shifts in zygoty for each tumor sample and to create bed files to facilitate identification of the start and end of LOH for each sample.

VAF representation

VAF plots in **Figure 4B**, **Supplementary Figures 1A, 2A, 3A, 3B** were plotted with alleleFreqs: <https://github.com/nriddiford/alleleFreqs> using the LOH calls from VarScan2.

Rainfall plots

To generate rainfall plots in **Figure 6F** and **Supplementary Figure 6C**, filtering was carried out by using the “high quality” somatic calls produced by VarScan2.

We specified minimum coverage of 20 for both the tumour and normal sample and a somatic p-value<0.05. These calls were represented on rainfall plots. Some samples showed no SNVs and are thus not represented in the **Supplementary Figure 6C**.

Mapping of regions of recombination

First, on IGV, we confirmed that at this region, all the informative SNPs in the head are heterozygous and have gone homozygous in the tumour. Using a tool we have developed to identify regions of LOH in matched tumour normal pairs (nf-lohcator, see above), we were able to determine where the first homozygous SNP (from the centromere) is located, we then manually located the first heterozygous SNP on IGV relative to that for the mitotic recombination events. It is important to note that both coverage and tumour purity play an important role in confidently mapping these regions, with more emphasis placed on purity. We determined values for both coverage and purity for each mapped region (see **Supplementary Table 1**). We were unable to map the region of recombination for 3 samples (Samples F8, F10 and F26) because of EC contamination. These samples however allowed us to benchmark what is deemed “too impure” for the mapping of regions of recombination.

Association of mapped breakpoints with genomic regions

To assess whether mapped breakpoints were enriched for *The Histone Locus Cluster* (**Figure 5B**), and other genomic features (**Supplementary Figure 4A-D**), permutation tests were performed using regioneR [94] to determine the significance of the overlap between sequence features and our mapped breakpoint regions. In order to compare observed counts between real and shuffled data, we restricted permutations to within the genomic locus where mitotic recombination can be detected between the *Su(H)* locus and the centromere chr2L:15039488-23512838, and performed 10,000 permutations.

Calculating coverage

Mosdepth was used to calculate genome-wide sequencing coverage with default parameters (<https://github.com/brentp/mosdepth>).

Copy number analysis

Copy number analysis was performed using a read-depth-based approach with CNV-seq [95]. CNVPlotter <https://github.com/nriddiford/cnvPlotterR.git> was then used to generate copy number plots shown in **Supplementary Figure 2B** and **Supplementary Figure 3B'**.

Calculating EC contamination

The approach described in [60] and implemented in <https://github.com/nriddiford/winlow> was used to determine the likely contamination of tumor samples with EC cells (**Supplementary Table 2**).

Data Availability:

The data have been deposited with links to BioProject accession number PRJNA858414 in the NCBI BioProject database (<https://www.ncbi.nlm.nih.gov/bioproject/>).

Acknowledgements

610 The authors would like to thank members of the Bardin team, Sarah Lambert, Yohanns Bellaïche, Pierre-Antoine Defossez, Evan Dewey, and Jeff Sekelsky for helpful discussions and comments on the manuscript. Primary grant support for this work was from Worldwide Cancer Research (WCR 20-0147 grant to A.J.B.). Additional financial support came by grants from Fondation pour la Recherche Médicale (A.B., DEQ20160334928), as well as funding 615 from the program "Investissements d'Avenir" launched by the French Government and implemented by ANR, ANR SoMuSeq-STEM (A.B and N.S.), Labex DEEP (ANR-11-LBX-0044), IDEX PSL (ANR-10-IDEX-0001-02 PSL) and ICGex STEM-MITO-REC (A.B.). High-throughput sequencing has been performed by the ICGex NGS platform of the Institut Curie supported by the grants ANR-10-EQPX-03 (Equipex) and ANR-10-INBS-09-08 (France 620 Génomique Consortium) from the Agence Nationale de la Recherche "Investissements d'Avenir" program), by the Canceropole Ile-de-France and by the SiRIC-Curie program—SiRIC Grant "INCa-DGOS- 465". Salary support of A.J.B by CNRS; L.A-Z received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 666003 and the French foundation for medical 625 research (FRM, FDT201904008016).

Author Contributions

L.A. A.J.B. designed the study and analyzed the data. N. Riddiford developed software and performed most of the bioinformatic analysis, with N. Rubanova. performing additional 630 bioinformatic analyses supervised by N.S. L.A. and M.S. performed the experiments. M.B. from the ICGex NGS platform optimized low-input DNA sequencing. L.A. and A.J.B. wrote the manuscript with contribution from the other authors.

The authors declare no competing interest.

References

1. Al Zouabi L, Bardin AJ. Stem Cell DNA Damage and Genome Mutation in the Context of Aging and Cancer Initiation. *Cold Spring Harb Perspect Biol.* 2020. Epub 2020/01/15. doi: 10.1101/cshperspect.a036210. PubMed PMID: 31932318.
2. Behjati S, Huch M, van Boxtel R, Karthaus W, Wedge DC, Tamuri AU, et al. Genome sequencing of normal cells reveals developmental lineages and mutational processes. *Nature.* 2014. Epub 2014/07/22. doi: 10.1038/nature13448. PubMed PMID: 25043003.
3. Martincorena I, Roshan A, Gerstung M, Ellis P, Van Loo P, McLaren S, et al. Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science.* 2015;348(6237):880-6. Epub 2015/05/23. doi: 10.1126/science.aaa6806. PubMed PMID: 25999502.
4. Li R, Di L, Li J, Fan W, Liu Y, Guo W, et al. A body map of somatic mutagenesis in morphologically normal human tissues. *Nature.* 2021;597(7876):398-403. Epub 2021/08/27. doi: 10.1038/s41586-021-03836-1. PubMed PMID: 34433965.
5. Stern C. Somatic Crossing over and Segregation in *Drosophila Melanogaster*. *Genetics.* 1936;21(6):625-730. Epub 1936/11/01. PubMed PMID: 17246815; PubMed Central PMCID: PMC1208727.
6. Siudeja K, Bardin AJ. Somatic recombination in adult tissues: What is there to learn? *Fly (Austin).* 2017;11(2):121-8. Epub 2016/11/12. doi: 10.1080/19336934.2016.1249073. PubMed PMID: 27834607; PubMed Central PMCID: PMC5406163.
7. Jonkman MF, Scheffer H, Stulp R, Pas HH, Nijenhuis M, Heeres K, et al. Revertant mosaicism in epidermolysis bullosa caused by mitotic gene conversion. *Cell.* 1997;88(4):543-51. Epub 1997/02/21. PubMed PMID: 9038345.
8. Jongmans MC, Verwiel ET, Heijdra Y, Vulliamy T, Kamping EJ, Hehir-Kwa JY, et al. Revertant somatic mosaicism by mitotic recombination in dyskeratosis congenita. *American journal of human genetics.* 2012;90(3):426-33. Epub 2012/02/22. doi: 10.1016/j.ajhg.2012.01.004. PubMed PMID: 22341970; PubMed Central PMCID: PMC3309184.
9. Choate KA, Lu Y, Zhou J, Choi M, Elias PM, Farhi A, et al. Mitotic recombination in patients with ichthyosis causes reversion of dominant mutations in KRT10. *Science.* 2010;330(6000):94-7. Epub 2010/08/28. doi: 10.1126/science.1192280. PubMed PMID: 20798280; PubMed Central PMCID: PMC3085938.
10. Kiritsi D, He Y, Pasmooij AM, Onder M, Happel R, Jonkman MF, et al. Revertant mosaicism in a human skin fragility disorder results from slipped mispairing and mitotic recombination. *J Clin Invest.* 2012;122(5):1742-6. Epub 2012/04/03. doi: 10.1172/JCI61976. PubMed PMID: 22466645; PubMed Central PMCID: PMC3336993.
11. Venugopal P, Moore S, Lawrence DM, George AJ, Hannan RD, Bray SC, et al. Self-reverting mutations partially correct the blood phenotype in a Diamond Blackfan anemia patient. *Haematologica.* 2017;102(12):e506-e9. Epub 2017/10/04. doi: 10.3324/haematol.2017.166678. PubMed PMID: 28971907; PubMed Central PMCID: PMC5709125.
12. Jongmans MCJ, Diets IJ, Quarello P, Garelli E, Kuiper RP, Pfundt R. Somatic reversion events point towards RPL4 as a novel disease gene in a condition resembling Diamond-Blackfan anemia. *Haematologica.* 2018;103(12):e607-e9. Epub 2018/09/15. doi: 10.3324/haematol.2018.200683. PubMed PMID: 30213830; PubMed Central PMCID: PMC6269308.
13. Garelli E, Quarello P, Giorgio E, Carando A, Menegatti E, Mancini C, et al. Spontaneous remission in a Diamond-Blackfan anaemia patient due to a revertant uniparental disomy ablating a de novo RPS19 mutation. *Br J Haematol.* 2019;185(5):994-8. Epub 2018/11/22. doi: 10.1111/bjh.15688. PubMed PMID: 30460677.
14. Hirschhorn R, Yang DR, Puck JM, Huie ML, Jiang CK, Kurlandsky LE. Spontaneous in vivo reversion to normal of an inherited mutation in a patient with adenosine deaminase

- deficiency. *Nat Genet.* 1996;13(3):290-5. Epub 1996/07/01. doi: 10.1038/ng0796-290. PubMed PMID: 8673127.
15. Revy P, Kannengiesser C, Fischer A. Somatic genetic rescue in Mendelian haematopoietic diseases. *Nat Rev Genet.* 2019. Epub 2019/06/13. doi: 10.1038/s41576-019-0139-x. PubMed PMID: 31186537.
16. Howarth K, Ranta S, Winter E, Teixeira A, Schaschl H, Harvey JJ, et al. A mitotic recombination map proximal to the APC locus on chromosome 5q and assessment of influences on colorectal cancer risk. *BMC Med Genet.* 2009;10:54. Epub 2009/06/12. doi: 10.1186/1471-2350-10-54. PubMed PMID: 19515250; PubMed Central PMCID: PMC2705358.
17. Cottrell S, Bicknell D, Kaklamanis L, Bodmer WF. Molecular analysis of APC mutations in familial adenomatous polyposis and sporadic colon carcinomas. *Lancet.* 1992;340(8820):626-30. doi: 10.1016/0140-6736(92)92169-g. PubMed PMID: 1355210.
18. Cavenee WK, Dryja TP, Phillips RA, Benedict WF, Godbout R, Gallie BL, et al. Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. *Nature.* 1983;305(5937):779-84. Epub 1983/10/02. PubMed PMID: 6633649.
19. Serra E, Rosenbaum T, Nadal M, Winner U, Ars E, Estivill X, et al. Mitotic recombination effects homozygosity for NF1 germline mutations in neurofibromas. *Nat Genet.* 2001;28(3):294-6. doi: 10.1038/90148. PubMed PMID: 11431704.
20. Garcia-Linares C, Fernandez-Rodriguez J, Terribas E, Mercade J, Pros E, Benito L, et al. Dissecting loss of heterozygosity (LOH) in neurofibromatosis type 1-associated neurofibromas: Importance of copy neutral LOH. *Hum Mutat.* 2011;32(1):78-90. Epub 2010/10/30. doi: 10.1002/humu.21387. PubMed PMID: 21031597; PubMed Central PMCID: PMC3151547.
21. Nichols CA, Gibson WJ, Brown MS, Kosmicki JA, Busanovich JP, Wei H, et al. Loss of heterozygosity of essential genes represents a widespread class of potential cancer vulnerabilities. *Nat Commun.* 2020;11(1):2517. Epub 2020/05/21. doi: 10.1038/s41467-020-16399-y. PubMed PMID: 32433464; PubMed Central PMCID: PMC7239950.
22. Lee PS, Greenwell PW, Dominska M, Gawel M, Hamilton M, Petes TD. A fine-structure map of spontaneous mitotic crossovers in the yeast *Saccharomyces cerevisiae*. *PLoS genetics.* 2009;5(3):e1000410. Epub 2009/03/14. doi: 10.1371/journal.pgen.1000410. PubMed PMID: 19282969; PubMed Central PMCID: PMC2646836.
23. Symington LS, Rothstein R, Lisby M. Mechanisms and regulation of mitotic recombination in *Saccharomyces cerevisiae*. *Genetics.* 2014;198(3):795-835. Epub 2014/11/09. doi: 10.1534/genetics.114.166140. PubMed PMID: 25381364; PubMed Central PMCID: PMC4224172.
24. Lisby M, Rothstein R. Cell biology of mitotic recombination. *Cold Spring Harb Perspect Biol.* 2015;7(3):a016535. Epub 2015/03/04. doi: 10.1101/cshperspect.a016535. PubMed PMID: 25731763; PubMed Central PMCID: PMC4355273.
25. Jinks-Robertson S, Petes TD. Mitotic recombination in yeast: what we know and what we don't know. *Curr Opin Genet Dev.* 2021;71:78-85. Epub 2021/07/27. doi: 10.1016/j.gde.2021.07.002. PubMed PMID: 34311384; PubMed Central PMCID: PMC8671248.
26. Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW. The double-strand-break repair model for recombination. *Cell.* 1983;33(1):25-35. Epub 1983/05/01. PubMed PMID: 6380756.
27. Haber JE. DNA Repair: The Search for Homology. *Bioessays.* 2018;40(5):e1700229. Epub 2018/04/01. doi: 10.1002/bies.201700229. PubMed PMID: 29603285; PubMed Central PMCID: PMC6238635.
28. Wu X, Malkova A. Break-induced replication mechanisms in yeast and mammals. *Curr Opin Genet Dev.* 2021;71:163-70. Epub 2021/09/05. doi: 10.1016/j.gde.2021.08.002. PubMed PMID: 34481360.
29. Bizard AH, Hickson ID. The dissolution of double Holliday junctions. *Cold Spring Harb Perspect Biol.* 2014;6(7):a016477. Epub 2014/07/01. doi: 10.1101/cshperspect.a016477. PubMed PMID: 24984776; PubMed Central PMCID: PMC4067992.

30. West SC, Blanco MG, Chan YW, Matos J, Sarbajna S, Wyatt HD. Resolution of Recombination Intermediates: Mechanisms and Regulation. *Cold Spring Harb Symp Quant Biol.* 2015;80:103-9. Epub 20150914. doi: 10.1101/sqb.2015.80.027649. PubMed PMID: 26370409.
31. LaRocque JR, Stark JM, Oh J, Bojilova E, Yusa K, Horie K, et al. Interhomolog recombination and loss of heterozygosity in wild-type and Bloom syndrome helicase (BLM)-deficient mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America.* 2011;108(29):11971-6. Epub 2011/07/07. doi: 10.1073/pnas.1104421108. PubMed PMID: 21730139; PubMed Central PMCID: PMC3141969.
32. LaFave MC, Sekelsky J. Mitotic recombination: why? when? how? where? *PLoS genetics.* 2009;5(3):e1000411. Epub 2009/03/14. doi: 10.1371/journal.pgen.1000411. PubMed PMID: 19282976; PubMed Central PMCID: PMC2648873.
33. LaFave MC, Andersen SL, Stoffregen EP, Holsclaw JK, Kohl KP, Overton LJ, et al. Sources and structures of mitotic crossovers that arise when BLM helicase is absent in *Drosophila*. *Genetics.* 2014;196(1):107-18. Epub 2013/11/01. doi: 10.1534/genetics.113.158618. PubMed PMID: 24172129; PubMed Central PMCID: PMC3872178.
34. Carvajal-Garcia J, Crown KN, Ramsden DA, Sekelsky J. DNA polymerase theta suppresses mitotic crossing over. *PLoS Genet.* 2021;17(3):e1009267. Epub 2021/03/23. doi: 10.1371/journal.pgen.1009267. PubMed PMID: 33750946; PubMed Central PMCID: PMCPMC8016270.
35. Kuo HK, McMahan S, Rota CM, Kohl KP, Sekelsky J. *Drosophila* FANCM Helicase Prevents Spontaneous Mitotic Crossovers Generated by the MUS81 and SLX1 Nucleases. *Genetics.* 2014. Epub 2014/09/11. doi: 10.1534/genetics.114.168096. PubMed PMID: 25205745.
36. Moynahan ME, Jasin M. Loss of heterozygosity induced by a chromosomal double-strand break. *Proceedings of the National Academy of Sciences of the United States of America.* 1997;94(17):8988-93. Epub 1997/08/19. PubMed PMID: 9256422; PubMed Central PMCID: PMC22995.
37. Prakash R, Sandoval T, Morati F, Zagelbaum JA, Lim PX, White T, et al. Distinct pathways of homologous recombination controlled by the SWS1-SWSAP1-SPIDR complex. *Nat Commun.* 2021;12(1):4255. Epub 2021/07/14. doi: 10.1038/s41467-021-24205-6. PubMed PMID: 34253720; PubMed Central PMCID: PMCPMC8275761.
38. Boumard B, Bardin AJ. An amuse-bouche of stem cell regulation: Underlying principles and mechanisms from adult *Drosophila* intestinal stem cells. *Curr Opin Cell Biol.* 2021;73:58-68. Epub 2021/07/05. doi: 10.1016/j.ceb.2021.05.007. PubMed PMID: 34217969.
39. Micchelli CA, Perrimon N. Evidence that stem cells reside in the adult *Drosophila* midgut epithelium. *Nature.* 2006;439(7075):475-9. Epub 2005/12/13. doi: nature04371 [pii] 10.1038/nature04371. PubMed PMID: 16340959.
40. Ohlstein B, Spradling A. The adult *Drosophila* posterior midgut is maintained by pluripotent stem cells. *Nature.* 2006;439(7075):470-4. Epub 2005/12/13. doi: 10.1038/nature04333. PubMed PMID: 16340960.
41. Apidianakis Y, Pitsouli C, Perrimon N, Rahme L. Synergy between bacterial infection and genetic predisposition in intestinal dysplasia. *Proceedings of the National Academy of Sciences of the United States of America.* 2009;106(49):20883-8. Epub 2009/11/26. doi: 10.1073/pnas.0911797106. PubMed PMID: 19934041; PubMed Central PMCID: PMC2791635.
42. Buchon N, Broderick NA, Poidevin M, Pradervand S, Lemaitre B. *Drosophila* intestinal response to bacterial infection: activation of host defense and stem cell proliferation. *Cell Host Microbe.* 2009;5(2):200-11. Epub 2009/02/17. doi: S1931-3128(09)00028-6 [pii] 10.1016/j.chom.2009.01.003. PubMed PMID: 19218090.

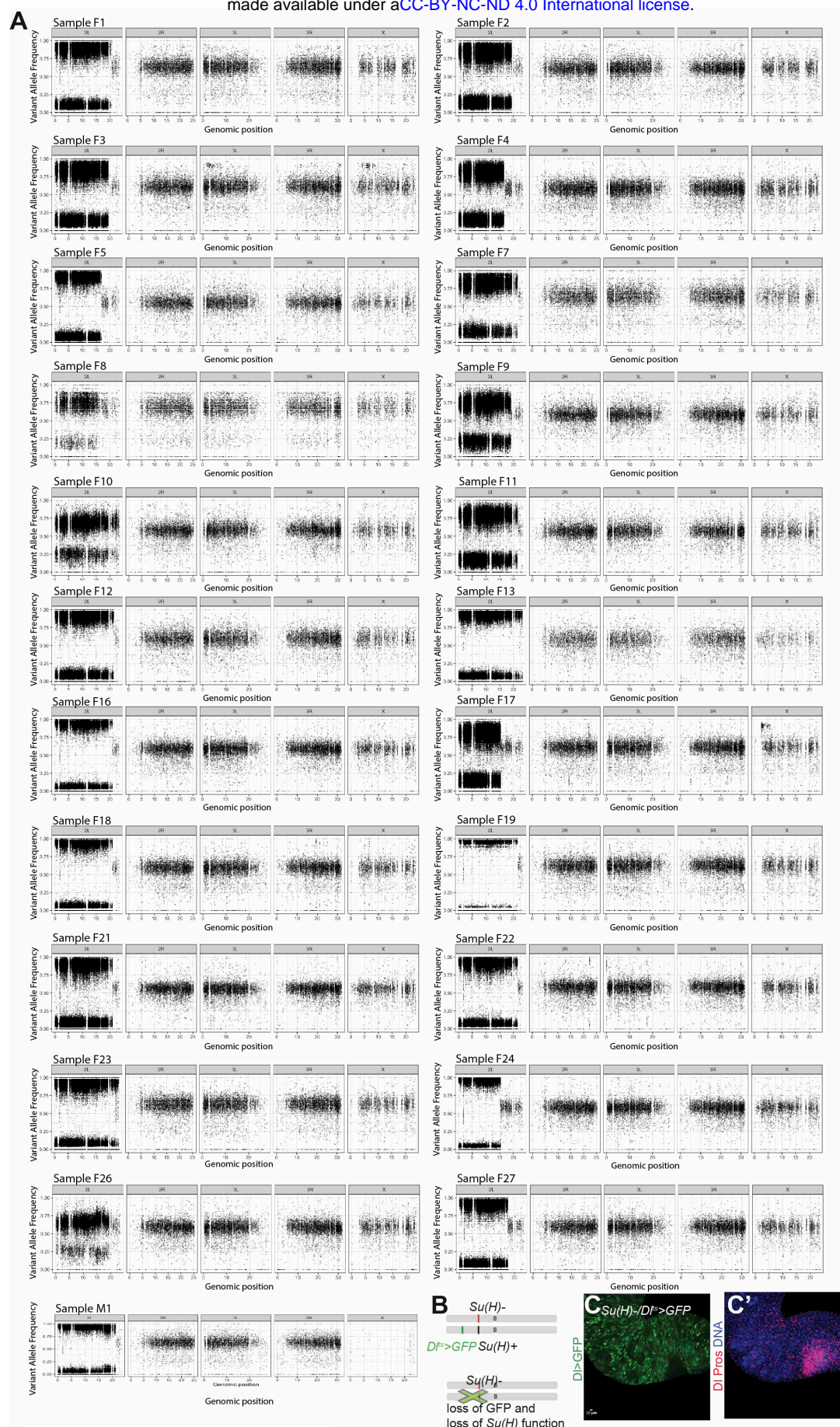
43. Jiang H, Patel PH, Kohlmaier A, Grenley MO, McEwen DG, Edgar BA. Cytokine/Jak/Stat signaling mediates regeneration and homeostasis in the *Drosophila* midgut. *Cell*. 2009;137(7):1343-55. Epub 2009/07/01. doi: S0092-8674(09)00571-6 [pii] 10.1016/j.cell.2009.05.014. PubMed PMID: 19563763.
44. Biteau B, Hochmuth CE, Jasper H. JNK activity in somatic stem cells causes loss of tissue homeostasis in the aging *Drosophila* gut. *Cell Stem Cell*. 2008;3(4):442-55. Epub 2008/10/23. doi: S1934-5909(08)00392-5 [pii] 10.1016/j.stem.2008.07.024. PubMed PMID: 18940735.
45. Amcheslavsky A, Jiang J, Ip YT. Tissue damage-induced intestinal stem cell division in *Drosophila*. *Cell Stem Cell*. 2009;4(1):49-61. Epub 2009/01/09. doi: S1934-5909(08)00570-5 [pii] 10.1016/j.stem.2008.10.016. PubMed PMID: 19128792.
46. Ferguson M, Petkau K, Shin M, Galenza A, Fast D, Foley E. Differential effects of commensal bacteria on progenitor cell adhesion, division symmetry and tumorigenesis in the *Drosophila* intestine. *Development*. 2021;148(5). Epub 2021/02/18. doi: 10.1242/dev.186106. PubMed PMID: 33593820.
47. Zhou J, Boutros M. JNK-dependent intestinal barrier failure disrupts host-microbe homeostasis during tumorigenesis. *Proc Natl Acad Sci U S A*. 2020;117(17):9401-12. Epub 2020/04/12. doi: 10.1073/pnas.1913976117. PubMed PMID: 32277031; PubMed Central PMCID: PMC7196803.
48. Campbell K, Rossi F, Adams J, Pitsidianaki I, Barriga FM, Garcia-Gerique L, et al. Collective cell migration and metastases induced by an epithelial-to-mesenchymal transition in *Drosophila* intestinal tumors. *Nat Commun*. 2019;10(1):2311. Epub 2019/05/24. doi: 10.1038/s41467-019-10269-y. PubMed PMID: 31127094; PubMed Central PMCID: PMC6534551.
49. Tamamouna V, Rahman MM, Petersson M, Charalambous I, Kux K, Mainor H, et al. Remodelling of oxygen-transporting tracheoles drives intestinal regeneration and tumorigenesis in *Drosophila*. *Nat Cell Biol*. 2021;23(5):497-510. Epub 2021/05/12. doi: 10.1038/s41556-021-00674-1. PubMed PMID: 33972730.
50. Singh SR, Zeng X, Zhao J, Liu Y, Hou G, Liu H, et al. The lipolysis pathway sustains normal and transformed stem cells in adult *Drosophila*. *Nature*. 2016;538(7623):109-13. Epub 2016/09/30. doi: 10.1038/nature19788. PubMed PMID: 27680705.
51. Bensard CL, Wisidagama DR, Olson KA, Berg JA, Krah NM, Schell JC, et al. Regulation of Tumor Initiation by the Mitochondrial Pyruvate Carrier. *Cell Metab*. 2020;31(2):284-300 e7. Epub 2019/12/10. doi: 10.1016/j.cmet.2019.11.002. PubMed PMID: 31813825; PubMed Central PMCID: PMC7004878.
52. Cordero JB, Stefanatos RK, Myant K, Vidal M, Sansom OJ. Non-autonomous crosstalk between the Jak/Stat and Egfr pathways mediates Apc1-driven intestinal stem cell hyperplasia in the *Drosophila* adult midgut. *Development*. 2012;139(24):4524-35. Epub 2012/11/23. doi: 10.1242/dev.078261. PubMed PMID: 23172913.
53. Suijkerbuijk SJ, Kolahgar G, Kucinski I, Piddini E. Cell Competition Drives the Growth of Intestinal Adenomas in *Drosophila*. *Current biology : CB*. 2016;26(4):428-38. Epub 2016/02/09. doi: 10.1016/j.cub.2015.12.043. PubMed PMID: 26853366; PubMed Central PMCID: PMC4771920.
54. Zhai Z, Kondo S, Ha N, Boquete JP, Brunner M, Ueda R, et al. Accumulation of differentiating intestinal stem cell progenies drives tumorigenesis. *Nat Commun*. 2015;6:10219. Epub 2015/12/23. doi: 10.1038/ncomms10219. PubMed PMID: 26690827; PubMed Central PMCID: PMC4703904.
55. Ngo S, Liang J, Su YH, O'Brien LE. Disruption of EGF Feedback by Intestinal Tumors and Neighboring Cells in *Drosophila*. *Curr Biol*. 2020;30(8):1537-46 e3. Epub 2020/04/04. doi: 10.1016/j.cub.2020.01.082. PubMed PMID: 32243854; PubMed Central PMCID: PMC7409949.
56. Patel PH, Dutta D, Edgar BA. Niche appropriation by *Drosophila* intestinal stem cell tumours. *Nature cell biology*. 2015;17(9):1182-92. Epub 2015/08/04. doi: 10.1038/ncb3214. PubMed PMID: 26237646.

57. Zhou J, Valentini E, Boutros M. Microenvironmental innate immune signaling and cell mechanical responses promote tumor growth. *Dev Cell*. 2021;56(13):1884-99 e5. Epub 2021/07/02. doi: 10.1016/j.devcel.2021.06.007. PubMed PMID: 34197724.
58. Siudeja K, Nassari S, Gervais L, Skorski P, Lameiras S, Stolfa D, et al. Frequent Somatic Mutation in Adult Intestinal Stem Cells Drives Neoplasia and Genetic Mosaicism during Aging. *Cell stem cell*. 2015;17(6):663-74. Epub 2015/11/27. doi: 10.1016/j.stem.2015.09.016. PubMed PMID: 26607382.
59. Siudeja K, van den Beek M, Riddiford N, Boumard B, Wurmser A, Stefanutti M, et al. Unraveling the features of somatic transposition in the *Drosophila* intestine. *EMBO J*. 2021;40(9):e106388. Epub 2021/02/27. doi: 10.15252/embj.2020106388. PubMed PMID: 33634906; PubMed Central PMCID: PMCPCMC8090852.
60. Riddiford N, Siudeja K, van den Beek M, Boumard B, Bardin AJ. Evolution and genomic signatures of spontaneous somatic mutation in *Drosophila* intestinal stem cells. *Genome Res*. 2021;31(8):1419-32. Epub 2021/06/26. doi: 10.1101/gr.268441.120. PubMed PMID: 34168010; PubMed Central PMCID: PMCPCMC8327918.
61. Bardin AJ, Perdigoto CN, Southall TD, Brand AH, Schweisguth F. Transcriptional control of stem cell maintenance in the *Drosophila* intestine. *Development*. 2010;137:715-24.
62. Gervais L, Bardin AJ. Tissue homeostasis and aging: new insight from the fly intestine. *Curr Opin Cell Biol*. 2017;48:97-105. Epub 2017/07/19. doi: 10.1016/j.ceb.2017.06.005. PubMed PMID: 28719867.
63. Lane ME, Sauer K, Wallace K, Jan YN, Lehner CF, Vaessin H. Dacapo, a cyclin-dependent kinase inhibitor, stops cell proliferation during *Drosophila* development. *Cell*. 1996;87(7):1225-35. Epub 1996/12/27. doi: 10.1016/s0092-8674(00)81818-8. PubMed PMID: 8980229.
64. Kohlmaier A, Fassnacht C, Jin Y, Reuter H, Begum J, Dutta D, et al. Src kinase function controls progenitor cell pools during regeneration and tumor onset in the *Drosophila* intestine. *Oncogene*. 2015;34(18):2371-84. Epub 2014/07/01. doi: 10.1038/onc.2014.163. PubMed PMID: 24975577.
65. Lucchetta EM, Ohlstein B. Amitosis of Polyploid Cells Regenerates Functional Stem Cells in the *Drosophila* Intestine. *Cell Stem Cell*. 2017;20(5):609-20 e6. Epub 2017/03/28. doi: 10.1016/j.stem.2017.02.012. PubMed PMID: 28343984; PubMed Central PMCID: PMCPCMC5419863.
66. Alecki C, Chiwara V, Sanz LA, Grau D, Arias Perez O, Boulter EL, et al. RNA-DNA strand exchange by the *Drosophila* Polycomb complex PRC2. *Nat Commun*. 2020;11(1):1781. Epub 2020/04/14. doi: 10.1038/s41467-020-15609-x. PubMed PMID: 32286294; PubMed Central PMCID: PMCPCMC7156742.
67. Marzluff WF, Koreski KP. Birth and Death of Histone mRNAs. *Trends Genet*. 2017;33(10):745-59. Epub 2017/09/05. doi: 10.1016/j.tig.2017.07.014. PubMed PMID: 28867047; PubMed Central PMCID: PMCPCMC5645032.
68. Ho CK, Mazon G, Lam AF, Symington LS. Mus81 and Yen1 promote reciprocal exchange during mitotic recombination to maintain genome integrity in budding yeast. *Mol Cell*. 2010;40(6):988-1000. Epub 2010/12/22. doi: 10.1016/j.molcel.2010.11.016. PubMed PMID: 21172663; PubMed Central PMCID: PMCPCMC3021384.
69. Wilson MA, Kwon Y, Xu Y, Chung WH, Chi P, Niu H, et al. Pif1 helicase and Poldelta promote recombination-coupled DNA synthesis via bubble migration. *Nature*. 2013;502(7471):393-6. Epub 2013/09/13. doi: 10.1038/nature12585. PubMed PMID: 24025768; PubMed Central PMCID: PMCPCMC3915060.
70. Kocak E, Dykstra S, Nemeth A, Coughlin CG, Rodgers K, McVey M. The *Drosophila melanogaster* PIF1 Helicase Promotes Survival During Replication Stress and Processive DNA Synthesis During Double-Strand Gap Repair. *Genetics*. 2019;213(3):835-47. Epub 2019/09/21. doi: 10.1534/genetics.119.302665. PubMed PMID: 31537623; PubMed Central PMCID: PMCPCMC6827366.
71. Bhandari J, Karg T, Golic KG. Homolog-Dependent Repair Following Dicentric Chromosome Breakage in *Drosophila melanogaster*. *Genetics*. 2019;212(3):615-30.

- Epub 2019/05/06. doi: 10.1534/genetics.119.302247. PubMed PMID: 31053594; PubMed Central PMCID: PMC6614899.
72. Li S, Wang H, Jehi S, Li J, Liu S, Wang Z, et al. PIF1 helicase promotes break-induced replication in mammalian cells. *EMBO J.* 2021;40(8):e104509. Epub 2021/01/21. doi: 10.15252/embj.2020104509. PubMed PMID: 33470420; PubMed Central PMCID: PMC68047440.
73. Smith CE, Llorente B, Symington LS. Template switching during break-induced replication. *Nature.* 2007;447(7140):102-5. Epub 2007/04/06. doi: 10.1038/nature05723. PubMed PMID: 17410126.
74. Deem A, Keszthelyi A, Blackgrove T, Vayl A, Coffey B, Mathur R, et al. Break-induced replication is highly inaccurate. *PLoS Biol.* 2011;9(2):e1000594. Epub 2011/02/25. doi: 10.1371/journal.pbio.1000594. PubMed PMID: 21347245; PubMed Central PMCID: PMC3039667.
75. Chakrabarti S, Liehl P, Buchon N, Lemaitre B. Infection-induced host translational blockage inhibits immune responses and epithelial renewal in the *Drosophila* gut. *Cell host & microbe.* 2012;12(1):60-70. Epub 2012/07/24. doi: 10.1016/j.chom.2012.06.001. PubMed PMID: 22817988.
76. Li J, Zhang AH, Wu FF, Wang XJ. Alterations in the Gut Microbiota and Their Metabolites in Colorectal Cancer: Recent Progress and Future Prospects. *Front Oncol.* 2022;12:841552. Epub 2022/03/01. doi: 10.3389/fonc.2022.841552. PubMed PMID: 35223525; PubMed Central PMCID: PMC8875205.
77. Gagnaire A, Nadel B, Raoult D, Neefjes J, Gorvel JP. Collateral damage: insights into bacterial mechanisms that predispose host cells to cancer. *Nat Rev Microbiol.* 2017;15(2):109-28. Epub 2017/01/04. doi: 10.1038/nrmicro.2016.171. PubMed PMID: 28045107.
78. Arthur JC. Microbiota and colorectal cancer: colibactin makes its mark. *Nat Rev Gastroenterol Hepatol.* 2020;17(6):317-8. Epub 2020/04/23. doi: 10.1038/s41575-020-0303-y. PubMed PMID: 32317778.
79. Arthur JC, Perez-Chanona E, Muhlbauer M, Tomkovich S, Uronis JM, Fan TJ, et al. Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science.* 2012;338(6103):120-3. Epub 2012/08/21. doi: 10.1126/science.1224820. PubMed PMID: 22903521; PubMed Central PMCID: PMC3645302.
80. Pleguezuelos-Manzano C, Puschihof J, Rosendahl Huber A, van Hoeck A, Wood HM, Nomburg J, et al. Mutational signature in colorectal cancer caused by genotoxic pks(+) *E. coli*. *Nature.* 2020;580(7802):269-73. Epub 2020/02/28. doi: 10.1038/s41586-020-2080-8. PubMed PMID: 32106218; PubMed Central PMCID: PMC68142898.
81. St Charles J, Petes TD. High-resolution mapping of spontaneous mitotic recombination hotspots on the 1.1 Mb arm of yeast chromosome IV. *PLoS genetics.* 2013;9(4):e1003434. Epub 2013/04/18. doi: 10.1371/journal.pgen.1003434. PubMed PMID: 23593029; PubMed Central PMCID: PMC3616911.
82. Yin Y, Petes TD. Genome-wide high-resolution mapping of UV-induced mitotic recombination events in *Saccharomyces cerevisiae*. *PLoS Genet.* 2013;9(10):e1003894. Epub 2013/11/10. doi: 10.1371/journal.pgen.1003894. PubMed PMID: 24204306; PubMed Central PMCID: PMC3814309.
83. Yin Y, Dominska M, Yim E, Petes TD. High-resolution mapping of heteroduplex DNA formed during UV-induced and spontaneous mitotic recombination events in yeast. *Elife.* 2017;6. Epub 2017/07/18. doi: 10.7554/eLife.28069. PubMed PMID: 28714850; PubMed Central PMCID: PMC5531827.
84. Sui Y, Qi L, Wu JK, Wen XP, Tang XX, Ma ZJ, et al. Genome-wide mapping of spontaneous genetic alterations in diploid yeast cells. *Proc Natl Acad Sci U S A.* 2020;117(45):28191-200. Epub 2020/10/28. doi: 10.1073/pnas.2018633117. PubMed PMID: 33106417; PubMed Central PMCID: PMC7668089.
85. Lu S, Wang G, Bacolla A, Zhao J, Spitser S, Vasquez KM. Short Inverted Repeats Are Hotspots for Genetic Instability: Relevance to Cancer Genomes. *Cell Rep.*

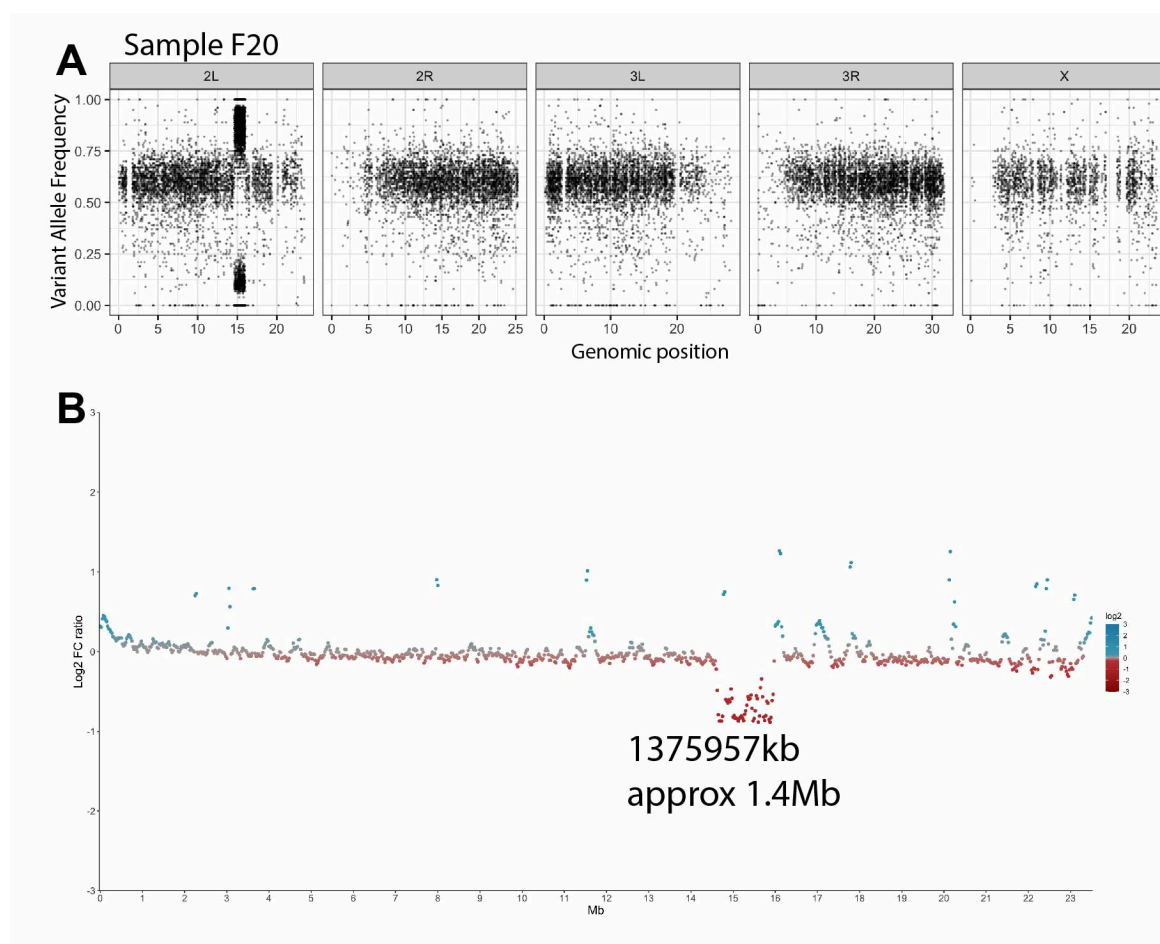
- 2015;10(10):1674-80. Epub 20150312. doi: 10.1016/j.celrep.2015.02.039. PubMed PMID: 25772355; PubMed Central PMCID: PMC6013304.
86. Boulianne B, Robinson ME, May PC, Castellano L, Blighe K, Thomas J, et al. Lineage-Specific Genes Are Prominent DNA Damage Hotspots during Leukemic Transformation of B Cell Precursors. *Cell Rep.* 2017;18(7):1687-98. doi: 10.1016/j.celrep.2017.01.057. PubMed PMID: 28199841; PubMed Central PMCID: PMC65318656.
87. Bulankova P, Sekulic M, Jallet D, Nef C, van Oosterhout C, Delmont TO, et al. Mitotic recombination between homologous chromosomes drives genomic diversity in diatoms. *Curr Biol.* 2021;31(15):3221-32 e9. Epub 2021/06/09. doi: 10.1016/j.cub.2021.05.013. PubMed PMID: 34102110.
88. Jia X, Zhang Q, Jiang M, Huang J, Yu L, Traw MB, et al. Mitotic gene conversion can be as important as meiotic conversion in driving genetic variability in plants and other species without early germline segregation. *PLoS Biol.* 2021;19(3):e3001164. Epub 20210322. doi: 10.1371/journal.pbio.3001164. PubMed PMID: 33750968; PubMed Central PMCID: PMC8016264.
89. Kurahashi H, Inagaki H, Yamada K, Ohye T, Taniguchi M, Emanuel BS, et al. Cruciform DNA structure underlies the etiology for palindrome-mediated human chromosomal translocations. *J Biol Chem.* 2004;279(34):35377-83. Epub 20040620. doi: 10.1074/jbc.M400354200. PubMed PMID: 15208332; PubMed Central PMCID: PMC602810964.
90. Paeschke K, Capra JA, Zakian VA. DNA replication through G-quadruplex motifs is promoted by the *Saccharomyces cerevisiae* Pif1 DNA helicase. *Cell.* 2011;145(5):678-91. doi: 10.1016/j.cell.2011.04.015. PubMed PMID: 21620135; PubMed Central PMCID: PMC3129610.
91. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114-20. Epub 20140401. doi: 10.1093/bioinformatics/btu170. PubMed PMID: 24695404; PubMed Central PMCID: PMC4103590.
92. Ewels P, Magnusson M, Lundin S, Kaller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics.* 2016;32(19):3047-8. Epub 20160616. doi: 10.1093/bioinformatics/btw354. PubMed PMID: 27312411; PubMed Central PMCID: PMC5039924.
93. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *Gigascience.* 2021;10(2). doi: 10.1093/gigascience/giab008. PubMed PMID: 33590861; PubMed Central PMCID: PMC7931819.
94. Gel B, Diez-Villanueva A, Serra E, Buschbeck M, Peinado MA, Malinverni R. regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics.* 2016;32(2):289-91. Epub 20150930. doi: 10.1093/bioinformatics/btv562. PubMed PMID: 26424858; PubMed Central PMCID: PMC4708104.
95. Xie C, Tammi MT. CNV-seq, a new method to detect copy number variation using high-throughput sequencing. *BMC Bioinformatics.* 2009;10:80. Epub 20090306. doi: 10.1186/1471-2105-10-80. PubMed PMID: 19267900; PubMed Central PMCID: PMC2667514.

Supplementary Figures



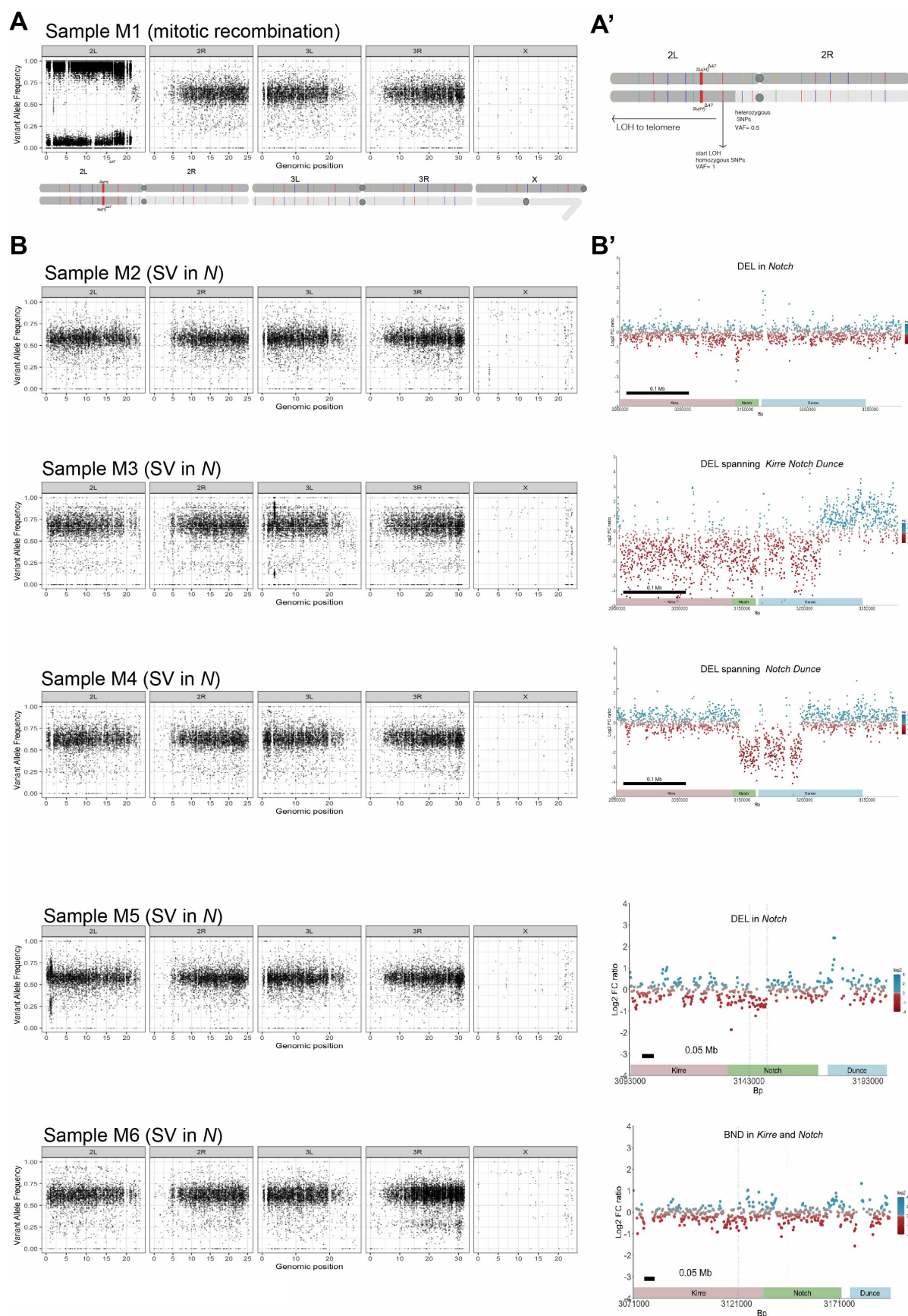
Supplementary Figure 1: Evidence for LOH driven by Mitotic Recombination on chromosome 2L

(A) Variant allele frequency plots of samples supporting LOH of chromosome 2L, with SNPs becoming homozygous (VAF>0.75). Samples F6, F14, F15 and F25 were excluded because of too low coverage. **(B-C')** The genetic background *Su(H)^{Δ47}/UAS-GFP; D^{Gal4}/+* was used to drive stem cell specific expression in the ISC, due to the *D^{Gal4}* driver. Neoplastic LOH clones where *Su(H)* is inactivated resulted in GFP negative clones due to the loss of the UAS-GFP transgene located on Ch2L more distal on the chromosome arm to *Su(H)* via mitotic recombination (GFP in green, DI and Pros in red, DAPI marking DNA in blue).



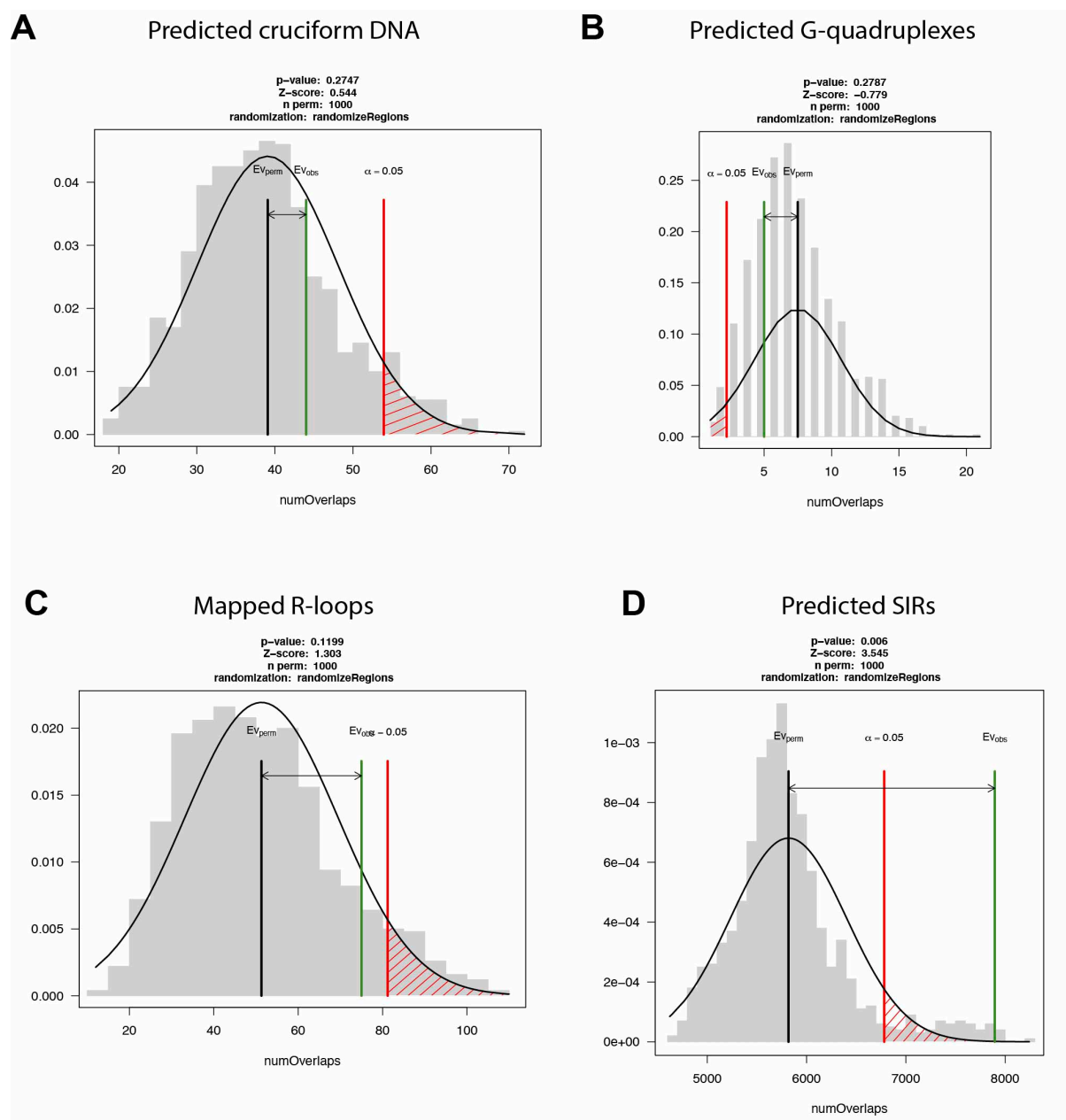
Supplementary Figure 2: Sample showing a deletion spanning *Su(H)* locus leading to LOH

(A) VAF plot of sample F20 showing a shift in allele frequency on Chr 2L where there is a deletion. (B) Copy number plot showing a large drop in coverage (approx 1.4Mb) in the tumor relative to the head control.



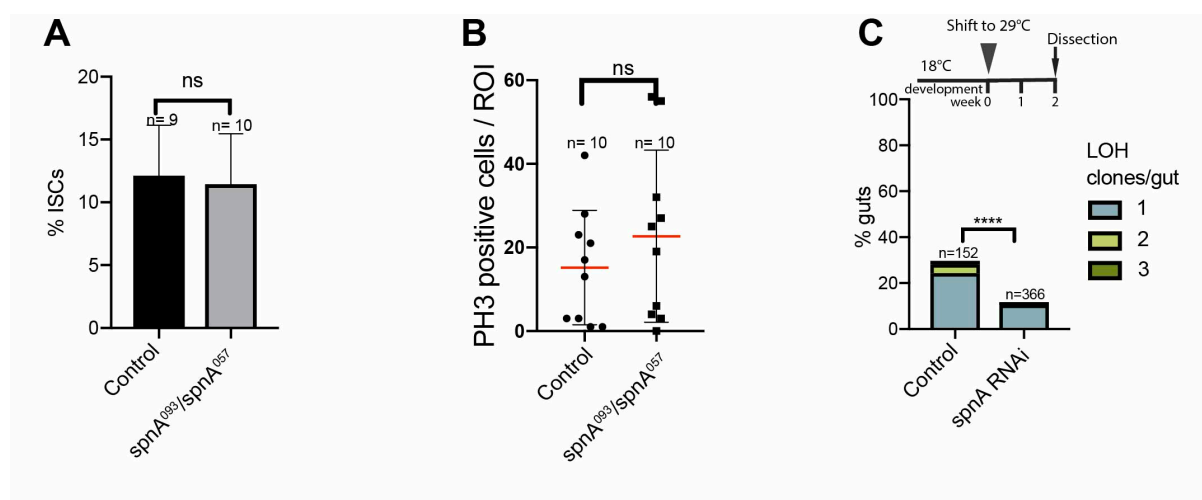
Supplementary Figure 3: VAF plots of all the male samples

*Plots of all male samples, showing all chromosomes. Only sample M1 (**A-A'**) shows that chromosome 2L has undergone an LOH event, with SNPs becoming homozygous (VAF >0.75). The remaining samples (**B-B'**) show no LOH on chromosome 2L (left panels) but rather show a structural variants (SV) spanning the Notch locus on the X chromosome (right panels). DEL is deletion, BND indicates “break-end class” of genomic aberration,.*



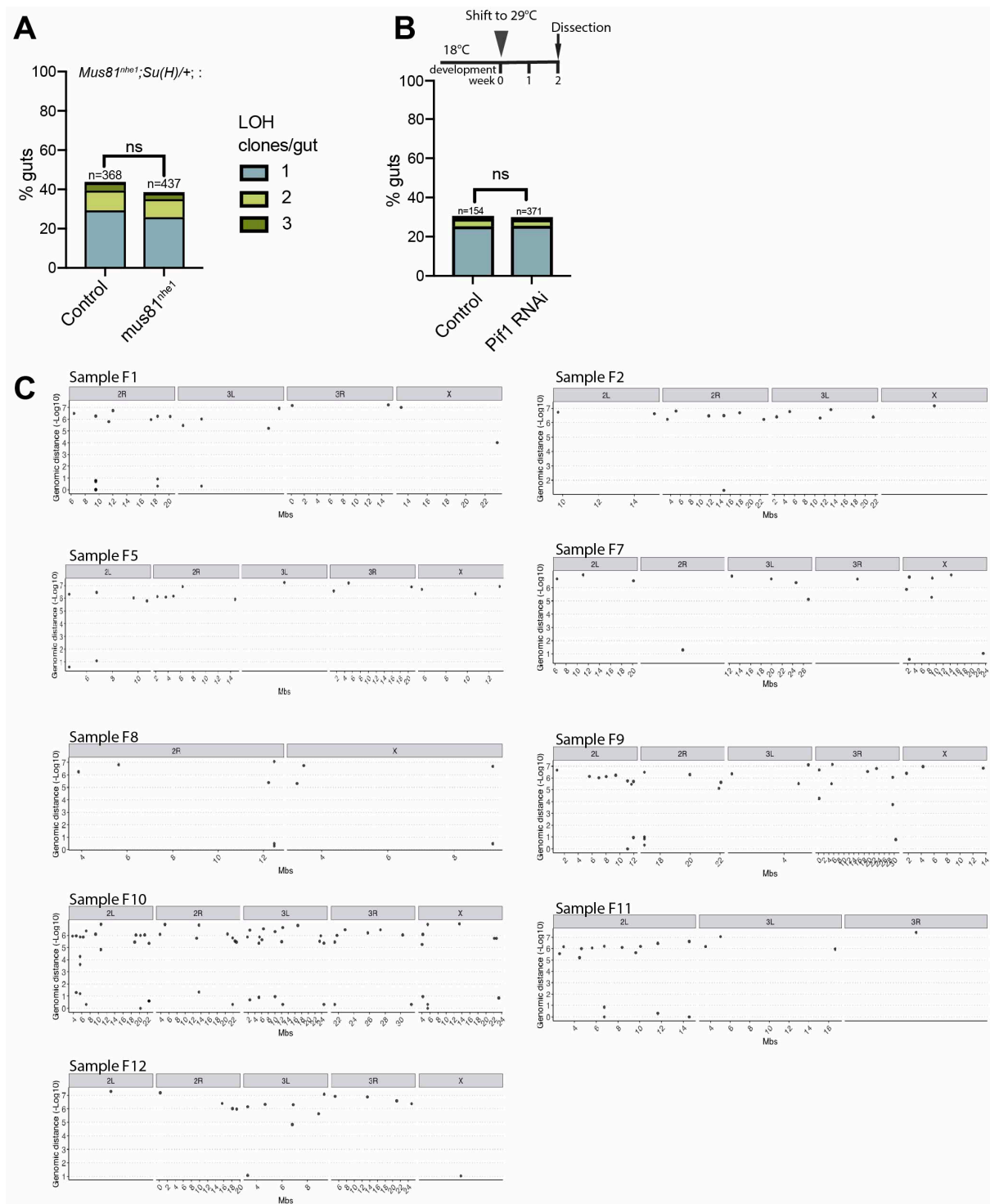
Supplementary Figure 4: Mapped recombination sites showed no significant overlap with non-B-form DNA such as cruciform DNA, G-quadruplexes and R-loops

Permutation test carried out using RegioneR. Test association between mapped recombination sites and (A) predicted cruciform DNA [89], (B) predicted G-quadruplexes [90]. (C) mapped R-loops [66], (D) predicted SIRs [85]. Shuffles: 1000 constricted to the region between Su(H) and the centromere where recombination can be detected.



Supplementary Figure 5: Evidence for *SpnA* promoting loss of heterozygosity

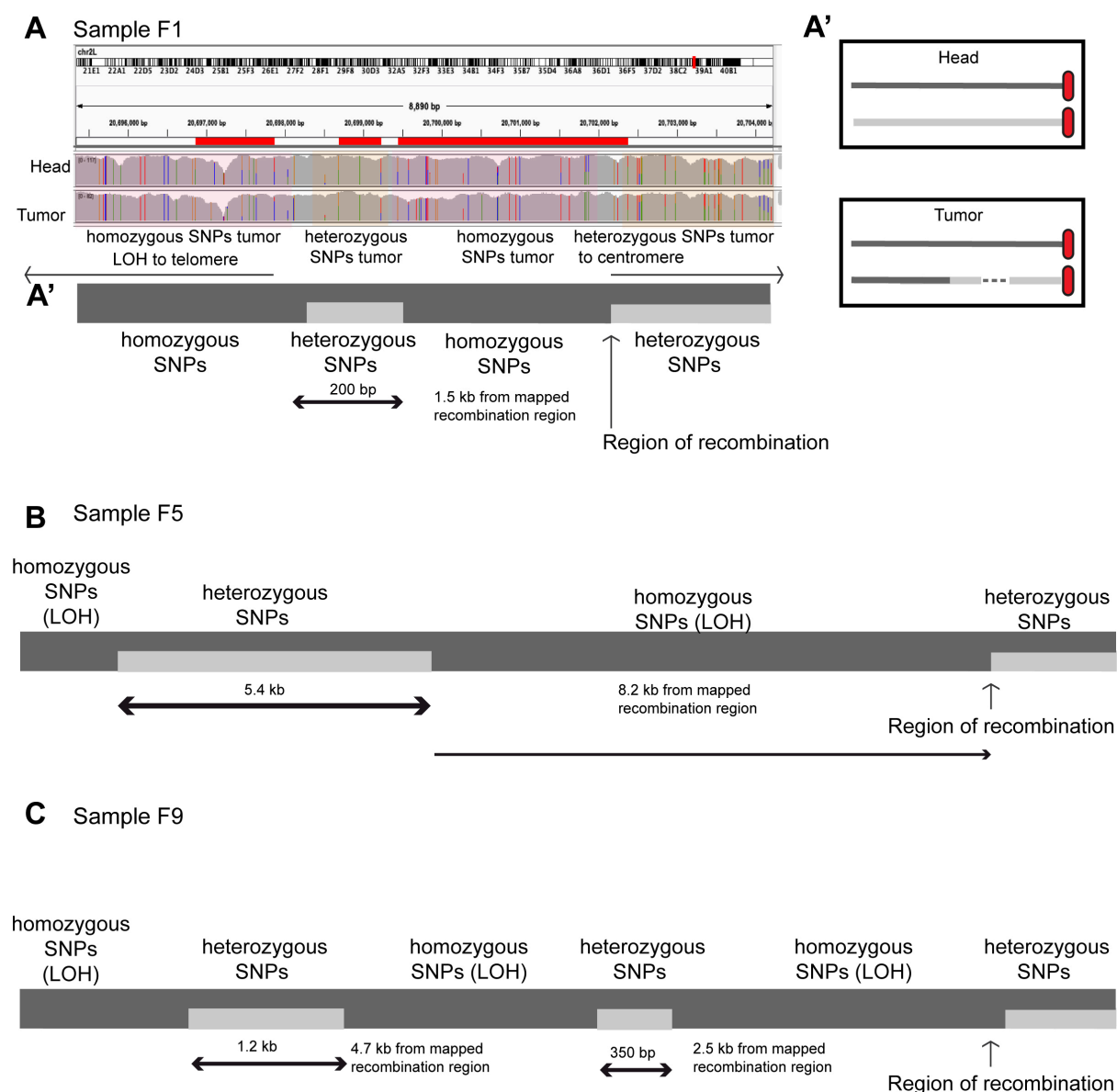
(A) No significant difference in %ISCs per region of interest (ROI) between control guts and *Su(H)^{A47/+}; SpnA⁰⁹³/SpnA⁰⁵⁷* (T-test with Welch's correction). **(B)** No significant difference in mitotic cells per ROI between control guts and *Su(H)^{A47/+}; SpnA⁰⁹³/SpnA⁰⁵⁷* (T-test with Welch's correction). **(C)** Frequency of LOH clones in control compared to *Rad51* knockdown in ISCs. Adult flies were shifted to 29°C to induce RNAi for 2 weeks then dissected. (Genotypes: *SpnA RNAi; Su(H)^{A47}/ UAS-GFP; Df^{Gal4} tubGAL80^{ts} /+* Vs *Su(H)^{A47}/ UAS-GFP; Df^{Gal4} tubGAL80^{ts} /+* control) (Fisher's exact test, two-tailed).



Supplementary Figure 6: Genetic and genomic data arguing against a BIR model.

(A) Frequency of LOH clones in *Mus81^{nhe1}; Su(H)/+* compared with *+/+; Su(H)/+* control (Fisher's exact test, two-tailed). (B) Frequency of LOH clones in control compared to *Pif1* knockdown in ISCs. RNAi was induced in adult

ISCs for 2 weeks prior to dissection. Genotypes : Su(H)^{A47}/ UAS-GFP; D^{Gal4} tubGAL80^{ts} / UAS-Pif1 RNAi Vs Su(H)^{A47}/ UAS-GFP; D^{Gal4} tubGAL80^{ts} /+ control. Fisher's exact tests (ns). (C) Rainfall plots of point mutations in the sequenced LOH samples. Each point represents the genomic distance between two SNVs (point mutations). A mutational pileup would be detected by an accumulation of points vertically. We observed no mutational hotspots on chromosome 2L, arguing against a BIR model. Samples not shown are samples which had no SNVs.



Supplementary Figure 7: Schematic showing samples with conversion tracts

(A) IGV view showing SNP evidence of a conversion tract 1.5kb away from the mapped region of recombination. While the control head sample representing the germline (top IGV tract) shows all heterozygous SNPs, the bottom LOH tumor sample shows a shift from heterozygous \rightarrow homozygous SNPs, then goes from homozygous \rightarrow heterozygous, before going back from homozygous \rightarrow heterozygous throughout the chromosome arm. A schematic of the cell of origin in head and tumor is shown on the right panel A'. (A') Schematic of sample F1 conversion tract denoted by shift from homozygous to heterozygous SNPs in the tumour sample for ~200 bp before again becoming homozygous until the telomere. (B) Schematic of sample F5: After the initial recombination site (indicated by arrow) there is a shift from heterozygous \rightarrow homozygous SNPs in the tumor, a region

approximately 8.2kb away showed a DNA tract marked by a shift from homo-het SNPs for ~5.4kb before again becoming homozygous until the telomere. (C) Schematic of sample F9 where 2 tracts were identified having heterozygous SNPs within the larger LOH region. Approximate length of each segmented is noted.