Mechanisms of spontaneous loss of heterozygosity in adult stem cells

Lara Al-Zoubi

▶ **To cite this version:**

Lara Al-Zoubi. Mechanisms of spontaneous loss of heterozygosity in adult stem cells. Cellular Biology. Université Paris sciences et lettres, 2020. English. NNT : 2020UPSLT017 . tel-03598037

HAL Id: tel-03598037
https://theses.hal.science/tel-03598037

Submitted on 4 Mar 2022

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

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

PSL UNIVERSITÉ PARIS

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à **l'institut Curie**.

Mécanismes de la perte spontanée d'hétérozygotie dans les cellules souches intestinales

Mechanisms of Spontaneous Loss of Heterozygosity in Intestinal Stem Cells

Soutenue par
Lara Al-ZOUBI
Le 19 Novembre 2020

École doctorale n° 515
Complexité du Vivant

Spécialité
Biologie cellulaire

Composition du jury :

Dr. Valérie, Borde Directrice de recherche, Institut Curie – CNRS	*Président*
Dr. Jean-René, Martin Directeur de recherche, Institut NeuroPSI – CNRS	*Rapporteur*
Dr. Alexandre, Djiane Chargé de recherche, IRCM – INSERM	*Rapporteur*
Dr. Michael, Rera Chargé de recherche, CRI – CNRS	*Examinateur*
Dr. Allison, Bardin Directrice de recherche, Institut Curie – CNRS	*Directrice de thèse*

institutCurie | PSL

Dedicated to Iman Al-Kaisi

Acknowledgements

My first big thank you goes to my supervisor Allison Bardin, who has gone above and beyond to support me throughout this process. Allison helped me venture through my unique journey with more ease, giving me boosts along the way and has relentlessly provided me with what I consider “lifelines”. Allison is not just a supervisor, she has been my friend, my role model, cheerleader and the reason I picked myself up after every dip (there were quite a few). A “thank you” will never suffice. So, Allison, I promise, that in addition to appreciating everything you have done for me, from “zamboniing” my thoughts/writing, to tolerating my (optimistic) time-management, I promise to take all what I have learnt from you and permanently instil it within me as I continue to evolve. It has all been a steep learning curve and you’ve shaped me into a far better person than when I started.

Second, I would like to thank the people who were integral to contributing towards this work. From the bioinformatics side, Nick Riddiford, who developed packages important for the analysis of my experiments and patiently helped me run them myself. Nick never minded re-explaining something to me and always had a wonderfully clear way of simplifying all things complicated. Nick has been so generous with his time making a real difference to the quality of my PhD. Thank you, Nick, I’m eternally grateful. A big thank you to Annabelle Suisse for making collaborative work fun and stimulating and for the opportunity to add a beautiful section on aneuploidy in my thesis. An important dimension of science is collaboration, so Annabelle, I’m grateful to have discovered what that is like with a person like you.

From the experimental side, Marine Stefanutti was my superstar saviour, getting things up and going post-COVD19 confinement. Marine helped with dissections and staining of hundreds and hundreds of guts post-confinement and has contributed a lot pre-confinement too. It would have been impossible without Marine. From the very beginning I’ve admired Marine’s dedication and love for her job. Her work ethic is one I wish to emulate. Luckily, Marine was not just my work colleague and lab-mate, she became a close friend, one I confided in, travelled with, and trusted. She has been the source of many smiles and the reason behind many laughs. Also my heartfelt thanks to the NGS platform for contributing to an important component of my experiments with the library preparations for the sequencing as well as sequencing my precious samples.

Of course, I have a lot of gratitude towards each and every person in the lab. So, I want to thank the following people: Benjamin, thank you for being my partner in crime from the very beginning. Thank you for the great discussions, for sharing your latest readings and brilliant quirky taste in music. Also, for letting me speak out loud and guiding me through my tangled-up thoughts. You have a beautiful mind and an incredible vision and ability to see things from different angles when I struggled. Manon, for exemplifying what good organisation looks like and for being a wonderful friend. Thank you for the chats, the boosts in motivation (with the help of a small penguin). I look forward to a life-long friendship with you! Kasia, thank you for answering many of my questions when things didn’t click with me just yet and for the manuscript feedback. Kasia was my *Droso* teacher and thanks to her brilliant work, I was able to do my work. I want to thank Louis for manuscript feedback, his kindness and for normalising my problems, when I expressed them. He always reassured me if something went ok, and those meant the world to me in those moments. Hojun, for making me want to strive for beauty in visual science-communication and for solving my illustrator concerns. Natalia, for being my weekend buddy and for always being that smiling face who answered

my informatics concerns. The lab, as a whole collective entity for always believing in me in times when my confidence was wavering. It has been difficult at times, but in my mind's eye, all the difficulty has been overshadowed by the incredible fun we've had. Most importantly, it has been an environment for me to grow and make mistakes (a lot) with no judgement from others. It's rare to see this much compassion in the work-place. It felt like every fall was somewhat a "cushioned fall" yet the environment was stimulating enough for me to fall (the perfect balance!).

I would also like to express my sincere gratitude to Yohanns, Françoise and Delphine who have offered insights during my thesis advisory committee meetings. A big thank you to my IC-3i family, who have made it an incredible PhD journey, I am proud to have gone through these years with such a diverse and exquisite cohort.

On a more personal level, I want to thank Hala and Kousay who have entrusted me with a pair of precious wings and told me to soar anywhere I want with them. Not everyone gets the opportunities I got and I owe it all to them. I want to thank my brothers Tommy and Rami, of course. Tommy for always being my saviour and Rami, for continuously being curious about what I'm up to in the lab (and asking brilliant questions), they both make me a very proud sister.

I want to thank Taha, Kahtan and Rabee for being my family in Paris, giving me moral support at Maison du Cambodge and countless good times and laughs at CitéU and the beautiful streets of Paris. They gave me a taste of Syria during these years, which inevitably comes with a lot of bittersweetness. Thank you for cooking for me, making sure I'm ok, that I've been introduced to all your friends and that I don't miss out on a single event, filling my days with music, food, lots of love and my favourite memories. You've made Paris my home away from home.

In addition to a wonderful support system in Paris, I also had active remote cheerleading from across the channel and also across the pond! Thank you to my soul sisters Yomn and Saryah, my support system from London (my eternal support system, my "constants"). Hania also in London, for getting me active again with her brilliant zoom fitness classes, it's the longest time I've stayed committed to a fitness class! Ilyas gets a big thank you for remotely helping me from NYC with all my informatics concerns. Clarissa, also in the US, for being integral to my mental health during COVID19 lockdown and post-lockdown and all its challenges. And, this may sound weird, but I deem it necessary... I want to say "thank you very mouche" to the mouches. Thank you *Drosophila*, for teaching me so much in the lab, for teaching me so much about how biology works, and for getting me closer to the essence of who we are. You've made me very unique amongst my non-science friends, I became "the girl who works with flies" and I'm very proud of that.

Table of Contents

Table of Figures

Supplementary Figures

Abbreviations

alt-EJ: Alternative end joining
AMPs: Antimicrobial peprides
APC: Adenomatoud Polyposis Coli
ATM: Ataxia Telangiectasia Mutated
bHLH : basic-helix-loop-helix
BIR: Break-induced replication
Blm: Bloom helicase
CCR: Copper cell region
CO: Crossover
CRC: Colorectal cancer
DBA: Diamond Blackfan Anemia
DDR: DNA Damage Response
Dl: Delta
D-loop: Displacement loop
DNMT3: DNA methyltransferase 3A
dNTP: deoxynucleoside triphosphate
dHJ: double Holliday junction
DSB: Double strand break
DSBR: Double strand break repair
EB: Enteroblast
ECs: Enterocytes
EEPs: Enteroendocrine precursors
EEs: Enteroendocrine cells
EGFR: Epidermal growth factor receptor
ESCC: Esophogeal squamous cell carcinoma
***Ecc15*:** Erwina carotovora carotovora 15
Gcn2: General control nonderepressible 2
HFSCs: Hair follicle stem cells
HR: Homologous recombination
HSCs: Hematopoietic stem cells
IBD: Irritable bowel disease
IGF-1: Insulin-like growth factor 1
IR: Irradiation
ISCs: Intestinal stem cells
JNK: c-Jun N-terminal kinase
***KRT10*:** Keratin 10
Kif1: Kinesin-like protein
LCR: Low copy repeats
LOH: Loss of heterozygosity
MCM4: Minichromosome Maintenance Component 4
MMEJ: Microhomology-mediated end joining
MR: Mitotic recombination
NCO: Non crossover
NF1: Neurofibromatosis 1
NHEJ: Nonhomologous end joining
NSCs: Neural stem cells
PCR: Polymerase chain reaction
PH3: Phospho-Histone H3

***Pe* :**	Pseudomonas entomophila
***Pros* :**	Prospero
RFLP :	Restriction fragment length polymorphism
ROS :	Reactive oxygen species
RPA:	Replication protein A
Rb:	Retinoblastoma
SDSA:	Synthesis dependent strand annealing
SJ:	Septate junction
SNP:	Single nucleotice polymorphism
SNV:	Single nucleotide variation
Su(H) :	Suppressor of hairless
TCJ :	Tricellular junctions
TET2 :	Ten eleven translocation 2
TMEJ :	Theta-mediated end joining
TOR:	Target of rapamycin
UV:	Ultra-violet
Upd3:	Unpaired 3

Chapter 1
Introduction

Chapter 1 : Introduction

Prelude

There are three major themes I wanted to cover in my introduction. 1. The importance of better understanding the altered genomes of stem cells. 2. The gaps in our knowledge regarding important mechanisms driving genome alterations. 3. The use of the Drosophila intestine as a model to address questions regarding mechanisms of genome alteration in stem cells.

Therefore, in chapter 1.1, I will begin with a general overview of DNA damage in stem cells and how it could lead to genome alteration, potentially driving aging phenotypes and cancer. In chapter 1.2, I will focus more on a common cause of genome alteration: loss of heterozygosity, highlighting its contribution to disease and particularly cancer initiation. Finally, in chapter 1.3, I will introduce the Drosophila model system that I use, to answer important questions about loss of heterozygosity in stem cells.

Chapter 1.1 is largely based on a review that I wrote in my third year (Al zouabi and Bardin 2020) with the goal in mind of using it for my thesis introduction: Al zouabi L and Bardin AJ (2020 Jan 13) Stem Cell DNA Damage and Genome Mutation the Context of Ageing and Cancer Initiation - Cold Spring Harbor Perspectives in Biology.

It has, however, been slightly modified to accommodate a better flow for the thesis.

1.1 DNA damage and mutation in stem and progenitor cells in the context of aging and cancer

Stem cells and tissue dynamics

Many adult metazoan tissues maintain long-term function through the ongoing elimination of terminally differentiated cells and the replacement of these cells by the newly divided progeny of cycling cells. An understanding of this process began almost 70 years ago through early lineage tracing studies of Charles Philippe Leblond using tritiated-thymidine injection in mice to reveal the turnover rates of labeled cells (Leblond and Walker 1956). These seminal studies initiated the stem cell theory of renewal and laid the foundation for modern labeling studies of cell turnover that confirmed age mosaicism of adult tissues (Arrojo e Drigo et al. 2019; Spalding et al. 2013). Importantly, this work raised important conceptual questions regarding how stem cells may endure the process of aging.

Aging is associated with an alteration in stem cell functionality and kinetics of tissue renewal in many tissues such as blood, skin, muscle, and the brain (Kuhn et al. 1996; Morrison *et al.*, 1996; Conboy *et al.*, 2003; Nishimura et al. 2005). An imbalance in tissue dynamics due to deregulated self-renewal or cell turnover rates can compromise tissue function. Understanding how tissue dynamics are altered during aging or in pathological contexts is an important, yet highly complex question. At a molecular level, these changes may be induced by genetic, epigenetic or metabolic alteration. Potential causes may include changes in stem-cell-intrinsic factors, alteration of niche properties, or modification of systemic signals. Here, the focus will be on stem cell-intrinsic alteration through DNA damage and genetic mutation. For recent reviews on the impact of epigenetic and metabolic changes on

aging stem cells, please see (Brunet and Rando 2017; Booth and Brunet 2016; Chandel et al. 2016).

DNA damage and how it leads to mutation

All cells, including stem cells, are faced with the challenges of protecting their DNA from erosion. DNA damage is a deviation from the normal DNA structure with the introduction of damaged sites in the base-pairing or backbone structure. Multiple exogenous agents such as UV light, ionizing radiation and chemical mutagens, such as hydrocarbons present in tobacco smoke, can damage DNA. In addition, endogenous factors such as reactive oxygen species, telomere erosion, and replication errors can also be a source of damage. DNA replication, for example, is an opportune time for error, as the replication fork can slow down, or collapse, due to topological challenges including limiting nucleotides, repetitive sequences, non-B-form DNA, and collisions with transcription machinery.

It has been estimated that tens of thousands of lesions are experienced by a mammalian cell per day, with single-strand lesions making up the majority of this number (Lindahl and Nyberg, 1972; Lindahl, 1974; Lindahl and Barnes, 2000). Damage involving a single-strand can be accurately repaired using the other strand as a template. Small base lesions that do not significantly change the DNA helix structure are repaired by base-excision repair (Lindahl 1974). As for misincorporated bases, they are corrected by the mismatch repair pathway (Lahue *et al*. 1989). On the other hand, lesions involving bulky adducts, and dimers are repaired by nucleotide-excision repair (Sancar 1993) whereas interstrand crosslinks require the Fanconi anemia pathway (Zhang and Walter 2014).

Additionally, DNA double-strand breaks also arise in the cell, often through replication fork collapse. This type of damage is more dangerous because more error-prone repair mechanisms are used that can lead to the loss of genetic information, contributing to genome instability. DNA double-strand break repair is primarily orchestrated by two pathways: if the cell has gone through S phase, duplicating its chromosomes, providing a template for the repair of the damaged chromosome, homologous recombination (HR) is usually used. HR is regarded as a mechanism of “high-fidelity” due to the presence of another template from either the sister or homologue within which the bases that have been broken are still intact, losing no genetic information (this mechanism will be covered more thoroughly in chapter 1.2).

If, on the other hand, the cell is in G1, the more erroneous non-homologous end joining (NHEJ) pathway is used, which is an efficient and quick repair mechanism that does not make use of an external homologous template. NHEJ instead involves the bridging of broken ends by a DNA dependent protein kinase consisting of a Ku heterodimer and a ligation step via DNA ligase IV, which comes with the introduction of small deletions as a result.

There is also an additional third pathway called alternative-end joining (alt-EJ), which is considered a “backup pathway”. Alt-EJ was initially discovered through inactivating the NHEJ pathway in cells and it was found that these cells can undergo repair without the use of Ku or ligase IV (Shima et al. 2003). Alt-EJ relies on short homologies of 5-25 bp near the broken ends to align the broken strands before joining with a ligase. Alt-EJ is also sometimes referred to as microhomology-mediated end joining (MMEJ) or theta-mediated end joining TMEJ, as it has been

shown that DNA Polθ has important roles in stablising the end-joining intermediate (McVey and Lee 2008).

Despite the existence of these strategies to safeguard the integrity of DNA, glitches in the system arise frequently leading to sequence variants, structural variants, or aneuploidy. Sequence variants include indels and point mutations, arising, for example, through deamination of 5-methylcytosine in CpG nucleotides in vertebrates, resulting in a C—>T substitution (Razin and Riggs 1980). Structural variants involve more large-scale changes to the DNA sequence and therefore are more likely to alter gene function. These include amplifications, deletions, and translocations, which can be caused by recombination and replication-based mechanisms, erroneous DNA double-stand break repair, or be a result of transposable element mobility (Carvalho and Lupski 2016; Bourque et al. 2018). Most genes are haplosufficient (Huang et al. 2010) and therefore inactivation of one copy may not impair cell function. However, problems may arise when one allele in the genome is already inactive in the germline and the second allele is inactivated somatically, leading to loss of heterozygosity (LOH). LOH can be driven by the aforementioned mutagenetic processes as well as recombination with the homologous chromosome, also known as “mitotic recombination”, which upon cell division, leads to segregation of two mutant alleles into one daughter cell. (LOH is discussed in more depth in Chapter 1.2.) Thus, DNA damage in stem or progenitor cells can alter the genome in numerous ways and potentially radically disrupt tissue function over the course of aging.

Some of the first evidence suggesting a potential causal link between cellular DNA damage and organismal aging came from the realization that inactivation of DNA repair genes such as in Fanconi anaemia and Werner syndrome in humans

lead to early aging or “progeroid” syndromes (Carrero et al. 2016; Moskalev et al. 2013). Due to reduced ability to repair DNA damage, DNA lesions persist, and somatic mutations accumulate. Patients with these syndromes exhibit accelerated aging and present symptoms of loss of proper tissue renewal such as skin atrophy, loss and graying of hair, and higher susceptibility to cancer development. While these studies suggest sufficiency of DNA damage to drive early aging phenotypes, they do not provide evidence that endogenous levels of DNA damage or mutation can impact aging. How DNA damage and genome mutation may impair stem cell function will be further discussed below. First the mechanisms that can mitigate the effects of DNA damage in stem cells will be examined.

Mechanisms protecting the stem cell and tissue from the effects of DNA damage

While coping with DNA damage is important for all cells, it is particularly vital for adult stem cells that renew tissues throughout adult life. What are the ways in which stem cells and tissues avoid the negative impact of DNA damage and mutation? Here, a number of important protection mechanisms acting at the stem cell and tissue level will be discussed.

Protecting the stem cell: DNA damage responses and repair

Evidence suggests that at least some stem cells employ distinct mechanisms from their downstream differentiated or more committed progenitor cells to prevent the accrual of genetic lesions, which can be detrimental to homeostasis of the tissue. DNA damage is managed via the DNA damage response (DDR), which is an evolutionarily conserved signaling pathway where sensors, mediators and effectors orchestrate DNA repair or by the elimination of the damaged cell by apoptosis or by

exiting the cell cycle. Interestingly, adult mouse hematopoietic stem cells (HSCs) and hair follicle stem cells (HFSCs) of the bulge were found to have increased radioresistance with minimal apoptotic response and accelerated DNA repair compared to their more differentiated progeny (Mohrin et al. 2011; Sotiropoulou et al. 2010; Beerman et al. 2014). Aged HSCs are even more resistant than young HSCs to DNA-damage induced apoptosis (Gutierrez-martinez et al. 2018). This likely helps prevent depletion of the stem cell pool but could be at the cost of accumulating mutations.

Stem cells also differ in a tissue-dependent manner in terms of strategies used that help limit passing mutations to progeny, with some favoring robust repair (**Figure 1.1A**), others apoptosis (**Figure. 1.1B**), or terminal differentiation (**Figure 1.1C**). The small intestine for instance is sensitive to apoptosis driven by DNA damage, whilst stem cells of the colon are resistant to apoptosis (Potten and Grant 1998; Merritt et al. 1995). Intrinsic differences in cell cycle properties could explain why stem cells differ widely in their DNA repair mechanisms between tissues. As shown in the hematopoietic system, when a DSB arises in a quiescent cell, DNA repair is mediated by the efficient NHEJ mechanism, which acts quickly and does not need the presence of a homologue for repair, but is error-prone (Mohrin et al. 2011). Proliferating HSCs on the other hand use high-fidelity HR to repair DSBs but have an increased likelihood of accumulating damage during S-G2/M (Mohrin et al. 2011). Alternatively, another strategy of protecting the tissue from propagating a mutation is employed by melanocyte stem cells that differentiate upon DNA damage (Nishimura et al. 2005; Inomata *et al.*, 2009). Similarly, during the aging process, it is thought that the HFSC are gradually lost due to differentiation upon repeated DNA damage acquisition during hair follicle cycles (**Figure 1.1C**) (Matsumura et al. 2016).

Caught between balancing the need to maintain tissue function and the need to block the propagation of mutations, stem cells have evolved diverse modes to cope with DNA damage and repair, often sacrificing immediate survival of a given stem cell for the expense of long-term maintenance of genome-integrity in the tissue. Further studies are important to better understand the sensitivity and resistance of adult stem cells to damage, the repair mechanisms employed, and age-related changes in this process.

Protecting the stem cell: A quiescent state

One way to limit DNA damage is simply to avoid undergoing cell division, which would restrict replicative and chromosome segregation errors (**Figure 1.1C**). Indeed, many populations of adult stem cells including hematopoietic, muscle, and neural stem cells remain in a non-proliferating quiescent state of G0 (van Velthoven and Rando 2019; Cho et al. 2019). Evidence suggests that quiescence serves a protective role in these contexts as these populations of stem cells become depleted or “exhausted” when driven into the cell cycle upon transplantation, due to stress, or upon genetic manipulation (Chen *et al.*, 2000; Harrison, 1978; Gan *et al.*, 2010; Sacco *et al.*, 2010; Schaniel *et al.*, 2011; Staber *et al.*, 2013; Cavallucci et al. 2016; Yue *et al.*, 2016; Baumgartner *et al.*, 2018; Singh *et al.*, 2018; Kamminga et al. 2006)**.** Stem cell exhaustion in these contexts may be due to loss of niche

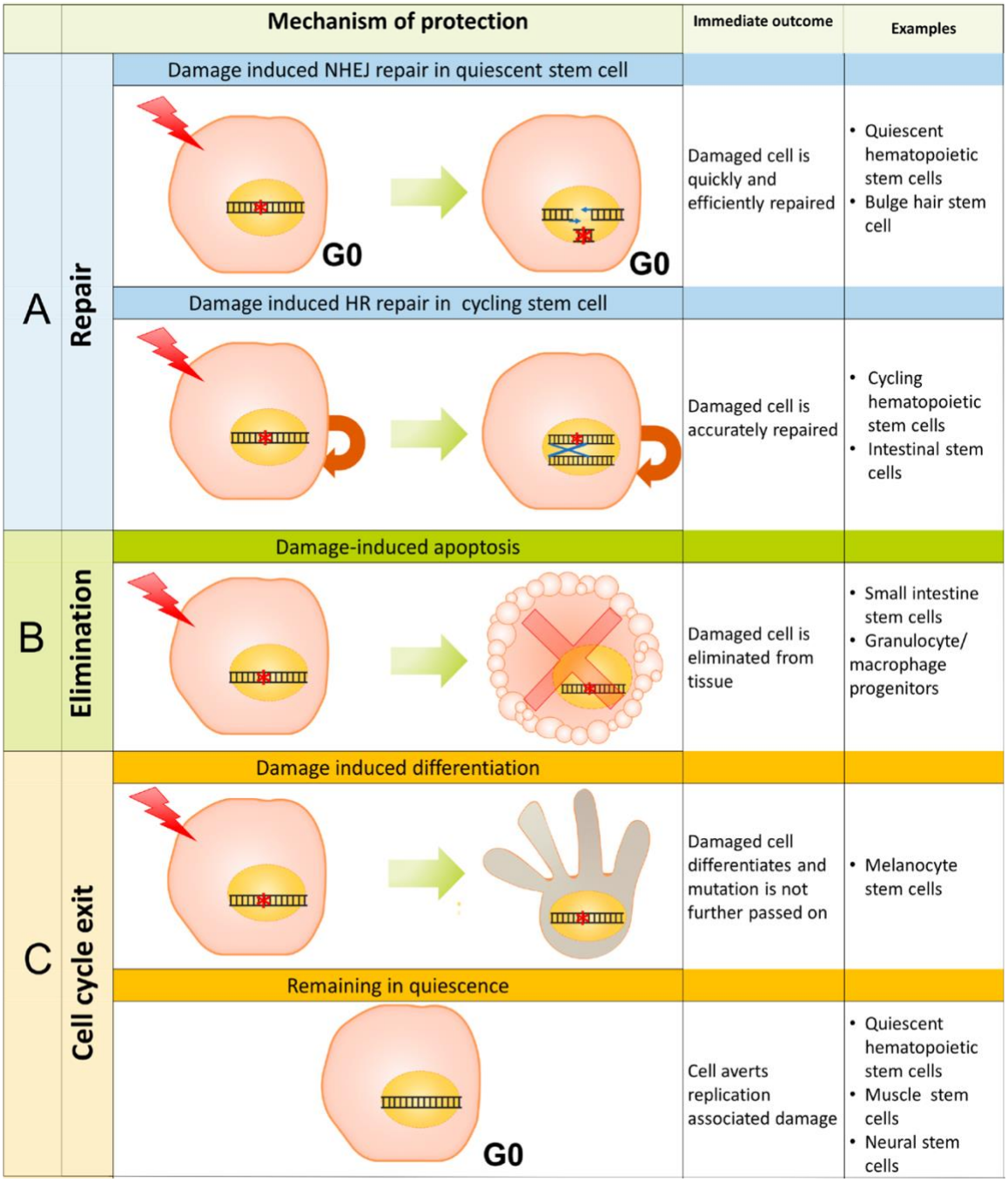


Figure 1.1: Mechanisms of stem cell protection from DNA damage by repair, elimination or cell cycle exit.

(A) Repair: Depending on the cell cycle status of the cell, the cell undergoes repair by either non-homologous end-joining (NHEJ); (top panel) or homologous recombination (HR); (bottom panel). NHEJ is the quick and efficient mechanism employed by quiescent stem cells when they are faced with damage. It involves the ligation of the broken ends and often results in the introduction of small deletions, but can also lead to translocation and genome rearrangements. HR is employed if the cell is cycling and goes through S phase, duplicating its chromosomes, providing a template for the repair of the damaged chromosome. This is usually more accurate repair than NHEJ, though erroneous choice of the homologous chromosome, rather than the sister, can lead to LOH.

(B) Elimination: by apoptosis. Some cells undergo apoptosis rather than repair. If this

mechanism is preferentially employed in the stem cell, there is a higher chance of stem cell depletion.
(C) Cell cycle exit: by differentiation (top panel) upon DNA damage, or remaining in a state of quiescence.

signals. Alternatively, these studies raise the possibility that increased DNA damage or an increased mutational burden upon loss of quiescence may lead to stem cell functional decline during aging (Sharpless and DePinho 2007). Consistent with this notion, when mouse HSCs were forced repeatedly out of quiescence they acquired DNA damage and became depleted (Walter et al. 2015). Nevertheless, the extent to which stem cell exhaustion is related to increased DNA damage or acquisition of mutations is not entirely clear and may differ depending on stem cell type. Additional potential links between DNA damage and stem cell senescence will be discussed later in this chapter.

Protecting the tissue: Competition between cells and lineages

In addition to stem-cell intrinsic mechanisms of protection mentioned above, tissue-level protection also helps to ensure the survival of the fittest lineage. This may be especially important for stem cells such as the Crypt Basal (Lgr5+) intestinal stem cells (ISCs) and those of the skin epidermis, that actively divide. One such mechanism is neutral competition in which ISCs undergo dynamic stem cell replacement shown, both in the mouse (Lopez-Garcia et al. 2010; Snippert et al. 2010) and the *Drosophila* intestine (De Navascués et al. 2012), which likely helps to prevent their loss.

Aside from neutral competition, biased cell competition also occurs between cells. Initially described in *Drosophila*, biased cell competition is a phenomenon whereby differences in cellular fitness allow selection of “winner” cells, while weeding out less-fit “loser” cells (**Figure 1.2A**) (Morata and Ripoll 1975). A large

body of work in *Drosophila* has revealed that this process plays an important role in shaping adult tissues and has elucidated many molecular mechanisms underpinning this process. For a review of the topic see (Levayer 2019). Competition between stem cells also occurs and can result in greater niche occupancy of a given genotype with selective advantage or greater production of progeny (Nystul and Spradling 2007; Jin et al. 2008; Issigonis et al. 2009; Kolahgar et al. 2015; Zhang and Kalderon 2001; Amoyel and Bach 2014).

An important question, however, is what types of fitness differences are being sensed during cell competition? Could stem cells with DNA damage or with less fit mutant genomes be selected against? Interestingly, mouse HSCs that have been treated with low dose ionizing radiation (IR) are less competitive than non-irradiated HSCs in a manner that is dependent on Trp53 levels and last for weeks (**Figure. 1.2B**) (Bondar and Medzhitov 2010). This implies that a memory of the irradiation stress was kept, which is proposed to be linked to a Trp53-dependent long-term mark acting as a cellular memory for DNA damage. A recent study of the mouse skin demonstrated that stem cell lineages compete based on levels of the hemidesmosome component, COL17A1 (**Figure 1.2C**) (Liu et al. 2019). Interestingly, like HSCs, exposure of epidermal stem cells to IR triggers a long-lasting memory of genomic stress, resulting in the proteolytic degradation of COL17A1. How this memory is achieved and whether it also relies on p53, unrepaired DNA damage, or could be linked to genomic mutations or epigenetic mechanisms, is not clear. Another recent study of the mouse skin epidermis showed, as previously demonstrated in *Drosophila* (de la Cova, 2004; Moreno and Basler 2004), that during epidermal stratification cell lineages with higher levels of

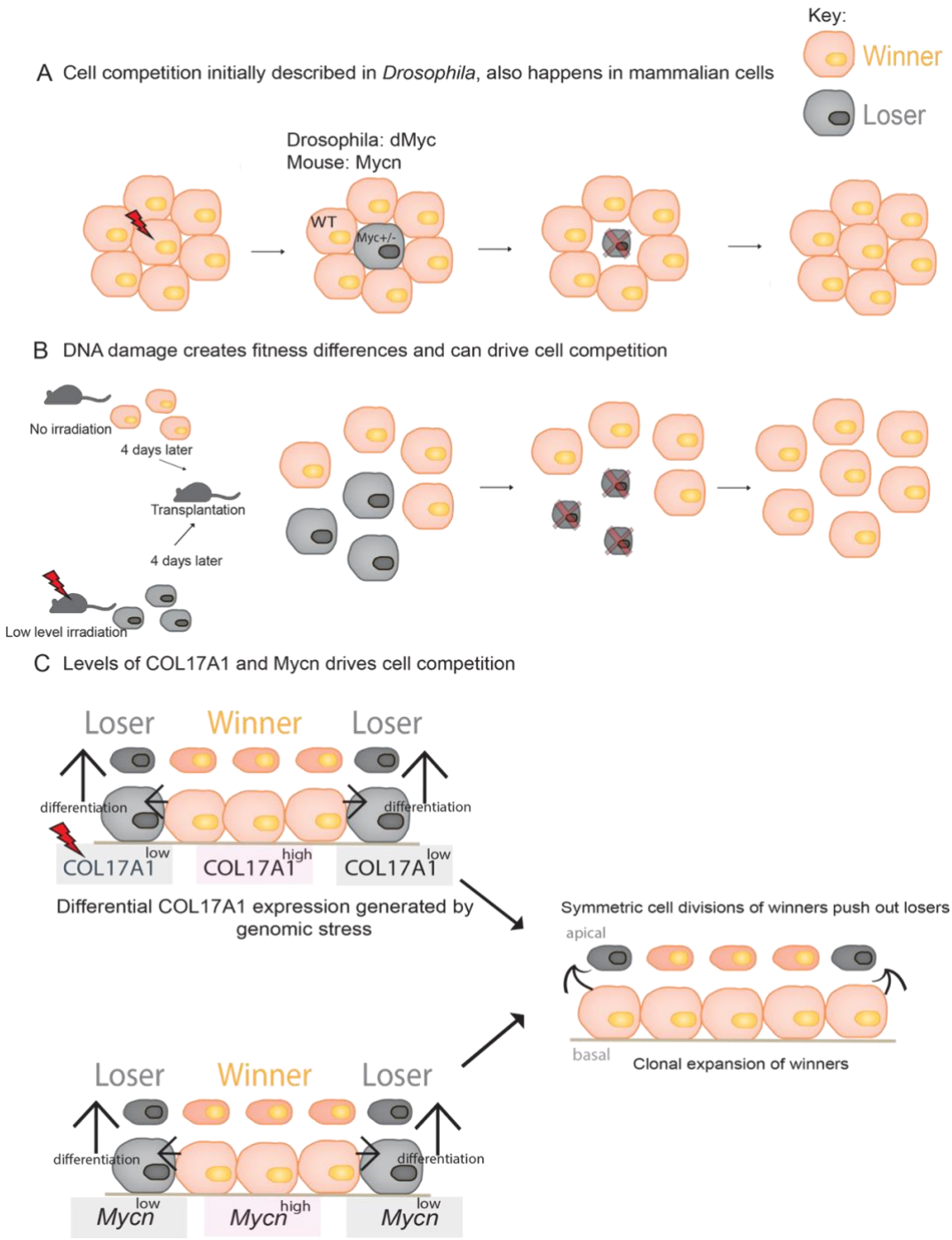


Figure 1.2: Cell competition selects for "winner" cells and weeds out less-fit "loser" cells.

(A) Differential Myc levels drive cell competition in both Drosophila (progenitors in the disc) and mammalian cells (mouse developing epidermal cells and adult epidermis). This figure illustrates that Myc mutants are "losers" and are outcompeted by adjacent WT cells. Extra levels of Myc also renders cells "winners" compared to wild-type cells (not shown).

(B) DNA damage creates differences in fitness and this can drive cell competition. Irradiation of mouse hematopoietic stem and progenitor cells creates DNA damage which results in increased p53 levels. When these cells are transplanted into mice with non-irradiated cells (4 days after the irradiation), the irradiated cells are outcompeted by the non-irradiated cells that have lower p53 levels and a higher expression of more competitive signalling molecules. Thus, DNA damage via irradiation creates a long-lasting "loser" cell status by

inducing p53-mediated apoptosis or cell cycle arrest.
(C) Stem cell lineages with higher levels of COL17A1 and Mycn become “winners” in mouse skin epidermis. Genomic stress leads to the proteolytic degradation of COL17A1 and thus results in differential levels of COL17A1 expressed in the epidermis. Cells with higher levels of COL17A1 outcompete the cells expressing lower levels via symmetric cell division and the elimination of the losers. The higher expression of COL17A1 maintains a healthy skin phenotype, whereas COL17A1 deficiency causes skin atrophy, fragility, dyspigmentation and alopecia. Similarly, skin lineages with higher levels of Mycn outcompete the cells expressing lower levels of Mycn, but it remains unclear whether genomic stress is what drives differential Mycn expression.

Mycn, a bHLH transcription factor, become winners (**Figure 1.2A, C**). It is not currently clear whether, like COL17A1, *Mycn* might respond to altered genomic stress and how these two mechanisms might overlap. Interestingly, mechanisms that may be akin to cell competition can also expunge aberrant tissues with altered tissue architecture, such as those expressing oncogenic *Hras* GTPase, as demonstrated in mouse hair follicle using live imaging (Brown et al. 2017). This is very reminiscent of early work in the fly showing elimination of tumorigenic cells via cell extrusion (Brumby and Richardson 2003; Vaughen and Igaki 2016). Thus, cell and lineage competition are mechanisms that can help to maintain integrity of adult tissues and are likely one means of eliminating cells with harmful DNA damage or mutant genotypes.

When protection mechanisms fail: acquisition of mutation

Despite the numerous mechanisms in place to protect stem cells from harmful effects of DNA damage, studies over the past 10 years revealed the extent to which genomic mutations arise in adult stem cells. Here we will present data demonstrating that genetic changes occurring in stem or progenitor cells contribute to tissue mosaicism. We will also highlight some of the recent literature from humans that has demonstrated that somatic genetic mosaicism is not a rare, pathological event, but a phenomenon present in many of our healthy adult tissues.

Evidence of surprising diversity in somatic genomes

Finding and studying somatic mutations in subsets of cells within a tissue is extremely challenging. While recent advances in genomic sequencing are beginning to unveil the extent to which somatic variation arises, classic genetic studies using visible marker phenotypes provided the first evidence of genetic mosaicism. Studies by Curt Stern using *Drosophila* first demonstrated spontaneous loss of heterozygosity (LOH) during development due to mitotic recombination between homologous chromosomes (Stern 1936b). Mitotic recombination is an important mechanism of LOH in cancer and other genetic disorders (Jonkman et al. 1997; Choate et al. 2010), though not yet well understood in healthy tissues. Somatic variation due to mobilization of transposable elements was later studied in maize by Barbara McClintock (1950). Evidence from reporter mice and DNA sequencing-based approaches suggest that Line1 element mobility contributes to genetic mosaicism in the nervous system (Coufal et al. 2009; Erwin et al. 2016; Upton et al. 2015; Muotri et al. 2005) and estimate a de novo *Line1* element insertion frequency of 0.2 events per neuron in humans (Evrony et al. 2012). See (Faulkner and Garcia-Perez, 2017) for a more extensive review of this literature. How somatic mobilization of transposable elements impact adult tissues, is only beginning to be understood.

Additional mutagenic processes also shape somatic mosaicism. Sequencing clonally expanded human adult stem cells using organoids has demonstrated that around 40 de novo point mutations are acquired per year in liver, colon, and small intestine (Blokzijl et al. 2016); 13 de novo point mutations mutations per year in muscle stem cells (Franco et al. 2018); and about 200-400 total point mutations impact neural precursors (Lodato et al. 2015). One prominent mutational signature found in both human and mouse precursors is C-to-T transitions at CpG

dinucleotides, thought to be due to deamination of 5-methylcytosine to thymine (Behjati et al. 2014; Blokzijl et al. 2016; Lodato et al. 2015). In addition, larger-scale gene deletion and rearrangements were detected using SNP array methodology, with around 14% of human colon crypts bearing a large-scale deletion or LOH event (Hsieh et al. 2013), which has been also documented in other tissues (O'Huallachain et al. 2012). Whole-genome sequencing of colon also recently confirms SNP and copy number changes in healthy tissue (Lee-Six et al. 2019). Aneuploidy and copy number variation in the brain and other tissues have similarly been reported, though frequencies vary depending on the detection technique (Rehen et al. 2002; Cai et al. 2014; O'Huallachain et al. 2012). Thus, it is now abundantly clear that human tissues have high degrees of genetic mosaicism. It is, therefore, critical to perform functional studies to understand the full impact of mosaicism on young, aged, healthy and diseased adult tissues.

Clonal expansion in blood and solid tissues

Mosaic patches of adult tissue, or "clones", can result from a long-lived stem or progenitor cell acquiring a mutation driving positive selection due to increased fitness, or from neutral drift of an alteration with no impact on fitness (Snippert et al. 2010; Traulsen et al. 2013). Evidence for age-dependent clonal expansion of mutant stem cell lineages in the blood dates back to the 90s where probes for the inactive X-chromosome were used and detected its skewing during aging (Fey et al. 1994; Busque et al. 1990). More recently, the study of "healthy" control blood using sequencing-based approaches led to surprising evidence for clonal expansion of lineages having somatic mutation in the genes *TET2*, *DNMT3a,* and *ASLX1* during adult aging (Busque et al. 2012; Jacobs et al. 2012; Laurie et al. 2012; Holstege et

al. 2014; Welch et al. 2012). The physiological implications of blood clonality will be discussed further below but for an extensive review on clonal haematopoiesis see (Jaiswal and Ebert 2019).

Mounting evidence similarly indicates that solid tissues also have a high degree of genetic mosaicism with mutant progenitor cells giving rise to expanding mutant lineages under positive selection. In the 90s, it was recognized with PCR and through whole-mount tissue staining that sun-exposed normal human skin acquires clones of mutant *TP53* (Nakazawa et al. 1993; Jonason A S et al. 1996). In recent years, these finding were greatly extended using targeted deep sequencing of 74 cancer driver genes on biopsies of normal sun-exposed eyelid epidermis and normal esophagus tissue. Frequent mutation of genes was found, including in *NOTCH1* and *TP53,* that expand clonally and accumulate with age (Martincorena *et al.*, 2015, 2018; Yokoyama *et al.*, 2019). Additional recent evidence for large clonal expansions across numerous tissues including breast and lung has been demonstrated with mutational analysis of RNAseq data (Yizhak et al. 2019). Furthermore, other tissues show clear examples of somatic mutation-driven clonal expansion. In humans, megaencephaly syndromes leading to a clonal overgrowth of part of the brain arise through activating mutations of the AKT/PI3K pathway that can be due to somatic mutations arising in neural precursor cells (Lee et al. 2012; Rivière et al. 2012; Lodato et al. 2017; Poduri et al. 2012). Interestingly, somatic mutations activiating PI3K have also been found to lead to Proteus syndrome, with patients having overgrowth of fibrous and adipose tissues (Lindhurst et al. 2012). Thus, positive selection of mutant lineages is prevalent in human tissues. The implications on cancer initiation of somatic mutations in driving early lineage expansion and selection will be further discussed below.

DNA damage and somatic mutation in adult tissues: roles in cancer initiation and aging

What is the impact of these mutations on tissues? Clearly cancer initiation is one detrimental consequence, but not all mutations lead to cancer. Here the functional implications of somatic genetic mosaicism will be highlighted.

Somatic mutations and cancer initiation

For over a hundred years, it has been recognized that cancer cells are distinct from normal ones due to the presence of aberrant genomes (Boveri 1914).Therefore, recent revelations that normal tissues harbor extensive mutations, raise important questions about the relationship between apparently healthy tissue and cancer: do mutations that provide positive selection in a tissue actually promote the eventual acquisition of additional genetic mutations leading to cancer as described in a classical multistep carcinogenesis model? Alternatively, in some instances, might these be two distinct selection processes with cancer requiring a divergent path from one that optimizes growth within an otherwise healthy tissue? As previously discussed, multiple modes of cell and lineage competition actively shape the nature of selection within a tissue and, in theory, could respond differently to expanding mutant lineages versus precancerous clones.

Evidence from clonal hematopoiesis supports a multistep process where a first mutation in healthy tissue precedes additional mutation (**Figure 1.3A-D**), increasing cancer risk. Indeed, longitudinal studies of patients with clonal hematopoiesis detected by SNP arrays support a strong increased risk of developing not only hematological cancer (Laurie et al. 2012; Jacobs et al. 2012; Welch et al. 2012; Genovese et al. 2014; Jaiswal et al. 2014; Coombs et al. 2017),

but also lung and kidney cancers (Jacobs et al. 2012). Exome sequencing revealed that known tumor suppressor genes of myeloid cancers such as *TET2, DNMT3A* and *ASXL1*, were mutated in apparently healthy blood (Busque et al. 2012; Genovese et al. 2014; Jaiswal et al. 2014; McKerrell et al. 2015; Coombs et al. 2017). Thus, the acquisition of these mutations in healthy blood is thought to represent the earlier phase in the development of leukemogenesis and suggests a period of latency that precedes it. Therefore, an understanding of how processes such as stem cell competition for niche occupancy may influence the switch from a premalignant state to a malignant one is important (**Figure 1.3B,D**).

Recent studies in the skin and esophagus support the idea of healthy tissue acquiring premalignant drivers, but also suggest the intriguing possibility that healthy tissues may have distinct selective pressures than those in cancer. Targeted deep sequencing of normal oesophageal epithelium from young and old donors revealed that the number of detectable mutations and the sizes of mutant clones increased with donor age (Martincorena, *et al.*, 2018). *NOTCH1* and *TP53*, canonical drivers of Esophageal squamous cell carcinoma (ESCC), were found to be under selection in normal tissue (Martincorena *et al.*, 2018b; Yokoyama *et al.*, 2019). Thus, the presence of clonal expansions in the normal epithelium suggests that these clones have a premalignant capacity and their persistence can lead to cancer initiation (**Figure 1.3B,D**). These data strongly support the concept of “field cancerization” (Slaughter and Southwick 1953), previously proposed to predispose the esophagus to development of subsequent multiple tumors via initial precancerous drivers such as p53 (Tian et al. 1998). Nevertheless, an intriguing finding is that mutations in *NOTCH1* and *PPM1D* are much more prevalent in normal skin than in cancer (Martincorena *et al.*, 2018b; Yokoyama *et al.*, 2019). This suggests that different

fitness of certain mutations exist in “normal” tissue versus cancer, complicating the notion of a linear multistep mutation accumulation process. Future studies will be necessary to understand these fitness differences and potentially capitalize on them for clinical benefit.

An impact of mutations and DNA damage on aging?

Aside from initiating and driving cancer evolution, what impact do somatic mutations have on aging? Here some of the potential detrimental consequences of mutation on tissues will be discussed.

Studies from clonal hematopoiesis have demonstrated a collapse of clonal diversity with very few stem cells contributing to the aging blood (**Figure 1.3C**). This results from “winner” HSC clones expanding and, in an apparent zero-sum game, “loser” HSCs failing to contribute to blood. This was strikingly demonstrated from sequencing the blood of a hematologically asymptomatic supercentenarian (aged 115 years old) revealing that approximately 65% of her healthy blood compartment was dominated by the progeny of two hematopoietic stem cell (HSC) clones (Holstege et al. 2014). Extending on earlier work discussed above (Busque et al. 2012; Jacobs et al. 2012; Laurie et al. 2012; Welch et al. 2012), a study using whole-genome sequencing from the peripheral blood of ~11,000 Icelanders of different ages found that a striking 50% of patients older than 85 had clonal hematopoiesis (Zink et al. 2017a). Thus, abundant evidence indicates that mutations arise in HSCs (or in very upstream precursor cells) during aging and lead to selection of mutant lineages, however, the functional impact of collapse of clonal diversity is still not fully understood. One feature of the aging hematopoietic system in humans and mouse is a bias towards myeloid lineages (Sudo et al. 2000; Ganuza et al. 2019; Yamamoto

et al. 2017). While unlikely to explain all of the myeloid bias of HSCs that occurs during aging, *TET2* deletion is sufficient in mouse to lead to a myeloid disorder (Li et al. 2011) and is strongly associated with myeloid dysplasia in humans (Buscarlet et al. 2018). Thus, a failure to maintain the repertoire of differentiated cell types present in youth can arise from a loss of clonal diversity. Interestingly, a reduction in the clonality of mouse muscle stem cells upon repeated injury was found (Tierney et al. 2018). While the role of mutation or DNA damage was not evoked in this study, it is feasible that increased replication stress might indeed drive some stem cell lineages to contribute less to the tissue, possibly explaining the observed collapse in clonality in the muscle.

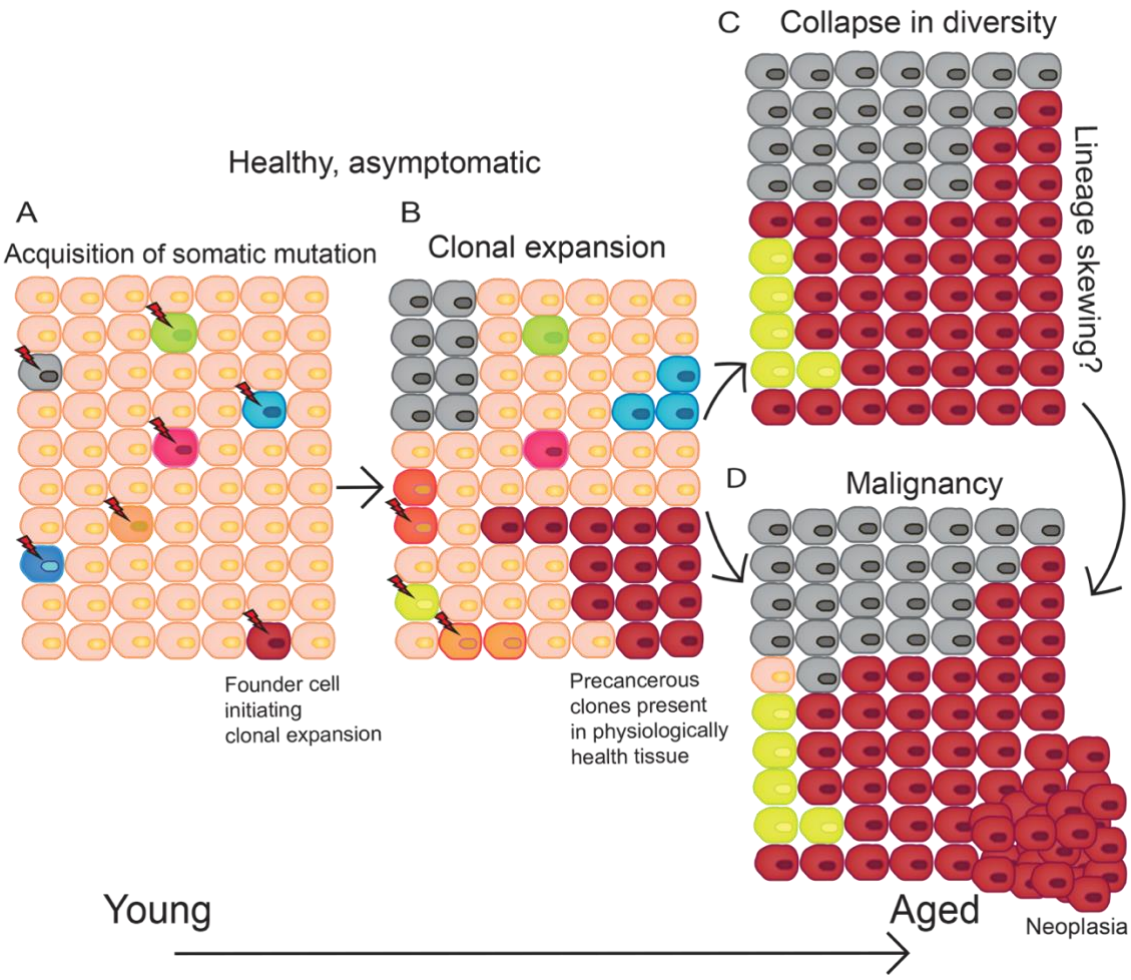


Figure 1.3: Somatic mosaicism with age

(A) Mutations arise in stem cells of young tissues.
(B) Age-dependent clonal expansion of mutant stem cells via positive selection or neutral drift give rise to mosaic patches. These may have a premalignant capacity and their persistence can lead to cancer initiation.
(C) Clonal expansion can lead to the age-related collapse in clonal diversity with very few stem cells contributing to the aging tissue. The functional impact of collapse of clonal diversity is still not fully understood but it can impact age-associated lineage skewing, in some cases.
(D) Clonal expansion of cancer driver genes can lead to cancer initiation.

Hypercompetitive lineages may render other lineages “losers”, but deleterious mutations may also create “loser” lineages cell-autonomously through suboptimal growth, stem cell functional decline, or loss from the tissue of the stem cell or lineage. Is there evidence for this? Quantifying deleterious mutations is a difficult task as these mutations will be either lost or only be present in a few cells.

As a work-around, techniques from evolutionary biology have been applied to look at negative selection of point mutations within somatic tissues. By considering the normalized ratio of non-synonymous to synonymous mutations, one can deduce the amount of detrimental mutations which had been lost. Strikingly, no evidence of negative selection was found in human tissues or in numerous types of cancer (Martincorena et al. 2017; Franco et al. 2018), arguing that the arising point mutations were not detrimental to the survival of the cell in which they arose. It is not yet clear how other types of mutational processes may create burdens on the cell or be selected against. For example, it is more likely that large-scale deletions or mitotic recombination-based LOH, both affecting hundreds to thousands of genes, would reduce cellular fitness. Similarly, de novo transposition events may also impair cellular function through transcriptional deregulation. The extent to which this occurs or might trigger cell death or cell selection mechanisms at the tissue level, is not yet known.

Contributions of persistent DNA damage to stem cell decline

A large body of literature using induced DNA damage has explored the effects of persistent DNA damage including on HSCs, NSCs, and muscle stem cells and has demonstrated the sufficiency of DNA damage to drive early aging phenotypes. For some excellent reviews of the subject (Williams and Schumacher 2017; Niedernhofer et al. 2018). While much of this work is not exclusively on stem cells, collectively these studies demonstrate that unrepaired DNA damage can perturb general cellular function in a number of ways including: 1. Leading to cell cycle arrest, 2. Driving apoptosis or cellular senescence, 3. Physically disrupting transcription (Garinis et al. 2009), 4. Causing large transcriptomic changes including

growth signaling and metabolic pathways (Edifizi et al. 2017), 5. Altering chromatin organization through relocalization of factors to DNA damage sites (Oberdoerffer et al. 2008).

This work raises the question of whether endogenous levels of DNA damage can impact aging and if so, by which mechanisms. Several studies demonstrate a link between increased cellular senescence and stem cell functional decline during aging. An increase in the expression of the senescence-associated cyclin-dependent kinase inhibitor, *p16^{INK4a}*, was observed during aging in HSCs, NSCs, pancreatic islet cells, and muscle stem cells, accompanied by a decreased functionality of these stem cell populations during aging that was ameliorated in *p16^{INK4a}*-/- mice (Janzen et al. 2006; Krishnamurthy et al. 2006; Molofsky et al. 2006; Sousa-Victor et al. 2014). While these data support the notion of *p16^{INK4a}*-dependent effects on stem cells, it should be noted that *p16^{INK4a}* need not be activated through endogenous DNA damage, but could be linked to one of the DNA damage-independent modes of *p16* activation, such as changes to chromatin (Martin and Beach 2014). Consistent with this, a loss of silencing via BMI1 repression of *p16^{INK4a}* was shown to underlie muscle stem cell senescence (Sousa-Victor et al. 2014).

Ongoing DNA damage may also result from alteration in replication kinetics. Interestingly, in mouse adult HSCs, diminished expression of MCM4 and MCM6 during aging resulted in delayed replication kinetics in aged HSCs causing replication stress. Induced replication stress in HSCs resulted in preferential killing of old HSCs, therefore providing a mechanism for functional decline of cycling HSCs during aging (Flach et al. 2014) and a likely explanation for previous observations in human and mouse HSCs, of increased marks of DNA damage during aging (Rossi et al. 2007; Rübe et al. 2011).

Despite these findings suggesting that DNA damage may impair stem cell activity, the effects on adult stem cells of persistent DNA damage versus genome mutation or DNA damage signaling, must be further teased apart. In addition, determining how different types of endogenous DNA damage or mutagenic processes impact adult stem cells will be important. Finally, future studies are needed to define tissue-specific differences in endogenous DNA damage and their effects on stem cells and niche signals.

Towards an understanding of DNA damage and mutation in adult tissues

With the recent influx of DNA sequencing of healthy human tissues with age, our views regarding the genomes of somatic cells have been radically challenged. While these studies provide a descriptive snapshot of evolving somatic genomes, the use of genetically amenable model systems will further improve our understanding of molecular causes and tissue-wide consequences of endogenous DNA damage and somatic mutations in adult stem cells.

Model systems to quantify and study spontaneous mutation in tissues

Early model system studies investigating spontaneous mutation accumulation with age *in vivo* did so exploiting a transgenic mouse and *Drosophila* lines with an integrated *LacZ* reporter gene allowing quantification of mutation at this locus (Dolle et al. 2000; Garcia et al. 2010; Busuttil et al. 2007; Giese et al. 2002; Dolle 2002). An age-dependent increase in spontaneous mutation and an intriguing tissue bias of *LacZ* mutations was found: while mostly point mutations were found in the small intestine, large genome rearrangements were found in the heart (**Figure 1.4A**) (Dolle et al. 2000). Though these studies only focus on a single, artificial

transgene, they provided an important foundation to begin to study spontaneous mutation *in vivo*.

Other studies using *in vivo* lineage tracing of mutant stem cells in the mammalian intestine allowed for the better understanding of stem cell dynamics and the fixation of mutations with age. Kozar and colleagues used mice containing a dinucleotide repeat tract within a reporter gene to mark if strand slippage happens during DNA replication that consequently resulted in an in-frame reporter gene, marking the cell. Intestinal crypts that are wholly populated by these marked mutations increased with age (**Figure 1.4B**) (Kozar et al. 2013). Similarly, in the human colonic epithelium, mutant stem cell dynamics were revealed by marking known spontaneous mutations that continuously label the ISCs. Interestingly, the authors found much slower kinetics of crypt clonality likely due to slower stem cell turnover in humans compared to mouse (Nicholson et al. 2018).

Our lab has recently developed a powerful model system to investigate spontaneously arising mutations in adult intestinal stem cells in *Drosophila* (Siudeja et al. 2015). A detailed description of the adult intestine and intestinal stem cells will be given below in Chapter 1.3. In our past studies, we observed that during aging, spontaneously arising intestinal neoplasia develop in around ~12% of adult males over a rapid period of 6 weeks of adult life. Through application of whole-genome sequencing, we could demonstrate that these arise largely due to structural variants deleting regions of the *Notch* gene. As *Notch* is X-linked and present in a single copy in males, loss of one copy is sufficient to fully inactivate *Notch* and block proper stem cell differentiation, thereby resulting in the accumulation of large clonal masses of stem cells (**Figure 1.4C**). In addition, we uncovered a second means of genome alteration through loss of heterozygosity (Siudeja and Bardin 2017a; Siudeja et al.

2015). Understanding the mechanisms regulating loss of heterozygosity will be the main subject of my results 2.1 below. These data suggest that spontaneous mutation occurs frequently in *Drosophila* adult intestinal stem cells, making them a useful model to decipher underlying causes and consequences of stem cell somatic mutation on adult tissues.

Important advantages of this model include the rapid acquisition of mutations over 6 weeks of aging, the application of whole-genome sequencing, abundant genetic tools, and the ability to alter environmental conditions.

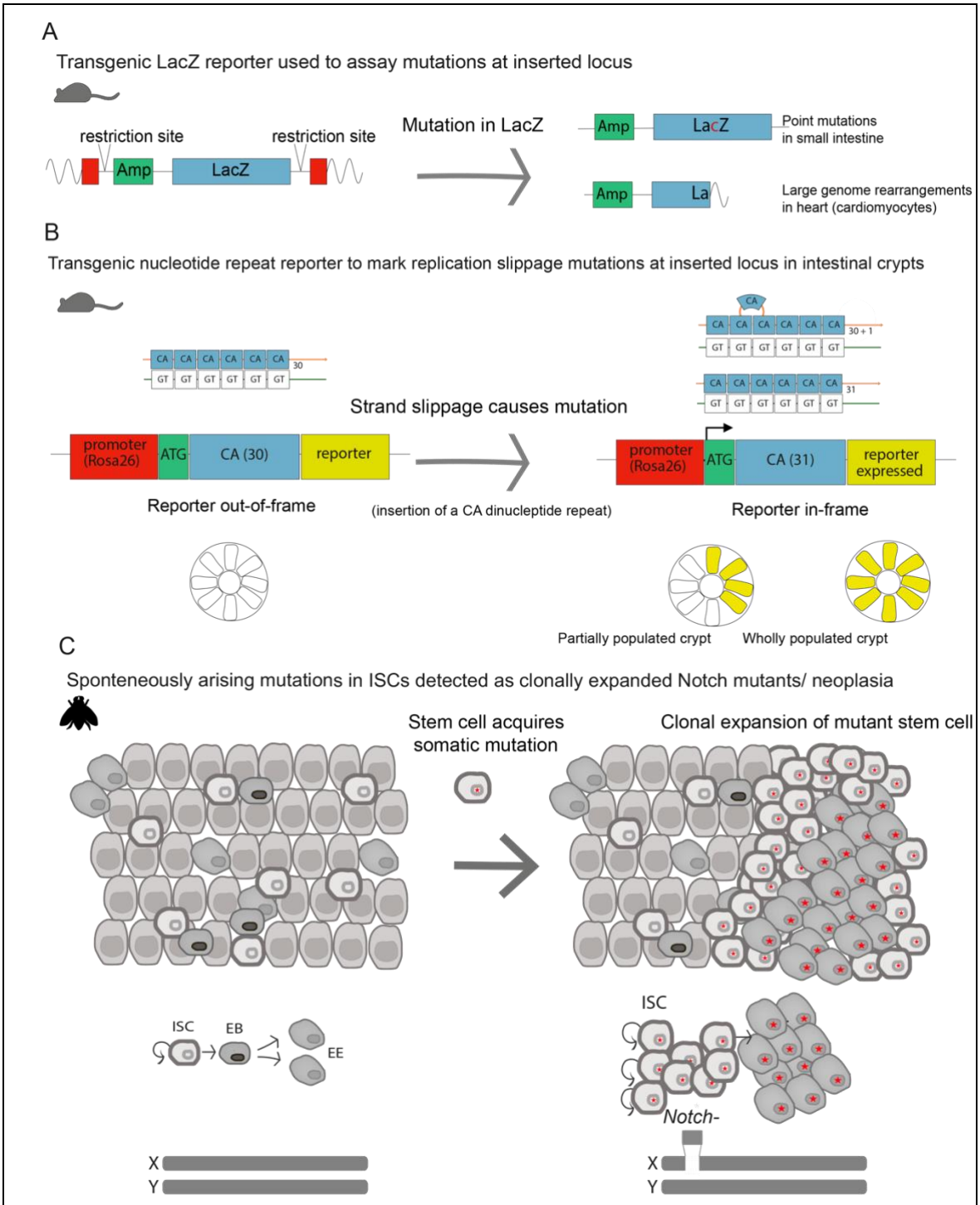


Figure 1.4: The use of model systems to assay spontaneous mutations in adult tissues.

(A) A LacZ reporter to assay spontaneous mutations (Dolle et al. 2000). In this assay, genomic DNA extracted from young/old hearts and small intestines of transgenic mice containing the LacZ reporter is assessed for spontaneous mutation. LacZ encoding plasmids are recovered in bacteria tested for intact LacZ activity. A significant age-related increase in point mutations was detected in small intestines, whereas a significant age-related increase in genomic rearrangements was detected in hearts.

(B) A transgenic nucleotide repeat reporter assay to mark spontaneous mutations in mouse intestinal crypts (Kozar et al. 2013). In this system, spontaneous slippage of the cassette

during replication allows for expression of the reporter. Wholly populated crypts indicative of the fixation of mutations were shown to increase in an age-dependent manner.

(C) Spontaneously arising somatic Notch mutations in Drosophila intestinal stem cells (ISCs) can be detected as clonally expanded Notch mutants/ neoplasia in aged flies (Siudeja et al. 2015). Inactivation of Notch leads to neoplasia with an accumulation of ISCs and enteroendocrine cells. Whole-genome sequencing of aged male neoplasia revealed that Notch is inactivated via deletions or structural rearrangements.

Alternative model systems: diverse evolutionary strategies of somatic genome stability

In addition to fly and mouse models, other models are providing important advances in our understanding of effects of somatic DNA damage and mutation on adult stem cells and tissues. Active work in *C. elegans* has led to insight into systemic effects of somatic DNA damage (Mueller et al. 2015; Williams and Schumacher 2017). Further investigation in alternative invertebrate models such as planaria, hydra, and non-traditional vertebrate models such as the naked-mole rat may provide surprising solutions to how organisms cope with DNA damage or somatic mutations. Hydra and planaria, for example, have stem cells with an unlimited capacity for self-renewal and do not show signs of aging (Boehm et al. 2013) and the naked-mole rat is a long-lived vertebrate that is cancer resistant (MacRae *et al.*, 2015; Petruseva et al. 2017). Probing into mechanisms in these models may yield unanticipated new insight into potential ways to mitigate the negative effects of mutation.

Concluding remarks

Alterations to the DNA of stem cells can disrupt their efficient self-renewal and differentiations, consequently changing the status quo of different tissues,

eventually impacting aging and cancer initiation. Cell- and tissue-specific mechanisms by which stem cells protect themselves from damage and mutations thus exist. Despite these protection mechanisms, damage and acquisition of mutation occur. Ironically, somatic mutation and errors in the DNA repair process are capitalized on in the soma in some instances such as the generation of antibody diversity in vertebrates, reviewed in (Li et al. 2004) or in programmed genome rearrangement occurring in lamprey, actively eliminating potentially harmful germline genes from the soma (Smith et al. 2018; Wang and Davis 2014).

Over recent years, advances in the technology to detect mutations and rare events in asymptomatic healthy tissues have revealed the sobering fact that our tissues are peppered with mutations. Healthy tissues are actually mosaics of cell lineages derived from mutant stem and precursor cells. Future studies will better define the forces of selection in healthy tissues and how these relate to cancer. An additional challenge will be to unveil the functional impact of accumulating mutations, linking genotype to diseases and aging phenotypes.

Fundamental questions remain regarding how DNA damage and somatic mutation of stem cells can be manipulated to slow down aging and delay or evade cancer initiation. How can genomic damage be prevented from accumulating in stem cells? Might mechanisms of cell competition be harnessed to replace potentially harmful mutant cells with therapeutic cells? Can tissue extrinsic factors such as changes in the environment be manipulated to control clonal expansions? Indeed, these questions remain open today and will be active areas of research benefitting from studying diverse genetic model systems.

In the next section, chapter 1.2, I will provide a closer insight into one of the consequences of DNA mutation: loss of heterozygosity (LOH), which has been alluded to earlier in this section.

1.2 Loss of heterozygosity (LOH): a common cause of genome alteration in somatic cells

As the cells in our tissues are being renewed throughout life, our somatic cells should faithfully preserve the integrity of DNA. In the previous chapter, I highlighted that the integrity of DNA is under constant threat, leading to many possible genome alterations in somatic cells. I also mentioned that despite the threats, there are a number of mechanisms at the cellular- and tissue-level maintaining a degree of protection, especially for the stem cell. I would like to mention here that at the broader organismal level, a degree of protection is also provided through diploidy, where having two copies of each gene, protects against the effects of somatic mutation (Crow and Kimura 1965; Perrot et al. 1991; Otto and Goldstein 1992; Mable and Otto 2001). If a mutation arises in one copy, the second wild-type copy provides a backup, maintaining function. Thus, the heterozygous state masks the effects of recessive deleterious mutations, (with the notable exception of haploinsufficient genes).

In this chapter, I explain the loss of the protective heterozygous state, which can lead to cancers, pathological disorders but also occurs in normal human tissues. In particular, I detail what is known about the mechanisms that can lead to the loss of heterozygosity (LOH). Importantly my PhD work aimed to elucidate some of how LOH occurs in adult stem cells.

What is LOH?

In strict molecular terms, “heterozygosity” refers to a state in which an allele has a different DNA sequence, however the “loss of heterozygosity” usually refers to the loss of the functional wild-type copy at a heterozygous locus. An individual

can be heterozygous at any given locus due to a germline mutation, or a somatic event. LOH is thus especially problematic if the heterozygous locus is a tumour suppressor gene, as LOH leads to unmasking its deleterious effects.

Fifty years ago, Alfred Knudson's observations of paediatric retinoblastomas led him to note that inherited forms of retinoblastoma occurred more frequently in younger individuals compared to sporadic retinoblastoma. He thus reasoned that tumours result from two "hits" (two mutations), whereby some individuals can be born with a germline mutation in the *Rb* gene and acquire the second mutation of the wild-type allele (the second hit) somatically; explaining the early onset. In the nonhereditary cases, the tumour could be explained by two successive somatic mutations of the same cell, explaining the later onset (Knudson 1995). This phenomenon was then established as "Knudson's two-hit hypothesis", providing an explanation that tumour development requires an additional step beyond inheritance of tumour susceptibility. It also provides another definition where "second hit" is synonymous with LOH. Since then, many studies have looked into the implications of LOH in cancer genomes.

How is the second wild-type allele somatically inactivated?

The "second hit" or LOH of the wild-type can arise by a number of means: (1) It can be inactivated via point mutation, (2) a deletion, (3) it can be lost through aneuploidy and (4) mitotic recombination (MR) resulting in the co-segregation of two mutant alleles into a daughter cell. While LOH via the direct inactivation of the wild-type allele via a point mutation or small deletions has been documented in various cancers, the mechanisms which give rise to these single locus events usually lead

to fewer genetic changes in the context of LOH than that of multi-locus chromosomal events such as large deletions, aneuploidy and MR.

It has been shown, that MR, plays a substantial role in both sporadic and familial cancers as well as other pathologies, thus I will begin by detailing what is already known about MR-driven LOH. Towards the end of this section, I will also touch on another multi-locus event LOH mechanism: aneuploidy.

Mitotic recombination-driven LOH

Mitotic recombination (MR), is homologous recombination (HR) that takes place during interphase of the mitotic cell cycle and not during *mitosis*, as the name suggests. It is defined as the homology-directed DNA exchange between sister or homologous chromosomes and comes into play to repair DNA double strand breaks (DSBs) using the intact chromosome as a template. The use of a template makes MR a high-fidelity DNA repair mechanism compared with alternative pathways of DSB repair such as NHEJ and alt-EJ, which introduce deletions along with the repair (explained in chapter 1.1). MR however, in spite of the more accurate repair, is clearly a double-edged sword. In cells heterozygous for a tumour suppressor gene (as mentioned above), MR corrects the DSB, but at the cost of recombining out the functional wild-type, thus leading to tumour suppressor gene inactivation establishing the very first steps of cancer.

Cancer initiation and mitotic recombination-driven LOH

In the particular case of *adenomatous polyposis coli* (*APC)* mutation carriers, patients develop “familial adenomatous polyposis” as the germline mutation in *APC* usually manifests in the appearance of polyps, which are small abnormal tissue growths on the surface of the colon. *APC* regulates β-catenin, a multifunctional

protein that plays a vital role in cell-cell communication, growth and signaling. Although these polyps are benign, it is the second hit LOH of the remaining *APC* copy that facilitates the generation of polyps and initiates the cascade of events attributed to the multistep carcinogenesis of the colon, where the oncogene *KRAS* gets activated leading to further subsequent inactivations of other tumour suppressor genes such as P53. Studies from human cell lines derived from familial polyposis patient tumours identified that the LOH of *APC* occurs via MR (Cottrell et al. 1992; Haigis et al. 2002; Thiagalingam et al. 2001; Howarth et al. 2009). Mouse models been developed modeling intestinal cancers using (Apc^{Min}/+) mice also show MR as a mechanism driving APC LOH (Haigis et al. 2002). Additional studies on human cell lines have also provided evidence of MR-driven LOH of other tumour suppressor genes such as retinoblastoma (*Rb*) (Cavenee et al. 1983), and neurofibromatosis *NF1* (Serra et al. 2001), also reviewed in (Tuna et al. 2009; Lapunzina and Monk 2011; Siudeja and Bardin 2017). MR is therefore a frequent means of tumour suppressor gene inactivation, particularly in familial cancers where a germline mutation is pre-existing.

Restoration of the wild-type genotype through MR-driven LOH

Interestingly, in the same way that MR can lead to the loss of the wild-type allele, it can also have a beneficial role in cases where a dominant mutant allele is lost instead, rescuing mutant phenotypes in heterozygotes. This phenomenon has been observed in Icthyosis (Choate et al. 2010), an autosomal dominant disease causing dry and scaly skin patches where the *KRT10* dominant mutation is spontaneously inactivated via MR and the wild-type is restored. The MR-driven spontaneous elimination of the mutant *KRT10*, and the restoration of wild-type,

gives rise to “revertant patches” leading to a natural form gene therapy. This phenomenon has been observed in other skin diseases too (Kiritsi et al. 2012; Jonkman et al. 1997). MR-mediated somatic reversion has also been well described in a metabolic disorder causing immunodeficiency (Hirschhorn et al. 1996) as well as blood disorders (Revy et al. 2019; Jongmans et al. 2012) including Diamond Blackfan Anemia (DBA) where case studies reveal disappearing features of anemia in patients as a result of MR-based somatic reversion to wild-type phenotypes (Jongmans et al. 2018; Venugopal et al. 2017) and a subsequent clonal expansion of the revertant cell, which lost its dominant mutation through MR.

Thus, a thorough understanding of the molecular mechanisms giving rise to MR and what drives it, particularly in cell types that can clonally expand such as stem cells, will provide important insight into cancer initiation for potential prevention strategies and also insight into somatic reversion strategies, rescuing pathogenic phenotypes. Presumably, MR is being driven by repair of a DSB. Important questions remain to be addressed: What leads to the DSB? Do DSBs happen in particular regions in the genome such as fragile sites? Are there environmental factors that increase the chance of DSB occurrence?

DSBs: Drivers of MR

While it is known that meiotic recombination is driven by programmed DSBs initiated by the topoisomerase-related Spo11 protein (de Massy et al. 1995; Keeney and Kleckner 1995; Liu et al. 1995; Bergerat et al. 1997; Keeney et al. 1997), the drivers of mitotic recombination are less known. Spontaneous cellular events leading to DNA damage are likely to be drivers, however whether these events leading to DNA damage are random or non-random is unclear. What is clear,

however, is that resultant spontaneous DSBs in mitosis are pathological, rather than physiological as in the case of programmed DSBs in meiosis.

There have been numerous efforts to understand how MR arises in somatic tissues, but, MR for the most part has been observed using low-resolution techniques such as restriction fragment length polymorphism (RFLP) markers or microsatellite loci along the chromosome arm and assayed using PCR-based methods or targeted sequencing. These markers have a significant distance between them and do not meet the required resolution to gain mechanistic insight from mapping recombination sites and learn more about where the DSB arose and the cause of repair. In one study by Howarth et al, that used SNP microarray to profile the SNPs in order to map sites of MR in human tumours caused from the LOH of *APC*, sites of MR were resolved to around 4.5 Mb. The authors claim that the sites were non-random and possibly associated with low copy repeats (LCRs), however the poor coverage of polymorphic markers did not allow finer-scale mapping (Howarth et al. 2009).

Finer mapping involves 1. a high density of polymorphic markers along the chromosome and 2. a good technique to detect the all the polymorphic markers possible along the chromosome. This can be achieved by 1. using organisms that have parental genotypes with a sufficient sequence divergence such as high informative SNP density and 2. assaying the informative SNPs using high-throughput sequencing. An increasing number of studies have consequently been produced in yeast producing fine maps of spontaneous MR due to the feasibility of achieving the aforementioned two points (St. Charles and Petes 2013; Lee et al. 2009; Sweetser et al. 1994; Yin et al. 2017). It was shown that in yeast, spontaneous MR sites were non-random, occurring near inverted repeats of Ty transposable

elements (St. Charles and Petes 2013; Sweetser et al. 1994). Whether these sequence features drive MR in somatic cells of higher eukaryotes remains elusive however. Though there is a common thread of “repetitive DNA elements” between the association found with LCRs in humans, and the association found with Ty inverted repeats in yeast. Repetitive DNA is capable of forming secondary structures, such as hairpin or cruciform non B-form DNA structures that can be an obstacle for a replication fork, causing it to stall or collapse, leading to a DSB. Other cis-acting elements such as highly transcribed DNA sequences can also present obstacles for the replication fork, reviewed in (Aguilera and Gómez-González 2008). Despite these studies, further insight is needed to define how underlying genomic sequences can impact DNA DSB and sites of MR.

Yeast: a paradigm for studying mechanisms of MR

In addition to the high-resolution mapping and potential insights into causes of DSBs, most of what we know about the succeeding steps after the broken strands comes from budding yeast *Saccharomyces cerevisiae.* By studying what happens after spontaneous as well as the synchronous induction of DSBs by site-specific endonucleases. Using a combination of genetic, molecular and cytological approaches, yeast has provided substantial insights into the different mechanisms of MR (Dé et al. 1999; Hicks et al. 2011; Baumgartner et al. 2018b). The possibility to recover products for a reciprocal crossover though clone sectoring events, analogous to meiotic tetrad analysis (Barbera & Petes 2006), has also led to yeast becoming such a powerful model system to understand MR.

Through studies in yeast, we know that LOH by MR comes in three flavours:

1. Gene conversion via synthesis dependent strand annealing (SDSA): which results in short tracks of LOH.
2. Break-induced replication (BIR) involving the recruitment of replication machinery to copy the homologous chromosome.
3. A cross-over (CO) event mediated by the formation of a double-holliday junction allowing for the exchange of arms between homologous chromosomes. The latter two modes, BIR and CO both result in long track LOH.

I will go into the three modes of MR in more detail below.

Molecular insights into the repair mechanisms of MR

A double strand break is when both strands of duplex DNA are severed. Here I describe how the severed/broken strands find a template and undergo repair via homologous recombination (HR). Though I would like to point out that HR is one of 3 pathways that can solve a DSB problem. Please see chapter 1.1 under “DNA damage and how it leads to mutation”, for a further description of the three pathways of DSB repair- HR, NHEJ and alt-EJ, whose utilisation depend largely on the timing of the DSB relative to the cell cycle stage.

Recombination mechanisms proceed as follows (**see Figure 1.5**): after DNA strands are broken, the resection of the broken ends by an endonuclease occurs in a 5’-3’ direction, resulting in 3’ overhangs (**Figure 1.5A**). These overhangs have a 3’-OH group exposed to prime DNA synthesis using a homologous region as a template. The subsequent search for homology relies on a recombinase protein called Rad51 that assembles on the overhangs and forms a nucleoprotein filament. The nucleoprotein filament facilitates Rad51 to search the entire genome for

homology until it finds its homologous match and initiates strand invasion to mediate the exchange of base-pairs. The strand invasion establishes the formation of a displacement loop, also known as a “D-loop” (**Figure 1.5A**). At this point, the choice is then made between the aforementioned three modes (SDSA/ BIR/ CO).

1. SDSA (**see Figure 1.5B**): The strand invasion leads to the copying of the template, but the new strand is displaced from the donor and captured by the second end of the DSB. Thus, through this pathway, a small patch of newly synthesised DNA is created to directly repair the DSB. This small patch of DNA is also called a patch of “gene conversion”. Gene conversion is defined as “the non-reciprocal transfer of genetic information between homologous sequences” (Haber 2007). LOH will only be detected in this case if the wild-type allele encompasses the region of the small patch of synthesis that corresponds to gene conversion.

2. BIR (**see Figure 1.5C**): BIR takes place when only one end of the DSB shares enough homology with donor region. DNA polymerase delta drives replication with Pif1 helicase unwinding the DNA and proceeds by a migrating D-loop in a long-leading strand intermediate, that is then completed by lagging strand synthesis (Saini et al. 2013). This has been shown to be a highly mutagenic process as the accumulation of ssDNA before the completion with lagging strand synthesis is sensitive to nucleases, resulting in mutation hotspots by the breakpoints. One study showed a median density of 1 mutation per 6.3kb, which is 900 times higher than the density of scattered mutations across the genome (Sakofsky et al. 2014) and another

study reported mutations increasing up to 2,800 fold compared to spontaneous events (Deem et al. 2011). Additionally, a higher rate of mutagenesis also extends far from the breakpoints, this is because dNTPs are elevated during BIR, which contributes to lower fidelity of the DNA polymerase (Deem et al. 2011).

3. CO (**see Figure 1.5D**): If the second strand of the DSB is captured, a structure called a double Holliday junction is formed. Small patches of DNA synthesis take place to fill in the gaps and the double-Holliday-junction moves via branch migration, which can then be dissolved forming non-crossover products and an outcome similar to that of SDSA, or the double-Holliday-junction can be cleaved by nucleases called resolvases in an orientation mediating crossover products. Intervening regions of conversion tracts near by the initiation breakpoint correspond to the small patches of DNA synthesis. The distance of these conversion tracts/ patches of non-reciprocal synthesis from site of recombination corresponds to how much branch migration took place as well as whether mis-match repair corrected heteroduplex sequences. Depending on the DNA strand, the resulting repaired chromosome can either contain one long LOH tract, or a long LOH tract with intervening regions of gene conversion nearby the initial breakpoint.

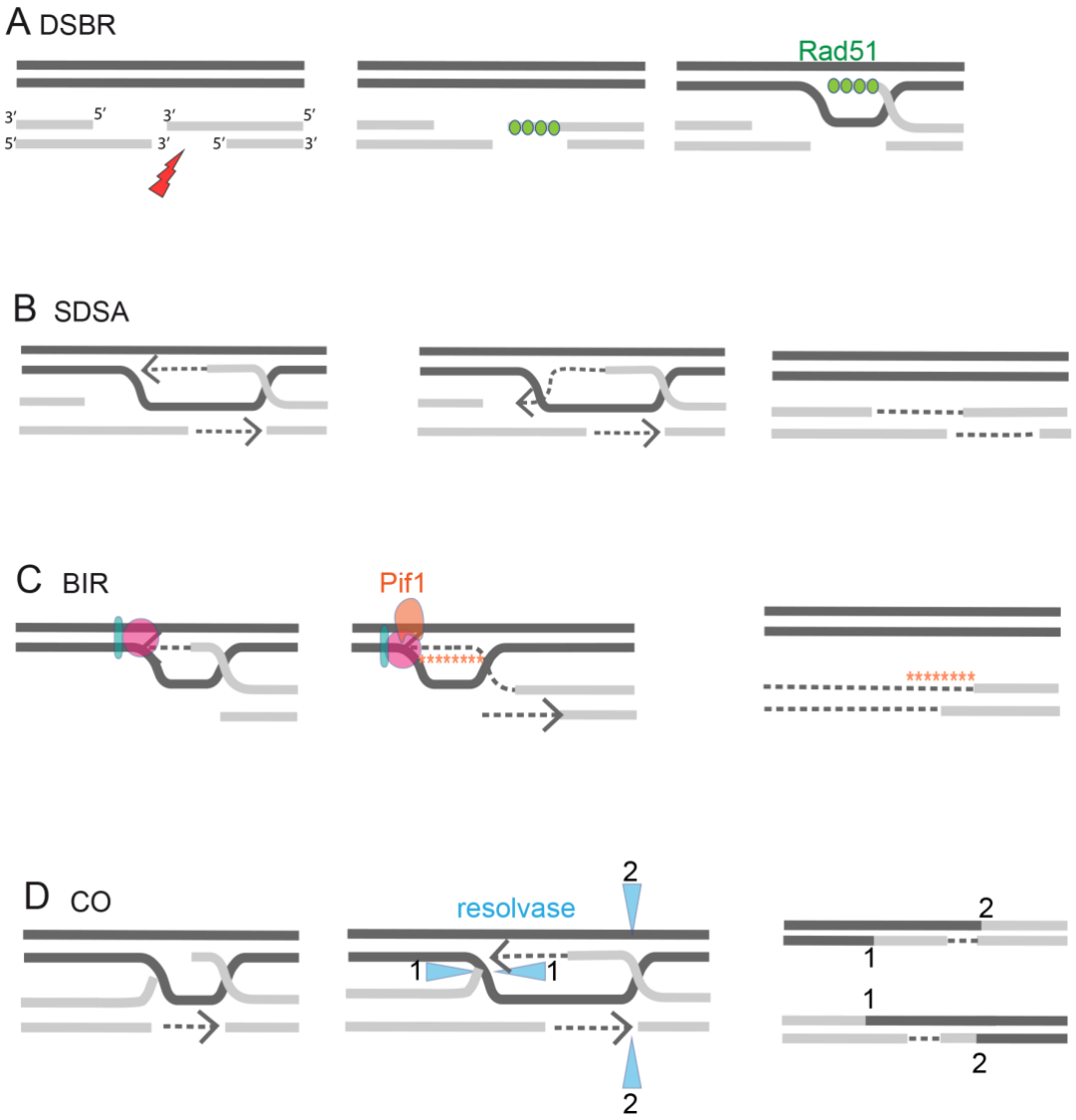


Figure 1.5: Double strand break repair pathways via HR

(A) The double strand break repair (DSBR) models all begin with the same substrate: 3' overhangs of ssDNA resulting from a DSB. Rad51 forms the nucleoprotein filament that initiates the search for DNA homology and strand invasion. A displacement loop (D-loop) is formed between the 3' overhang and the homologous chromosome strand (donor). These steps with shown here are common steps between all the DSBR models. The models then deviate after this point.

(B) In the SDSA pathway, the invading strand copies a short part of the donor through producing a patch of DNA synthesis (denoted by the dark grey dotted arrow). The most important step that distinguishes it from the other pathways is the fact that the new strand is displaced from the donor and is captured by the second end of the DSB (the 5' end). This leads to non-crossover products NCO, but a short patch of non-reciprocal DNA exchange within the region of synthesis.

(C) Break-induced replication (BIR) repairs the DSB by copying the donor chromosome. It relies on unwinding of the donor DNA by the Pif1 helicase. Leading strand synthesis copies the template followed by lagging strand synthesis off of the leading strand to produce a new end of the chromosome from where the DSB occurred. It is an error-prone process resulting in local de novo mutations (red stars) near the DSB origin.

(D) A CO mechanism is mediated by a double-Holliday junction 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.

While both BIR and CO mechanisms would lead to long stretches of LOH spanning the chromosome arm and depend on Rad51, they differ in several aspects: Firstly, since BIR involves the copying of the template with no reciprocal exchange, after chromosome segregation and cell division, the product will be one daughter with LOH and a second daughter that is heterozygous. As for the model involving CO, the double-Holliday junction resolution would result in 2 altered homologous chromosomes with reciprocal exchange events being inherited in the resulting daughter cells leading two daughters with LOH. Secondly, both models lead to different signatures after repair, with BIR showing mutation pileup by the breakpoints, as well as elevated levels of mutations across the chromosome) and half crossover products having an intervening stretch of conversion tract.

With this knowledge about MR mechanisms in yeast, an important question that ensues is: to what extent are these processes conserved in higher eukaryotes? It is known that there is high conservation in the proteins required for DNA repair from yeast to humans (Cromie et al. 2001), as the importance of DNA repair throughout evolution is clear, but which MR mechanism is more predominant in driving LOH by MR in humans for example? Attaining knowledge of this can have implications on targeting these mechanisms for potential therapies.

Another cause of LOH: Aneuploidy

Another somatic multi-locus event leading to LOH can arise from somatic aneuploidy. Aneuploidy is a state in which cells have an abnormality in ploidy, meaning that they have an unbalanced number of chromosomes. In the case of

chromosome loss, the “second hit”/LOH is driven by the loss of the wild-type copy of a heterozygous allele along with the whole chromosome loss (monosomy), alternatively, aneuploidy can also result in chromosome gain, where the notion of “zygosity” no longer applies with the increase in ploidy.

Whole chromosome loss is frequent in cancer genomes (Bignell et al. 2010; Lin et al. 2003; Duijf et al. 2013). Consistent with the potential of aneuploidy to drive cancer formation, mutations in mitotic checkpoint genes have been shown to give rise to spontaneous tumours through LOH driven by chromosome loss (Baker et al. 2009; Baker and Van Deursen 2010; Tighe et al. 2001).

Aneuploidy is the result of faulty mitosis, leading to improper segregation of whole chromosomes, and therefore arises in a very different manner than MR discussed above. While MR relies on the relationship between DSB and homologous template in S phase, aneuploidy on the other hand relies on the relationship between chromosome and spindle during mitosis. Here, I provide a quick overview of how aneuploidy arises, reviewed in (Chunduri and Storchová 2019; Funk et al. 2016).

Aneuploidy can occur via various means

Aneuploidy can arise from defects in the mitotic checkpoint. The mitotic checkpoint employs surveillance pathways to ensure that one copy of each chromosome is distributed into each daughter nucleus. At this checkpoint, if a replicated sister is not correctly attached to the spindle microtubule by the kinetochore, the irreversible transition into anaphase is delayed. If this checkpoint is “weakened” the cell fails to notice that a chromosome pair is not lined with the spindle apparatus properly and this could lead to an aberrant set of genetic material.

In addition, “merotelic attachments” can arise when a single kinetochore is attached to microtubules emanating from both spindle poles do not trigger checkpoint activation and are a major cause of lagging anaphase chromosomes (Cimini et al. 2003). Similarly, a hyperstablised kinetochore-microtubule interaction can keep chromosomes attached, resulting in a lag in segregation in anaphase, which if not corrected upon checkpoint activation, causes chromosome missegregation.

In addition to a lack of checkpoint activation, there is mounting evidence showing that mitosis can go wrong due to the deterioration of sister chromatid cohesion. Within each pair of chromosomes, the sister chromatids are held together by the cohesion complex at their centromeres and along the chromosome arms. A failure to maintain cohesion, can lead to attachment of both sister chromosomes to the same centrosome, resulting in missegregation. A final mechanism leading to aneuploidy, is centrosome amplification. The aberrant production of additional centrosomes, that capture chromosomes and cluster into the same cell, can cause defective chromosome segregation and aneuploidy.

All of the above are potential causes of abnormal numbers of chromosomes in the cell. The wrong chromosome complement can shift the dosage of genes in the cell and consequently impact the proteome. Although it has been shown that aneuploidy causes proteotoxic stress and reduces cellular fitness, cancer cells clearly tolerate aneuploidies, this is known as the “aneuploidy paradox”, reviewed in (Sheltzer and Amon 2011). Thus, while arising from defective cell divisions, aneuploidy can also lead to loss of heterozygosity and affect large regions of the genome, with additional consequences on the proteome.

Concluding remarks

In section 1.2 of my thesis, I introduced how LOH driven by MR can have a substantial impact on adult tissues. While there have been efforts to understand how MR arises in somatic tissues, MR for the most part has been observed using low-resolution techniques. Finer mapping provides an understanding of features in the genome that could drive spontaneous MR and its mechanistic signatures. Although yeast has illuminated sequence features and different mechanisms driving MR, the question remains of whether they are conserved and operate in adult stem cells that fuel tissue homeostasis (as discussed in 1.1). In addition to that, questions related to how factors in the environment can contribute to MR-driven LOH will be a valuable insight to the field. Thus, there is a need to fill the gaps in knowledge linking the findings from unicellular yeast and higher eukaryotes to understand pathways that promote and prevent MR in a complex tissue *in vivo*. An *in vivo* system in which we can test hypotheses and see the impact of changing variables will be important to elucidate MR from genomic initiation to tissue drivers.

Additionally, I introduced how LOH can be driven by aneuploidy and highlighted that whilst aneuploidy should, in theory, debilitate cellular proliferation, it paradoxically promotes tumour progression. This raises questions regarding compensatory mechanisms that may come into place amending the imbalance in genetic material and restoring proteomic homeostasis. Thus, a system is needed to test spontaneous aneuploidies and potential buffering mechanisms that could equip the cell with tolerance for an abnormal cell complement.

1.3 The *Drosophila* intestine: A model to study genome alterations such as LOH in stem cells

In 1.1, I told you about the importance of studying genome alteration in stem cells fueling tissue homeostasis, establishing tissue mosaicism that may contribute to aging phenotypes, premalignancy or malignancy. In 1.2, I covered a common way in which genome alteration can arise: LOH via mitotic recombination (or aneuploidy) and emphasised the questions that remain elusive in higher eukaryotes regarding genomic initiation of LOH, its mechanisms and what drives it in complex tissues. In this section, I move onto the model system we use in the lab, the Drosophila intestine, to address some of the questions regarding LOH in an in vivo stem cell model.

Here, I will briefly discuss the structure of the adult Drosophila intestine, its cell types, the cell-cell communication that specifies its cell lineage. I will then go on to describe some changes that are accompanied with aging and the extrinsic factors that can impact it such as the external environment. I will finally end this section highlighting the advantages of using this model system by briefly describing the genetic tools available making this system a powerful genetic model to address the questions in 1.1 and 1.2.

A dynamic tissue in a powerful *in vivo* model

Nutrient absorption and digestion are key components for growth and maintenance in all Animalia ranging from flatworms to humans (Hartenstein and Martinez 2019). The intestine is the portal for nutrient entry and is thus constantly exposed to food and in contact with pathogens, which can contribute to its wear and tear with time. Consequently, ensuring homeostasis is important to maintain its integrity for long-term function. The tissue needs to be tightly regulated to

compensate for the cells lost due to wear and tear and aging to satisfy the tissue's needs. The intestine is in fact the most highly regenerative organ in the mammalian body, renewing its epithelium every 3-4 days (Karin and Clevers 2016). Similar to the mammalian intestine, the *Drosophila* intestine also shares many common aspects of intestinal physiology and homeostasis and renews once every 1-2 weeks (Jiang and Edgar 2011). Studies with this model have made seminal contributions in the fields of stem cell biology, tissue homeostasis, organ physiology, and aging (Gervais and Bardin 2017). The lab has been using *Drosophila*, with its powerful genetic tools to address questions about stem cell regulation and deregulation during aging (**Figure 1.6A**).

Structure of the *Drosophila* intestine

The adult *Drosophila* intestine is a tube lined by an epithelial monolayer making up the foregut, midgut and hindgut (**Figure 1.6B**) comprised of different cell types (**Figure 1.6C, D**). The posterior midgut and hindgut are proposed to be homologous to the vertebrate small intestine and large intestine respectively (Miller 1994, 1965; Lehane and Billingsley 1996) reviewed in (Lemaitre and Miguel-Aliaga 2013; Miguel-Aliaga et al. 2018).

The midgut is the most studied intestinal region of the adult fly. It has three areas that are morphologically distinct: the anterior, middle and posterior midgut (Li et al. 2013). In addition to that, further compartmentalisation was evidenced in 2013 in a study that showed 6 regions (R0-R5, see **Figure 1.6B**) with distinct genetic properties, morphology and histology (Marianes and Spradling 2013; Buchon et al. 2013; O'Brien 2013). R3 for instance is known as the "copper cell region" (CCR) that has a higher acidity compared to the rest of the midgut (**Figure 1.6B**) and also

has different cell types (**Figure 1.6E**). The CCR controls the distribution and composition of the resident bacteria/ microbiota of the gut (Li et al. 2016).

Cells of the Drosophila midgut

The midgut is mostly made up of large absorptive polyploid cells called enterocytes (ECs), akin to their counterparts in the mammalian intestine, which are one of the two terminally differentiated cells in the tissue (**Figure 1.6C**) and absorb nutrients. The apical luminal surface of ECs is covered by microvilli that are rich in F-actin, α-spectrin, βH-spectrin, myosin-II (Baumann 2001). However, since ECs are the only polyploid cells in the tissue, their large nucleus distinguishes them from the other smaller diploid cells which are lodged in between.

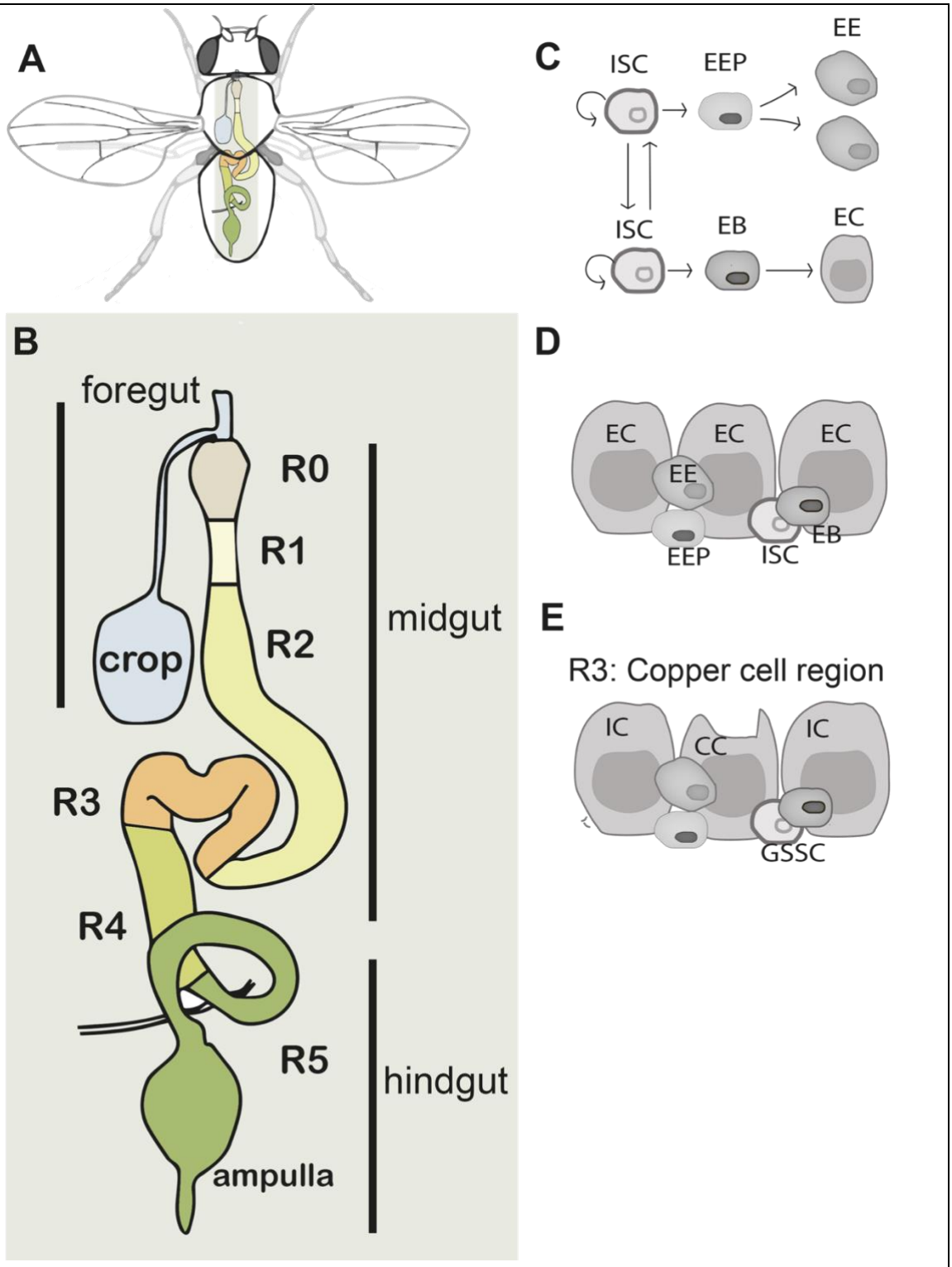


Figure 1.6: Structure and cells of the *Drosophila* midgut

(A) The *Drosophila* intestine is a regenerative tissue with distinct regions shown in (B).
(B) The *Drosophila* intestine is composed of a foregut, a midgut and hindgut and is subdivided into regions R0-R5.
(C) The intestinal stem cell (ISC) lineage. ISCs fuel tissue homeostasis of the midgut by dividing asymmetrically, self-renewing and giving rise to another cell

that is a progenitor; either an enteroblast (EB) or an eneteroendocrine precursor (EEP) cell that will further divide into two eneteroendocrine cells (EEs). The EB differentiates into an enterocyte (EC), which constitutes most of the intestine.
(D) The cells of the *Drosophila* midgut form a monolayer that is mostly made of large secretory polyploid ECs. The smaller diploid cells lodged in between are the secretory EEs, progenitors and ISCs.
(E) The R3 region of the *Drosophila* intestine is called the copper cell region (CCR). It is distinct from the rest of the region because of its higher acidity and it plays a major role in the distribution of resident bacteria.

The second differentiated cell type in the tissue is the diploid secretory enteroendocrine (EE) cell, similar to secretory EEs of the mammalian digestive tract. EEs translate local signals into systemic responses by secreting peptide hormones such as Allostatin and Tachykinin into the bloodstream (Lehane and Billingsley 1996; Ohlstein and Spradling 2006a; Scopelliti et al. 2014) and express the transcription factor Prospero (Pros) (Ohlstein and Spradling 2006a; Micchelli and Perrimon 2006a).

The midgut is maintained by intestinal stem cells (ISCs) that fuel the tissue homeostasis throughout the life of the fly. For some time, there was some uncertainty about the presence of ISCs as early reports created debate over whether there are somatic cell divisions occurring in the adult midgut (Bozcuk 1972). Bozcuk and colleagues showed that H-thymidine incorporation experiments revealed DNA synthesis in the *Drosophila* midgut but did not reveal mitotic figures with aging, suggesting that endoreplication and not DNA synthesis is taking place and that the midgut is composed of fixed, post-mitotic cells (Bozcuk 1972). This contrasted with other studies showing that many insects, morphologically similar to *Drosophila,* have midguts that comprise regenerative cells to maintain epithelial integrity through homeostasis (Snodgrass 1935; Baldwin and Hakim 1991). Later, another study in *Drosophila* stated a distinct smaller cell type interspersed in the intestinal epithelium that showed different subcellular characteristics and staining

than enterocytes. These cells were described as “regenerative cells” in analogy to other insects, though lacking functional data that they performed regenerative functions (Baumann 2001). The uncertainty was resolved however when two seminal papers were published in 2006 characterising the self-renewal properties governed by adult stem cells (Micchelli and Perrimon 2006b; Ohlstein and Spradling 2006a). These studies showed, for the first time, the characterisation of *Drosophila* midgut intestinal stem cells (ISCs) using lineage tracing techniques that showed mature cells were being lost and replaced by ISCs. This consequently opened up opportunities to utilise the *Drosophila* intestine as a model for homeostatic tissue in vivo with similarities to human/mammalian intestine thus establishing a new model system to study adult stem cells in the fly gut.

The ISC is the primary dividing cell type that, upon division, produces two cells, one that that self-renews and another cell that is a progenitor. In homeostatic conditions, the EB differentiates into an EC (80%) of the time and in 20% of the time, an ISC differentiates into an EE (Perdigoto et al. 2011; Ohlstein and Spradling 2007). More recent work showed that EE cells are produced by precursor cells called eneteroendocrine precursors (EEPs); Chen and colleagues show evidence through lineage analysis of a population of cells that undergo a second mitosis of the stem cell daughter that then go on to differentiate into EEs (Chen et al. 2018; He et al. 2018). Thus, the ISC lineage can be presented as follows (**Figure 1.6C**). But what determines this specificity? How is cell fate determined in homeostatic conditions? It all depends on the precise regulation of stem cell daughter fates which depends on cell-cell communication and Notch signaling. I will thus move on to describe how Notch signalling regulates cell fate.

The role of Notch in regulating cell fate

Notch signaling is a highly conserved paracrine cell-cell signaling pathway regulating a variety of cell fate decisions including mammalian and *Drosophila* ISCs. It has been shown that *Notch* is essential for the specification of the ISC lineage by examining loss of *Notch* activity in mosaic clones - i.e. induced mutant stem cells that generate clonal lineages - which led to an accumulation of diploid ISC and EE cells, showing clusters devoid of ECs (Ohlstein and Spradling 2006a; Micchelli and Perrimon 2006a).This is suggestive that *Notch* is required for ISC balance and differentation into ECs. Following this, other studies showed that the ligand for Notch, Delta (Dl), is endocytosed in the ISC, making it the “signal sending cell” activating the Notch receptor in the EB. It is the strength of this signal that determines the lineage choice of EC versus EE, with high *Notch* specifying EC fate and low *Notch* specifying EE fate (**Figure 1.7**) (Ohlstein and Spradling 2007).

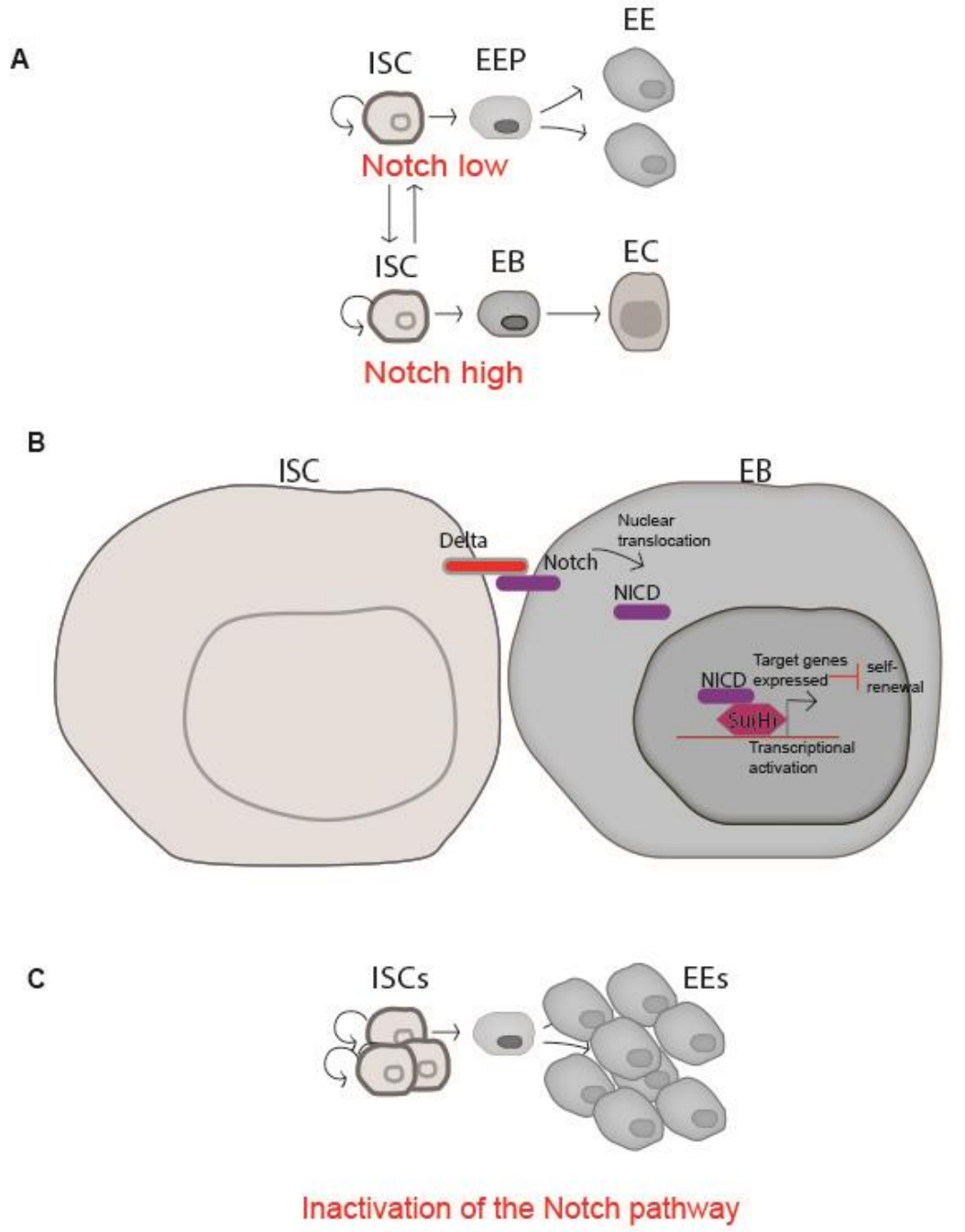


Figure 1.7: Role of the Notch pathway in specifying ISC lineage

(A) Low *Notch* signalling is required for EE cell fate acquisition whereas high *Notch* signalling specifies EC cell fate signalling.
(B) The Notch ligand Delta (RED) in ISC activates the Notch receptor (PURPLE) present in the EB (PINK) facilitating cleavage of the intracellular domain of Notch (N^{icd} - PURPLE) promoting transcription of downstream Notch target genes turning off stem cell self-renewal and activating an EC differentiation program.
(C) A loss of *Notch* signalling in stem cells will drive stem cell self-renewal and result in aberrant accumulation of EE cells and a loss of EC cells.

So far, I have highlighted how the *Drosophila* intestine functions to maintain a homeostatic state and how signaling pathways such as Dl/Notch signaling come to play. Dl/Notch signaling is important in orchestrating tissue homeostasis in order for the gut to carry out its functions. I will move on to describing what leads to changes in the homeostatic state of the gut.

The aging gut

In aging animals, cell signaling erodes, as repeated perturbations and repair lead to a decline in tissue function. The perturbations include changes in nutrient availability, temperature, oxygen levels and exposure to infectious or damaging agents, in addition to other internal changes including protein misfolding and DNA damage. This notion of “eroded harmony” in cell signaling with time, leads to aging phenotypes and is a common denominator of aging in different organisms. In recent years, model organisms have revealed that changes in signaling resulting in the disturbance of communication between cells, can augment these important stress response systems, altering lifespan and age-related changes. Examples of such stress response systems include highly conserved sirtuin, insulin/IGF-1, and TOR signalling pathways (Ayyaz and Jasper 2013). Although aging phenotypes differ from human to fly, the aging fly gut also undergoes age-related changes caused by alteration in stress response signaling pathways, which recapitulate the disturbance of communication between cells leading to aging. Here, I will discuss the observations made in the aging *Drosophila* midgut and how they contrast with that of a young midgut at different biological levels.

Age-related changes in the Drosophila midgut

A number of studies have examined how aging impacts the *Drosophila* midgut. As mentioned in 1.1, the depletion of stem cell reserves and/or diminished stem cell function seems to contribute to aging in some tissues. In addition, I described how DNA damage may be linked to the process of stem cell decline. Here I will primarily focus on alterations that impact stem cell properties which may contribute to their genomic damage.

Changes in ISC proliferation during aging

Impaired ISC activity can be observed at the tissue level (**Figure 1.8B**). It has been shown by Biteau and colleagues that old guts show an accumulation of progenitor cells and have an increase in ISC division with age. The authors show that this is caused by a disruption in JNK signaling with age that drives the proliferation of ISCs and leads to the alteration of tissue homeostasis in old guts through the accumulation of ISC daughter cells (Biteau et al. 2008). More recently, it was demonstrated that JNK interacts with a protein, Wdr62, at the spindle to promote planar spindle orientation by transcriptionally repressing Kif1, a kinesin. This in turn, was suggested to impact stem cell fate symmetry. In the aging fly, there was an overabundance of symmetric fates caused by the perturbations of the JNK/Wdr62/Kif1a axis (Hu and Jasper 2019). The observation of progenitor cells accumulating in old guts is consistent with a previous finding from Choi and colleagues where they also find an increase in proliferation in aged guts resulting in increased progenitor and stem cells with age. In addition to JNK signaling mentioned above, Choi et al show an increase in stress-responsive PVR signaling that is partially responsible for these aging phenotypes (Choi et al. 2008). Thus, aging is

associated with increased stress signaling that can promote enhanced ISC proliferation.

Evidence for increased DNA damage in ISCs during aging

In the aging *Drosophila* midgut, at the stem cell level, Park and colleagues have shown age-related accumulation of DNA damage marks with age (Park et al. 2012), a potential cause for perturbing important signaling. This was demonstrated with γH2Av, analogous to mammalian γH2Ax, which is a histone mark that forms when DNA double strand breaks (DSBs) appear (**Figure 1.8A**). DSBs instigate the DNA damage response (DDR) which recruits Ataxia Telangiectasia mutates (ATM) protein kinase leading to the phosphorylation of serine-139 of histone H2AX, turning it into γH2Ax, which can be visualized by immunofluorescence as foci adjacent to the DSB sites (Rogakou et al. 1998). In their study, Park and colleagues show that in young 10-day old flies, γH2Av is present in 10% of ISCs. With aging, this number increases to 41% of ISCs in 20-day old flies and 59% of ISCs in 45-day old flies (Park et al. 2012). Therefore another association with aging increased stem cell DNA damage.

Age-related increase in spontaneous mutation-induced neoplasia

Another important observation of age-related ISC functional decline at the tissue level, is the observation from our lab demonstrating spontaneously arising intestinal neoplasia in aged guts (mentioned in 1.1 above). These are caused by mutations in *Notch* pathway components blocking proper stem cell differentiation resulting in tumors characterised by the accumulation of diploid ISCs and EEs (Siudeja et al. 2015). While the aforementioned study by Parks et al showed an increase in DNA damage marks γH2Av with age, the question of whether the

increase in DNA damage translates into an increase in mutation was not known. Our lab found that old intestines had higher incidence of spontaneous arising mutant intestinal neoplasia, suggesting that either DNA damage is repaired less efficiently with age or the intestine is exposed to more DNA damage with age. The mutant neoplasia likely cause a decline in tissue function by promoting lineage skewing and accumulation of EE and ISC cells at the expense of ECs.

Changes in the gut epithelial integrity during aging

In addition to alteration of stem cell properties and tissues dynamics, various changes at the physiological level have been reported in the aging midgut too. Flies experience age-related increase in gut permeability due to gut barrier dysfunction. This has been shown using the “smurf” assay which puts the gut barrier integrity to test by feeding flies a non-absorbable blue dye. If the gut barrier is intact, the dye is retained within the gut. On the other hand, should there be a loss in barrier integrity, the blue dye is spread throughout the haemolymph, generating blue “smurf” flies (see **Figure 1.8D**) (Rera et al. 2011, 2012). This smurf phenotype is detected more frequently in aged flies and is an indication of impending death as loss of gut integrity is tightly coupled with age-related mortality (Bitner et al. 2020).

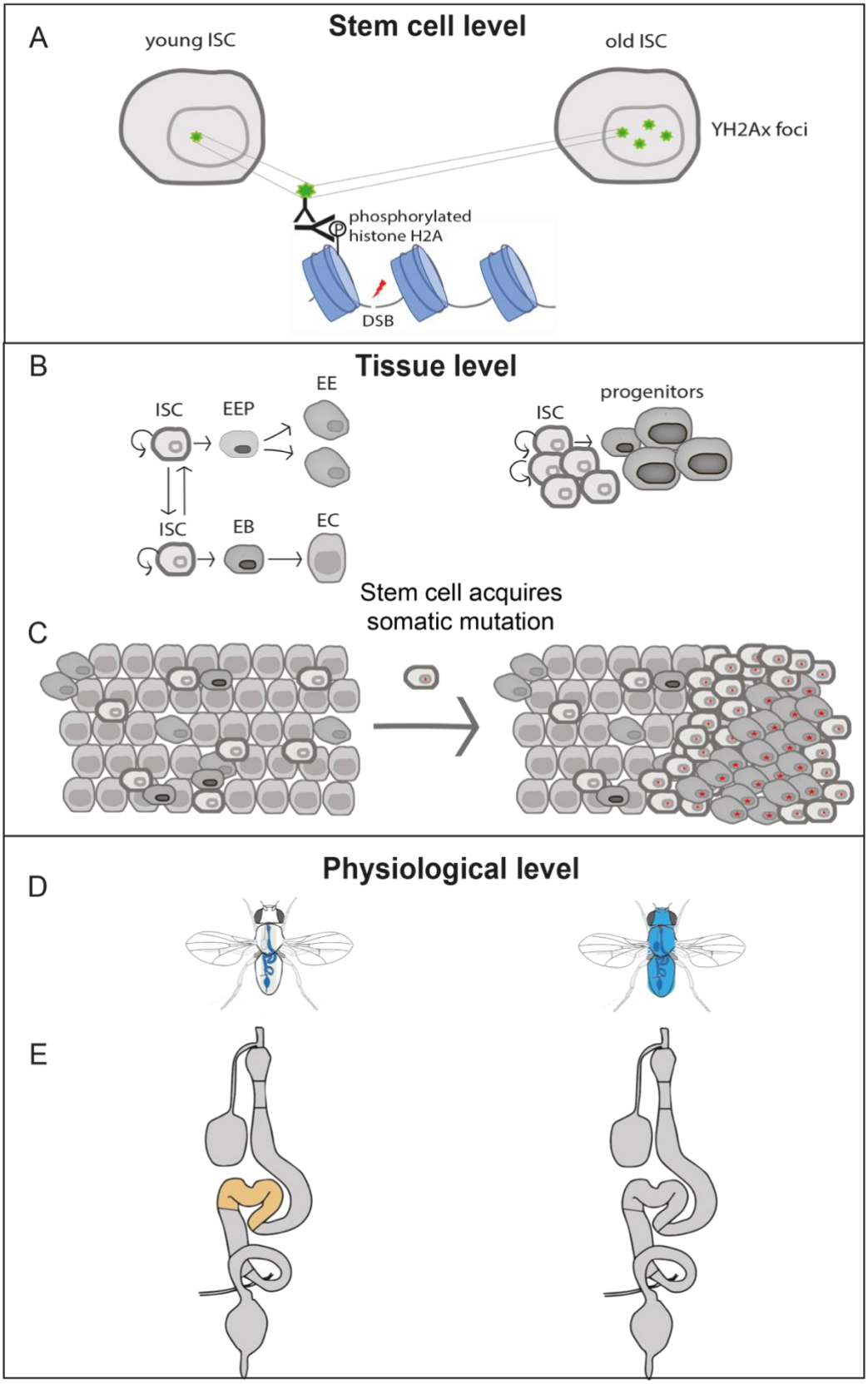


Figure 1.8: The aging *Drosophila* gut- changes at the stem cell level, tissue level and physiological level.

(A) Increase foci of marks of DNA damage, γH2Ax, are detected in ISCs upon aging.
(B) Increased ISC proliferation occurs with age, leading to changes in tissue structure with more newly formed progenitor cells in the tissue.
(C) Stem cell mutations occur during aging, with an inactivation of Notch causing a neoplastic phenotype.
(D) Physiological changes arise in adult life, with aged flies having permeable leaky guts and and changes in the CCR acidifying region.

One cause of loss of gut integrity is compromised tricelluar junction (TCJ) function in aged guts. In flies, TCJs seal the corners of three cells. It has been shown that the depletion of the TCJ protein, Gliotactin, from ECs induces age-related gut permeability and drives JNK signaling to drive ISC proliferation non-autonomously (Resnik-Docampo et al. 2017). Changes in gut regionalization have also been reported, with the middle CCR of the gut altering is cell composition and thereby modifying the gut pH leading to microbial (Li et al. 2016).

Impact of the environment on the *Drosophila* midgut

During aging, midguts undergo a functional decline and studies that have shown age-related changes in the microbiota in the form of dysbiosis of commensal bacteria (Clark et al. 2015; Rera et al. 2012; Guo et al. 2014). How an environmental factor such as the microbiota can contribute to an age-related functional decline is an interesting phenomenon, as it has the potential to be targeted and regulated, which in turn, has implications in extending lifespan for instance. Thus, it is paramount to gain a deeper insight into this tightly linked interaction between the intestinal epithelium and the microbiota to better understand stress and innate immune signaling in epithelial cells that can be used to develop therapies and preventative strategies for age-related diseases.

In humans, it has been documented that patients with inflammatory bowel disease (IBD), have an increased risk of developing colorectal cancer (Freeman 2008). This is likely due to the inflammation causing dysplasia. Indeed, there has been a rise in studies regarding microbiome-host interactions investigating the extent to which this can play a role in aggravating/relieving inflammation (Saus et al. 2019). Additionally, evidence that pathogenic bacteria have the ability to induce cancer initiation and progression was shown with an E.coli strain, NC101, which harbors a DNA-damaging toxin known as colibactin (Arthur et al. 2012). It was found that this strain was detected in 40% of IBD patients and in 70%of colorectal cancer patients. Thus, better knowledge of how bacteria can cause intestinal damage through inflammation, its genotoxicity, and other factors such as ROS production will be important for reducing CRC risk in patients with IBD for example. In my thesis, I aimed to address this point using the *Drosophila* model.

The *Drosophila* midgut is a particularly good model for addressing host-environment questions. Upon exposure to stress from the external environment the gut responds rapidly to accommodate the changes by inducing ISC proliferation to replenish the damaged gut cells and re-stablish tissue homeostasis (Amcheslavsky et al. 2009; Biteau et al. 2008; Buchon et al. 2009b, 2010; Jiang et al. 2009a; Cronin et al. 2014). Chemical damage for instance, can be caused by the anticancer drug DNA damaging agent, bleomycin. Bleomycin induces DNA damage specifically in the ECs but not in ISCs or EBs that are more basally located (Amcheslavsky et al. 2009). The bleomycin-induced damage in turn kills the ECs which consequently induces an increase in ISC proliferation through the insulin receptor (InR) signalling pathway and facilitates EBs to differentiate into new ECs to compensate for the loss (Amcheslavsky et al. 2009).

Bacterial damage on the other hand, can be caused by pathogenic infectious bacteria that are not indigenous to the normal gut microbiota (Lemaitre and Hoffmann 2007). Pathogenic bacteria of the fly intestine include the gram-negative *Erwina carotovara* subsp. *Carotovara 15* (Ecc15) and *Pseudomonas entomophila* (Buchon et al. 2009b; Jiang et al. 2009a; Vodovar et al. 2005; Zhu et al. 2019; Chakrabarti et al. 2012). Similar to bleomycin, these bacteria kill the ECs and, in turn, promote ISC proliferation (Jiang et al. 2009b; Buchon et al. 2009a, 2010). Infection-induced damage by bacteria causes the midgut to illicit an innate immune response via antimicrobial peptides (AMPs) as well as reactive oxygen species (ROS) production in order to attack the bacteria. ROS do not specifically target microbial structures, but rather damage proteins, lipids and nucleic acids through promoting oxidative degradation of the lipids in cell membranes. For this reason, anti-oxidant systems are mobilised in the gut in order to prevent damage on the host cell, namely the secretion of an extracellular immune related catalase (IRC), which neutralises ROS (Ha et al. 2005). With age however, it has been shown that the constant stimulation of immune resistant intestinal microbes results in excessive ROS accumulation, a likely contributing factor to the age-relates loss of tissue homeostasis (Buchon et al. 2009b). ROS production during aging could also conceivably impact DNA damage levels.

Drosophila midgut response to Ecc15 and Pe infection

Erwina caratovara (Ecc15) is a gram-negative bacteria, non-lethal pathogen that causes the host cell to produce ROS. Buchon and colleagues showed a ROS burst peaks 1 hour after infection with *Ecc15* (Buchon et al. 2009b). It has been suggested that this causes collateral effects causing stress in the host ECs, leading

them to delaminate from the epithelium. Despite the tight regulation to prevent damage to host gut cells (Ha et al. 2005), ECs are damaged and lost. The loss of cells in the epithelium instigates a proliferative response of the ISCs to compensate. A 10-fold increase of dividing cells has been shown in the guts of *Ecc15* challenged flies compared to unchallenged flies, after 16 hours of infection (Buchon et al. 2009b). This response is initiated by the release of the Upd3 cytokine from ECs which triggers JAK-STAT and the epidermal growth factor Keren. Induction of the JAK-STAT pathway in ISCs and EBs consequently promotes synthesis of Upd3 and another epidermal growth factor Spitz. Along with the epidermal growth factor ligand Vein, activated in the surrounding visceral muscle, these factors induce the EGFR pathway in ISCs to increase their rate of proliferation (Buchon et al. 2009b; Jiang et al. 2009a).

Pseudomonas entomophila (Pe), is another gram-negative bacterium that induces higher epithelial stress than *Ecc15*. Transcriptomic analysis found that more anti-microbial peptides are expressed after *Pe* infection compared to *Ecc15* infection (Chakrabarti et al. 2012). *Pe* infection also kills ECs (Vodovar et al. 2005) and induces a strong mitotic response in the midgut through JAK/STAT signalling that is required for ISC activation (Jiang et al. 2009a). Infection of a high dose of *Pe* induces an overwhelming stress response through high levels of ROS as well as the formation of a bacterial toxin called Monalysin (Chakrabarti et al. 2012). This, in turn, decreases global translation in ECs by inducing the kinase Gcn2 and decreasing Tor signalling. This consequently leads to the lack of tissue repair and death of the fly. Thus, the ingestion of diverse bacterial species can lead to dynamic stem cell responses, though differ in degrees and final outcomes of changes in tissue homeostasis.

Advantages of using the Drosophila midgut as a model system to study genome instability in adult stem cells

I will finally end chapter 1.3 highlighting the advantages of using the *Drosophila* gut as model system to study the impact of genome instability in adult stem cells discussed in chapter 1.1. Given the emerging evidence showing the increase in clonal expansions in mutant stem cell populations in aging tissues, highlighted in chapter 1.1, it has also been shown that the environment can affect the kinetics at which mutant lineages expand (Zhang *et al.*, 2001). As described in 1.1 above, clonal expansions can alter the homeostasis of the tissue and compromise its function, leading to aging phenotypes and cancer. A better understanding of the mechanisms underlying the acquisition of mutations, and factors that impact the initiation and expansion of mutant lineages is needed.

The *Drosophila* intestine is a well-suited model to better understand the molecular processes underlying somatic adult stem cell mutations and the factors that can impact their progression. Indeed, the *Drosophila* intestine is a regenerative tissue maintained by approximately 1000 multipotent intestinal stem cells that have been shown to undergo frequent spontaneous mutation. The readout of spontaneous mutations are the phenotypically visible neoplasias that are detected in aged flies. The lab showed that these are spontaneously arising mutant clones that arose from a mutant stem cell and that this system recapitulates clonal expansions of human tissues (Siudeja et al. 2015). The lab demonstrated through whole-genome next-generation sequencing that these neoplasias arise because of an inactivating deletion of the wild-type copy of the tumour suppressor gene *Notch*, which is on the X chromosome. Since males have a single copy of *Notch*,

inactivation of one copy is sufficient for the gut to acquire a Notch loss of function phenotype resulting in the neoplasia. In addition, this work demonstrated that heterozygous components of the Notch pathway undergo LOH and also lead to neoplasia (Siudeja et al. 2015). A mechanistic dissection of this process will be described below in my results 2.1. Overall, the Drosophila intestine can be used to determine how somatic mutations arise in adult stem cells.

Another advantage of this *Drosophila* model system is its small genome size. The *Drosophila* genome is ~175Mb compared with the 2.6Gb genome size of *Mus musculus*. This allows for more cost-effective sequencing. It also has a significantly shorter lifespan of 6 weeks, making aging experiments in flies faster than mouse aging experiments, which would take 3 years to acquire geriatric mice. I have highlighted that much is known about the intestinal stem cell lineage, how it is specified and signal pathways altering stem cell dynamics. Most importantly, what makes *Drosophila* a powerful model in general, is its genetic amenability with tools available for manipulation in vivo and the ability to alter environmental conditions. Despite the physiological divergence between *Drosophila* and vertebrates, the modelling of human intestinal diseases is possible because of the high degree of conservation in signaling pathways.

Chapter 2
Results

Chapter 2 : Results

Results Overview

In this chapter I will present the results in two sections. The first section (chapter 2.1) is the paper focusing on mitotic recombination as a mechanism driving spontaneous LOH in *Drosophila* intestinal stem cells.

The second section (chapter 2.2) presents another mechanism, *aneuploidy*, by which LOH can occur.

2.1 Mitotic Recombination as a Mechanism Driving Spontaneous Loss of Heterozygosity in *Drosophila* Intestinal Stem Cells **(article in preparation).**

Lara Al zouabi, Nick Riddiford, Marine Stefanutti, Mirka Uhlirova and Allison Bardin

In preparation

Mitotic Recombination as a Mechanism Driving Spontaneous Loss of Heterozygosity in *Drosophila* Intestinal Stem Cells

Lara Al zouabi[1,2], Nick Riddiford[1,2], Marine Stefanutti[1,2], Mirka Uhlirova[3] and Allison Bardin[1,2]

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

[2]Sorbonne Universités, UPMC University, Paris 6, France

[3]Institute for Genetics, University of Cologne, Cologne, Germany

Abstract

Somatic cells can undergo a genome alteration leading to loss of heterozygosity (LOH). This phenomenon occurs in normal human tissues, pathological disorders, and cancers. Although previous studies in yeast have provided substantial insight into different mechanisms of LOH, mechanistic details are lacking in multicellular organisms with complex tissues. Here we investigate the mechanisms giving rise to LOH, bridging the gap between unicellular yeast and higher eukaryotes using an *in vivo* stem cell model system in *Drosophila*. Through whole-genome sequencing of somatic LOH events, profiling copy number changes and changes in heterozygosity of single-nucleotide polymorphisms, we demonstrated that LOH arises via mitotic recombination. Consistent with this, we found involvement of the DNA repair enzyme Rad51 in LOH. Fine mapping of recombination sites did not reveal mutational pile-ups that commonly arise with a break-induced replication mechanism and instead showed clear examples of chromosomes arising from cross-over events generated by double-Holliday junction-based repair. The mapped recombination regions also provided insight into potential genomic sequence features that may promote mitotic recombination, including an association with the repeated region of the *Histone Locus Cluster* and regions previously mapped to form R loops. We further explored how environmental factors can influence this process and demonstrate that infection with the enteric pathogenic bacteria, *Ecc15*, increased LOH frequency. This study provides a better mechanistic understanding of how mitotic recombination arises in stem cells *in vivo,* and identifies intrinsic and extrinsic factors that can drive LOH, thus providing

important insight into cancer initiation and potential preventative and therapeutic strategies.

Introduction

Diploid organisms, with two copies of each gene, have an evolutionary advantage providing protection if a mutation arises in one copy, as the second wild-type allele maintains function. Thus diploidy provides redundancy, masking deleterious recessive mutations (Crow and Kimura 1965; Perrot et al. 1991; Otto and Goldstein 1992; Mable and Otto 2001). Despite this protective system, the genomes of somatic cells are under constant threat of DNA damage from both endogenous and external insults, thus the wild-type "backup" copy can be somatically inactivated through a process called loss of heterozygosity (LOH). LOH is one of the most common means of inactivation of tumour suppressor genes (Crabtree et al. 2003; Saeki et al. 2011; Knudson 1971; Xu et al. 1992; Shetzer et al. 2014) and reviewed in (Couto 2011; Wang 2018; Ryland et al. 2015). LOH can arise through different genetic mechanisms including point mutations, deletions, chromosome loss, and mitotic homologous recombination (MR)-based mechanisms resulting in co-segregation of two mutant alleles into a daughter cell. It has been shown that MR plays a substantial role in both sporadic cancers and familial cancers and is particularly problematic for individuals who have germline mutations in *adenomatous polyposis coli* (*APC*) (Cottrell et al. 1992; Haigis et al. 2002; Thiagalingam et al. 2001), *retinoblastoma* (*Rb*) (Cavenee et al. 1983) and *neurofibromatosis 1* (*NF1*) (Serra et al. 2001) as the subsequent somatic inactivation of the wild-type allele happens via MR, also reviewed in (Tuna et al. 2009; Lapunzina and Monk 2011; Siudeja and Bardin 2017).

MR, is a DNA double-strand break repair (DSBR) mechanism employed in mitotically dividing diploid cells to correct a DNA double strand break (DSB) using the intact homologue as a template and was first described by Stern in classic *Drosophila* experiments (Stern 1936a). MR is a double-edged sword however: in cells heterozygous for a tumour suppressor gene (such *APC*, *Rb*, *NF1* mentioned above), MR corrects a DNA lesion, but at the cost of recombining away the wild-type backup copy, thereby leading to cancer initiation. Thus, a thorough understanding of the molecular mechanisms giving rise to MR as well as cell-intrinsic and -extrinsic factors that can promote MR, will provide insight into cancer initiation and potential prevention strategies. Of particular importance is understanding these processes in adult stem cells, that maintain cell division capacity and plasticity in adult tissues. Indeed, adult stem cells are capable of producing large numbers of differentiated cells and therefore the transmission of mutant genomes can initiate tumour formation or profoundly affect tissue function.

While there have been numerous efforts to understand how MR arises in somatic tissues, MR for the most part has been observed using low-resolution techniques such as PCR-based methods, targeted sequencing and following markers with a significant distance between them using SNP and microsatellite arrays. In a study mapping sites of MR in human tumours caused from the LOH of *APC*, sites of MR were resolved to around 4.5 Mb. The authors claim that the sites were non-random and possibly associated with low copy repeats (LCRs), however the poor coverage of polymorphic markers did not allow finer-scale mapping (Howarth et al. 2009). Precise mapping provides an understanding of mechanistic signatures as well as features that could drive MR and has been achieved in yeast (Yin et al. 2017; St. Charles and Petes 2013). In addition to the polymorphic markers present in yeast,

it is possible to recover all MR products through analysis of clone sectors (Barbera and Petes 2006). For this reason, yeast has provided substantial insights into mechanisms of MR, which can arise by break-induced replication (BIR) and DNA DSBR leading to cross-over (CO) via double-Holliday junction resolution. Additionally, it has been shown that spontaneous MR sites are enriched near sites of inverted repeats of Ty transposable elements (St. Charles and Petes 2013). Whether these sequence features driving MR are conserved and operate in adult stem cells as well as which mode of MR, (BIR or CO) takes place, remain elusive. Thus, there is a need to bridge the gap between unicellular yeast and higher eukaryotes to understand pathways that promote and prevent MR in a complex tissue *in vivo*.

Drosophila is a well-established model that has made important contributions to our understanding of the dynamics of somatic tissues. The use of precise genetic tools has facilitated cancer modelling in *Drosophila* (Villegas 2019). In particular, studies using the adult *Drosophila* intestine (midgut) have defined cell signalling within tumour niches, and provided insight into cancer cachexia (Saavedra and Perrimon 2019, Gerlach and Herranz 2020, Patel et al. 2015). Additional studies have revealed environmental factors including a role for the microbiome in the progression of induced tumours in the midgut (Zhou and Boutros 2020; Ferguson et al. 2020). Whether the microbiome or other environmental factors play a role in initiating these tumours, is not clear. The *Drosophila* adult midgut is a regenerative tissue composed of approximately 10,000 cells renewed weekly by around 1000 multipotent intestinal stem cells (ISCs). ISCs are the primary dividing cell type in the tissue, and undergo asymmetric cell divisions to self-renew and give rise to an enteroblast (EB) and less frequently, an enteroendocrine precursor (EEP). These

progenitor cells give rise to the differentiated cell types the enterocytes (ECs) and enteroendocrine cells (EEs).

Previously, we had observed that tumour-like clusters of cells composed of ISCs and EEs appear spontaneously and frequently in the adult fly intestine (Siudeja et al. 2015). We demonstrated that the neoplastic clusters of cells are clonal and derive from spontaneous mutation of the *Notch* gene occurring in ISCs. The cell signalling receptor and transcription factor, *Notch*, acts as a tumour-suppressor gene in this system, crucial for limiting ISC and EE cell proliferation through controlling daughter cell fate decisions (Micchelli and Perrimon 2006a; Ohlstein and Spradling 2006b). *Notch*, is X-linked and therefore hemizygous in male flies, where it becomes spontaneously inactivated in some ISCs in ~10% of 6-week-old aged wild-type flies through deletion, structural rearrangement, and transposon insertion (Siudeja, 2015; Riddiford, 2020; Siudeja, 2020). Aged female flies, heterozygous for a mutant allele of *Notch*, were shown to develop spontaneous neoplasia in ~80% of animals. Collectively, these studies established an *in vivo* model system to investigate spontaneously arising gene inactivation events in adult stem cells with a simple phenotypic readout of neoplastic clones of ISC and EE cells.

Here we use the fly intestine as an *in vivo* model system to study spontaneously arising LOH in adult stem cells in a complex eukaryotic tissue. By exploiting the higher density of polymorphisms in *Drosophila,* twice as polymorphic than humans (Wang et al. 2015; Langley et al. 2012), combined with high coverage whole-genome sequencing (~50X), we could map MR events arising spontaneously in adult ISCs. Our data suggest that a majority of LOH events are driven by MR and not inactivation through point mutation, deletion or chromosome loss. Our findings further support CO as a primary mechanism model of MR, and argue against BIR

as a prevalent mode of MR. Cell-intrinsic DNA sequence features such as the repeated locus of the *Histone Cluster* and regions shown to map to R loops are associated with sites of recombination. Furthermore, cell-extrinsic interactions with the pathogenic bacteria *Ecc15*, known to promote rapid proliferation of ISCs, was found to increase DNA damage of ISCs and increase LOH frequency. We thus present a better mechanistic understanding of MR arising in stem cells *in vivo* and delineate underlying cell-intrinsic and extrinsic environmental features promoting LOH, providing important insight into cancer initiation, potential preventative, and therapeutic measures.

Results

Spontaneous loss of heterozygosity increases with age

In order to systematically study mechanisms and frequencies of the loss of heterozygosity (LOH), we wanted to use a genetic background in which an intermediate frequency of spontaneously arising LOH events occurred. As inactivation of the Notch pathway activity provides a robust readout, we assessed genes encoding Notch pathway components as potential tools to study LOH. We had previously shown a relationship between distance to the centromere and LOH frequency suggestive of a recombination-based mechanism of LOH, with *neuralized* heterozygous mutant flies having very few events (2.9% at 5 weeks of age) - and *Notch* heterozygous mutant flies - having many events - (81% at 5 weeks of age, most of which were multiple events) (Siudeja et al. 2015). A good candidate genetic background for an intermediate frequency of LOH, was the heterozygous flies for a null allele of *suppressor of hairless (Su(H)$^{\Delta 47}$)*, a encoding a null allele of a *Notch* pathway component, whose loss-of-function was previously shown to produce large

mutant clones when induced genetically with the FLP/FRT system (Bardin et al. 2010). *Su(H)* is 7.5Mb away from the centromere on chromosome 2L, an intermediate distance to that of *Notch* (21.8Mb) or *neur (*4.8Mb). We therefore predicted that *Su(H)*-/+ flies would have moderate frequencies of LOH suitable for detecting changes in frequencies in response to different conditions, and for avoiding multiple mutant spontaneous LOH clones per gut that may result in clone fusion, complicating downstream analysis.

We therefore assessed whether the $Su(H)^{\Delta 47}$/+ genetic background gave rise to spontaneously arising mutant clones in aged flies. Similar to wild-type and N^{55E11}/+ flies, the $Su(H)^{\Delta 47}$/+ flies presented an overall wild-type midgut appearance, composed of large polyploid enterocytes (ECs) with interspersed enteroendocrine (EE cells) (**Figure 2.1A- B'**). Upon aging, patches of tissue with a *Su(H)* loss-of-function phenotype arose, comprised of an accumulation of Delta (Dl) positive ISCs and Prospero+ (Pros+) EEs (**Figure 2.1C, C'**). These data strongly suggest that, as we previously showed for other Notch pathway components, the spontaneous inactivation of the wild-type allele of *Su(H)* occurs during aging **(Figure 2.1C')**. Since the primary dividing cell-type in the adult midgut is the ISC and inactivation of Notch pathway in ISCs is sufficient to generate neoplastic clones (Siudeja et al. 2015), the inactivation event (**Figure 2.1C')** likely occurred in ISCs.

Interestingly, we found that the frequency of detected spontaneously arising mutant LOH clones increased with age. In young 1-week old $Su(H)^{\Delta 47}$/+ flies, only 0.9% guts had mutant clones (n=108)

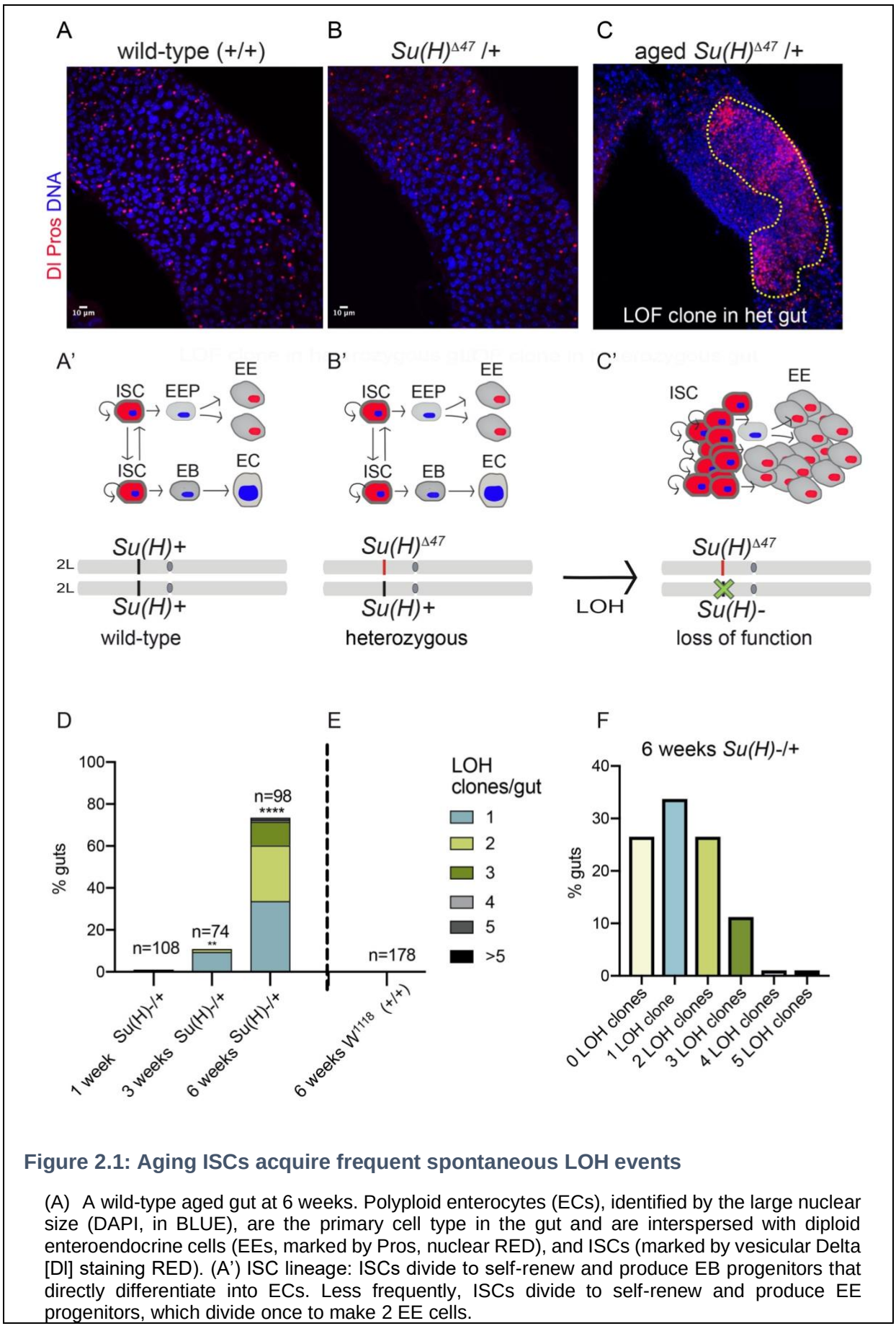


Figure 2.1: Aging ISCs acquire frequent spontaneous LOH events

(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, marked by Pros, nuclear RED), and ISCs (marked by vesicular Delta [Dl] staining RED). (A') ISC lineage: ISCs divide to self-renew and produce EB progenitors that directly differentiate into ECs. Less frequently, ISCs divide to self-renew and produce EE progenitors, which divide once to make 2 EE cells.

(B) A majority of tissue in the intestines of *Su(H)*$^{\Delta 47}$/+ flies was like that of wild-type flies (A' and (B') giving rise to large polyploid enterocytes (ECs) with interspersed Pros+ EEs at homeostatic proportions.
(C) Example of neoplastic *Su(H)* loss of function (LOF) clone (outlined in yellow), composed of an excess of Dl+ ISCs and EE cells (Pros+, nuclear RED).
(D) Frequency of LOH clones in *Su(H)*$^{\Delta 47}$/+ midguts at 1, 3 and 6 weeks of ages. The 1 week time point was used to calculate statistical significance.
(E) No midgut showed an LOH clone in wild-type intestines.
(F) Distribution of number of LOH clones per gut from *Su(H)*$^{\Delta 47}$/+ flies at 6 weeks.
***p < 0.001; ****p < 0.0001; n.s., not significant (Fisher's exact test, two-tailed).

(Figure 2.1D). At 3 weeks of age, the percentage of midguts with detected spontaneous clones increased to 10.8% (n=74), and further increased to 73% at 6 weeks of age (n=98). No mutant clones were observed in 6-week wild-type *w*^{1118} females (+/+) (n=178; **Figure 2.1E)**. Importantly, a majority of 6-week-old guts had 1 neoplastic growth (33.7%), whereas 26.5% had 2 neoplasia **(Figure 2.1F)**. We previously demonstrated that the Notch-/+ background had a majority of guts which contained more than 1 event, with clone fusion events therefore being highly likely. We conclude that the *Su(H)*$^{\Delta 47}$/+ background is a useful genetic background in which to further elucidate mechanisms of spontaneous LOH as it acquires mutant clones at intermediate frequencies, allowing modifications in frequency to be easily detected, and in which clone fusion is unlikely.

Whole genome sequencing to determine the mechanism of LOH

While our previous study hinted towards mitotic homologous recombination (MR) as a mechanism of LOH (Siudeja et al. 2015), as the frequency of LOH was diminished by balancer chromosomes known to suppress recombination, however, we could not rule out other mechanisms. In particular, a recent study suggested that an unusual chromosome segregation mechanism, amitosis, can lead to LOH (Lucchetta and Ohlstein 2018). In this mechanism, enteroblast progenitor cells of

the gut, which have undergone S phase and have a 4n chromosome ploidy, can undergo a reductive cell division under extreme stress conditions, leading to segregation of 2 chromosomes originating from the same parent, into a daughter cell (Lucchetta and Ohlstein 2018). Similarly, our previous study could not distinguish MR from chromosome loss (aneuploidy). Finally, whether MR might occur via cross-over - the exchange of material from homologous chromosomes - could not be distinguished from break-induced replication (BIR), where error-prone polymerases copy genetic material from the homologous chromosomes. Therefore, in order to differentiate between these mechanisms, we sought to carry out whole genome sequencing (WGS) of spontaneous *Su(H)* mutant clones in $Su(H)^{\Delta 47}$/+ flies to determine the molecular nature of inactivation events. In order to determine somatic mutation events, we adapted methodology previously developed in the lab (Siudeja et al. 2015) to compare genomic DNA from neoplastic gut tissue to genomic DNA from the head of the same individual using whole genome sequencing (**Figure 2.2A-D**).

To facilitate neoplasia detection and to obtain large enough quantities of genomic DNA from somatic mutation events, we aged flies and dissected at 5-6 weeks since by this time, neoplasia grew large enough to be isolated for sequencing. In order to identify the neoplastic LOH clones, we used a GFP reporter of EE cells, which aberrantly accumulate in the neoplasia ($Su(H)^{\Delta 47}$/+; $Pros^{V1Gal4}$/ *UAS-nlsGFP* genetic background; **Figure 2.2B, C**). In this context, *Gal4* expression is driven by the *pros* promoter leading to induction of *UAS-GFP*, allowing the characteristic accumulation of Pros+ EEs of the mutant clones to be identified as GFP+ clusters (**Figure 2.2A, B**). Midguts containing a large enough GFP+ cluster were identified and the region containing the mutant cells was manually micro-

dissected together with the head of the same fly that serves as the normal tissue control allowing for discerning somatic from germline events (**Figure 2.2C**). We refer to the mutant clone and its corresponding head as “tumour” and “normal” samples. Genomic DNA was then isolated from the “tumour” mutant clone and “normal” head and Illumina paired-end (125nt) sequencing was performed at a depth of ~50X on 16 female clones and 6 male clones with the *Su(H)*$^{\Delta 47}$/+; *ProsV1Gal4/ UAS-nlsGFP* genetic background. 2 of these female samples were excluded from further analysis due to low sequencing coverage (shown in **Table S1**). Importantly, the parental genotypes carried a large number of single nucleotide polymorphisms (SNPs), that we took advantage of in our bioinformatic analysis detailed below.

In order to distinguish between different mechanisms of LOH, we analysed the whole-genome sequencing (WGS) data of tumour/normal pairs for distinct features that could lead to inactivation of the wild-type copy of *Su(H)*, including copy number changes, structural variants, point mutations and changes to the zygosity of parental SNPs, i.e. those normally heterozygous could become homozygous resulting in a shift in zygosity indicative of a recombination process. We explored the following 5 possibilities (**Figure 2.2D**): (1) If LOH was due to a point mutation, there would be no change in zygosity or copy number between head and tumour. (2) If the *Su(H)* inactivation event was due to a deletion, there will be a copy number variant (CNV) removing *Su(H)* genomic sequence as well as a loss of the parental SNPs on that chromosome region.

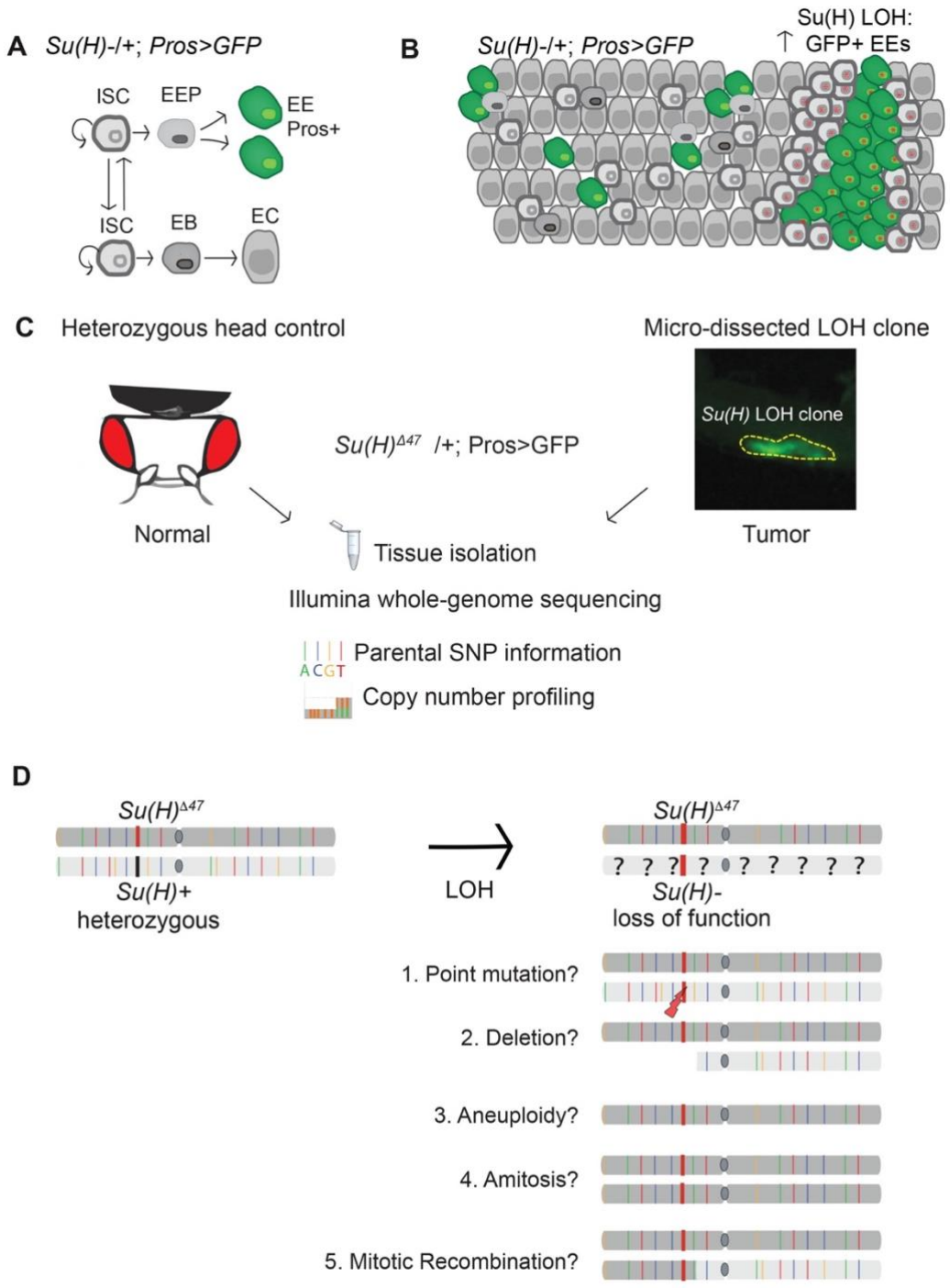


Figure 2.2: Sequencing setup and analysis predictions

(A) ISC lineage: ISCs divide to self-renew and produce EB progenitors that directly differentiate into ECs. Less frequently, ISCs divide to self-renew and produce EE progenitors, which divide once to make 2 EE cells. *ProsGal4 UAS-GFP* transgenes in the genetic background 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 recognised by the accumulation (marked by an arrow) of GFP+ EEs.
(C) The GFP+ LOH clones were microdissected with the head of the same fly for Illumina paired-end sequencing allowed for SNP (germline parental differences), SNV (somatic point mutations), and copy number profiling.

(D) SNP, SNV, and copy number profiling give allow to distinguish between the 5 potential mechanisms by which LOH can arise

(3) Aneuploidy would be revealed by a change in zygosity (heterozygous to homozygous [het->homo] SNPs of all of chromosome 2) and a copy number loss affecting the entire 2nd chromosome. (4) Amitosis would similarly result in a change in zygosity (het->homo SNPs of chromosome 2), however, there would be no copy number change as chromosome 2 would remain diploid. (5) Finally, mitotic recombination would reveal a shift in zygosity (het->homo SNPs of part of chromosome 2) with no copy number change. Therefore, our model system combined with WGS and analyses pipelines allow us to discern between the different LOH mechanisms.

LOH arises through mitotic recombination in both males and females.

With these 5 possible mechanisms in mind, we then analysed our sequencing datasets for the 13 female and 6 male neoplasia samples ("tumours"), compared to their respective head ("normal") tissue. A combination of in-house bioinformatic pipelines [(Riddiford 2020) and https://github.com/bardin-lab)] and available published pipelines that utilise split-read and mate pair-based evidence along with changes in genomic coverage, were used to assess altered copy number and structural variants along with point mutations and changes in SNP zygosity **(see Methods).**

In the 13 female samples, no structural variants were detected on chromosome 2 or point mutations in *Su(H),* ruling out deletion, point mutation, or loss of the chromosome as mechanisms of inactivation of *Su(H)*. We instead found changes in SNP heterozygosity, indicating that in 12 out of the 13 female tumours, LOH occurred by a mechanism consistent with MR (mitotic recombination). Thus, in

these samples the mutant *Su(H)* allele becomes homozygous, resulting in the neoplastic phenotype. A shift in zygosity going from a variant allele frequency (VAF) of ~0.5 to a VAF of 0.75 or above or 0.25 and below on chromosome 2L, confirmed that loss of heterozygosity (LOH) occurs between the *Su(H)* gene and the centromere and extends throughout the chromosome to the telomere **(Figure 2.3A;** for all 13 samples see **Figure 2.S1**). In a scenario where the tumour purity is 100%, the VAFs in LOH regions should be 1or 0, however given the manual nature of the tumour micro-dissections, it is likely that contaminating non-tumour cells make up part of the sequenced tumour resulting in deviations from this (for example see **Figure S1**, sample F10 is at 0.75 and 0.25).

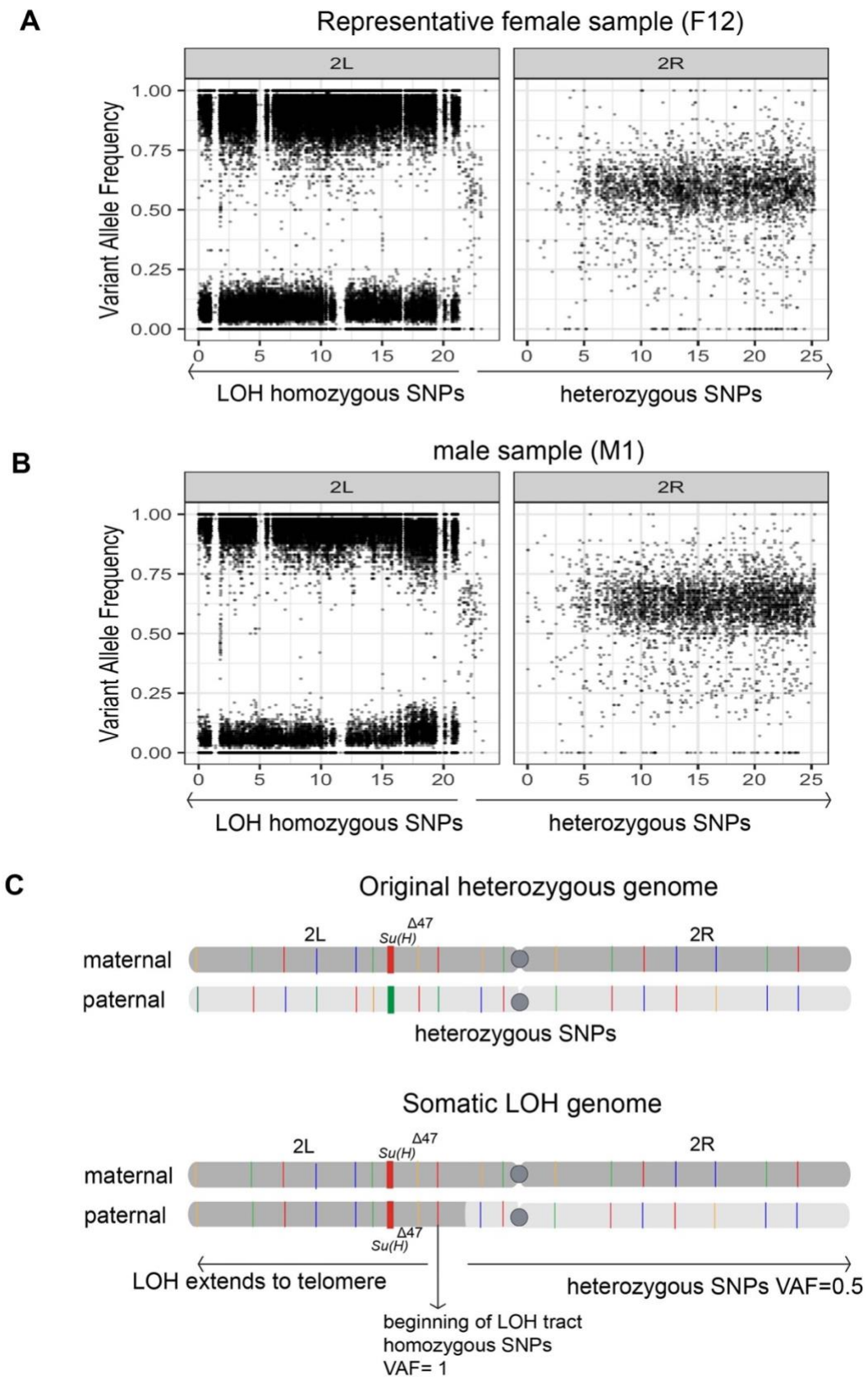


Figure 2.3: Figure 2.1 Sequencing *Su(H)* LOH clones reveals loss of heterozygosity through mitotic recombination in both males and females

(A) A representative Variant Allele Frequency (VAF) plot of a female sample (F11). In this LOH sample, chromosome 2R on the right panel had heterozygous SNPs, represented at ~0.5 VAF. In contrast, chromosome 2L has undergone an LOH event, with SNPs becoming homozygous (~1%) shown in the left panel.
(B) A representative Variant Allele Frequency (VAF) plot of a male sample (M1), in which LOH arose on chromosome 2L.
(C) Schematic representation of homologous chromosomes in which the bottom chromosome has undergone a mitotic recombination event, therefore resulting in LOH of *Su(H)*$^{\Delta 47}$ as well as the SNPs along the chromosome arm (shown as coloured bars)

Therefore, the mechanism of LOH in the female samples that we analysed can be attributed to MR, a DNA repair mechanism that comes into play to correct a double-strand break (DSB) likely arising on the chromosome arm. The cell uses the chromosome as a template to repair. In this particular case, the homologous chromosome arm harbouring the mutant *Su(H)* allele was used to repair the other homologous chromosome. Upon chromosome segregation, 2 copies of the mutant *Su(H)*$^{\Delta 47}$ are inherited in the same cell. Consequently, the DSB gets repaired but at the cost of losing the wild-type backup copy of the *Su(H)* gene.

Similarly, in 1 of the male samples, we found evidence for LOH via MR, with a shift in allele frequencies of SNPs on chromosome 2, resulting in LOH of a large portion of the chromosome arm, including the *Su(H*) locus **(Figure 2.3B)**. The remaining 5 out of the 6 males neoplasia samples showed structural variants of different lengths that inactivated the *Notch* locus located on the X chromosome and therefore hemizygous in males **(Figure 2.S2).** These events are consistent with the inactivation events in *Notch* that we previously described in wild-type male flies (Siudeja et al. 2015; Riddiford 2020).

The female sample (F6) for which we could not find evidence supporting LOH, showed significant non-tumour contamination thus impacting bioinformatic analysis **(Figure 2.S3A)**. We therefore used this sample to benchmark what is

deemed “too low” a tumour purity for the analysis. Since most of the adult intestine is composed of polyploid ECs (80%), it is likely that the detected contaminating reads in the tumour will be coming from ECs (non-tumour cell) **(Figure 2.S3B, B’)**. Therefore, we examined WGS generated from whole gut samples, comprised of a majority of EC cells. This sequencing, therefore, mostly reflects the polyploid genomes of ECs as they constitute most of the sequencing reads. Comparing whole guts with their respective heads revealed distinct genomic regions in the gut that are under-replicated (**Figure 2.S3C, C’**). These under-replicated genomic regions have a reduced copy number relative to the overall ploidy of the cell and are reflected as a loss of coverage in those regions and a drop in copy number (**Figures 2.S3B-C’**). This signature is characteristic of polyploid cells undergoing endoreplication as has been previously described (Yarosh and Spradling 2014; Spradling 2017), which we consequently named “EC signature”. One particularly obvious under-replicated region is the first 3Mb on chromosome 3R. We thus used this to assign a value for EC contamination in all the samples, (**Figures 2.S3B-D’** and **Table 2.S1)**.

These data provide evidence, for the first time, that spontaneous mitotic recombination occurs in ISCs. Importantly, our data exclude other mechanisms of LOH including (1) point mutations, (2) deletions, (3) aneuploidy, and (4) amitosis, at least for the 12 female and 1 male samples that we analysed. Furthermore, the long track LOH extending to the telomere rules out repair mechanisms that result short track LOH that do not lead to crossover, including gap repair and single-strand annealing.

LOH through mitotic recombination also happens on other chromosome arms

We also carried out similar analysis using two other Notch pathway components to determine whether MR also affected other chromosome arms. We tested chromosome 2R using heterozygous flies for a null allele of the Notch pathway component *O-fut1*. Chromosome 3R was assessed using an inactivating allele of Notch ligand, Dl – Dl^{Gal4}. We sequenced 3 neoplasia from $O\text{-}fut1^{4R6}$/+, and 4 neoplasia from Dl^{GAL4}/+, from which 2 *O-fut1* and 3 *Dl* passed our EC signature quality control. We detected evidence of LOH due to mitotic recombination based on the VAF frequency (**Figure 2.4A, B**) in 1 $O\text{-}fut1^{4R6}$/+ sample and 1 Dl^{GAL4}/+ sample, consistent with our above analysis of the $Su(H)^{\Delta 47}$/+ on the 2nd chromosome. Therefore, we conclude that spontaneous MR leads to LOH that can inactivate heterozygous genes and affect large chromosomal regions.

Interestingly, 3 additional samples pointed to novel mechanisms of LOH. One *Dl* sample, showed a structural variant at the *Dl* locus (data not shown), which likely led to inactivation of the wild-type allele. Two additional samples showed a loss of the entire second X chromosome (aneuploidy). One of these samples also had clear evidence of a somatic structural variant in *Notch*, whereas the other was too low of coverage to assess this (data not shown- see thesis section 2.2). We hypothesize that somatic loss of X co-occurred with somatic inactivation of *Notch*, in these samples, therefore leading to biallelic inactivation of *Notch*. Although this is not LOH of the wild-type alleles we were assaying, it nevertheless provides sequencing-based evidence for aneuploidy driven LOH which is another very important cancer initiating process. I will further discuss aneuploidy in section 2.2 below.

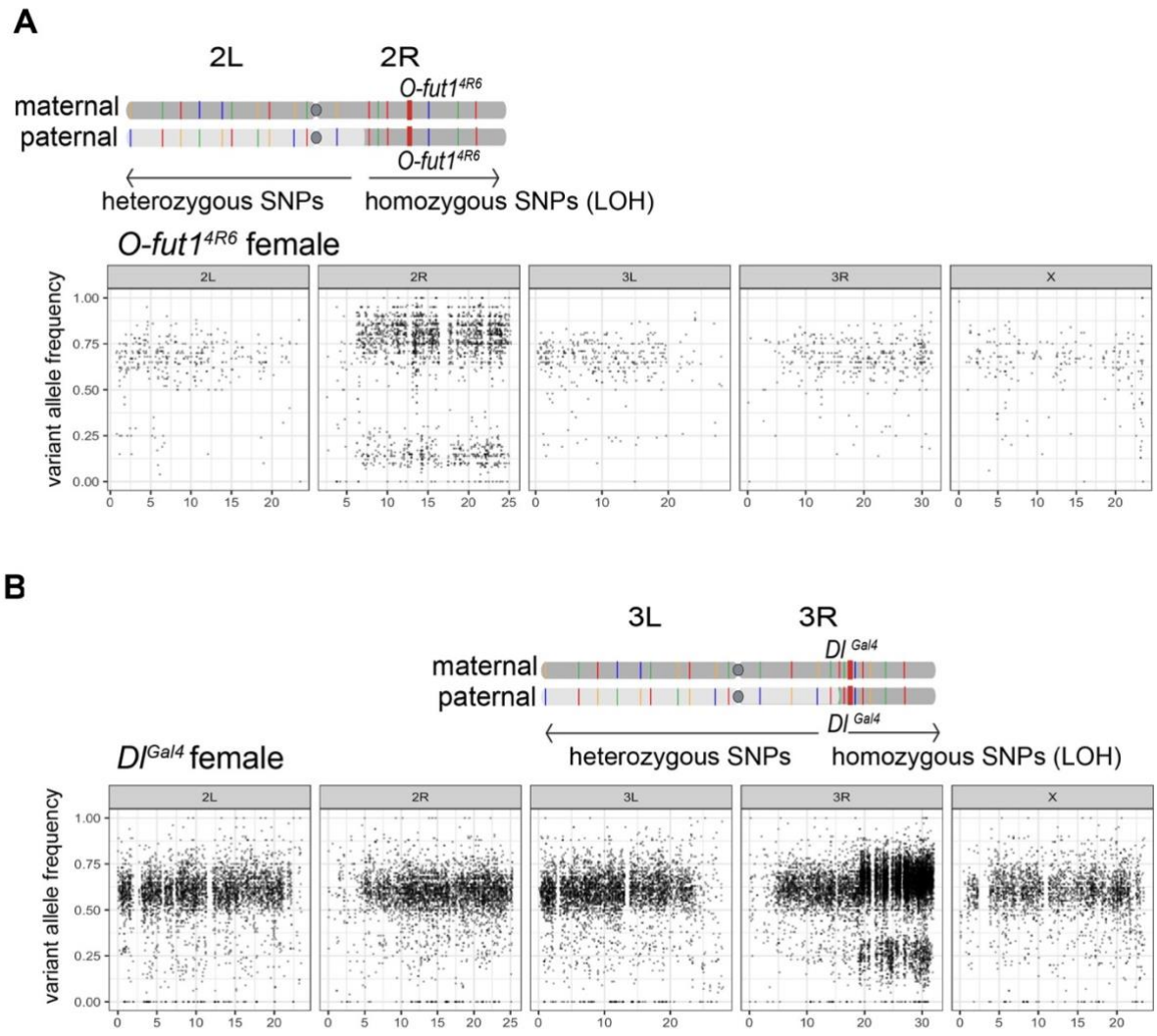


Figure 2.4: Loss of heterozygosity via mitotic recombination is also detected on other chromosome arms/ chromosomes

(A) Schematic of chromosome 2, showing LOH of chromosome 2R. VAF plots of an LOH sample from the *O-fut1*^{4R6}/+ genotype.
(B) Schematic of chromosome 3, showing LOH of chromosome 3R. VAF plots of an LOH sample from the *Dl*^{GAL4}/+ genotype.

Rad51 promotes loss of heterozygosity

We next wanted to investigate further the dependency on distinct sub-types of MR and to delineate genetic requirements for LOH leading to neoplasia formation. For MR to occur in a cell, there needs to be an interaction between the broken DNA and the homologous donor DNA, which typically relies on the Rad51 protein **(Figure**

2.5A). We therefore wanted to know whether knocking down *Rad51* (*spnA*) activity would decrease LOH frequency and neoplasia formation, suggesting that Rad51 is required for LOH formation, likely because damaged stem cells would be unable to repair and consequently die. To address this question, we expressed RNAi targeting *Rad51* in a stem cell-specific manner in $Su(H)^{\Delta 47}$ /+ heterozygous mutants and compared this with $Su(H)^{\Delta 47}$ /+ flies lacking RNAi expression. *Rad51* RNAi was expressed using Dl^{GAL4} combined with $tub\text{-}GAL80^{ts}$ in which ISCs express *Gal4* whose activity is controlled by the temperature sensitive, ubiquitously expressed *GAL80* (see **Figure S2.8** and **methods**). We found that LOH frequency decreased from 23% (n=113) in controls to 11.5% (n=312) upon expression of an RNAi targeting *Rad51* **(Figure 2.5D)**, suggesting that the MR in our system is, at least in part, dependent on *Rad51*. The remaining LOH detected upon *Rad51* knockdown may be explained by residual Rad51 activity. In addition, we cannot exclude an alternative, yet interesting, hypothesis that upon *Rad51* knockdown, other repair mechanisms substitute that would not generate LOH. Further studies will determine whether stem cells die or undergo different repair in absence of *Rad51* activity.

Two distinct mechanisms of MR could explain LOH in intestinal stem cells, resulting in the very long regions of chromosomes that were detected (Haber 2018): (1) Break-induced replication (BIR), in which an error-prone polymerase would copy material directly from the homologous chromosome **(Figure 2.5B)**. (2) Crossover, resulting from a classic double-strand break repair (DSBR) involving a double-Holliday junction structure (4 stranded DNA molecules between the homologous chromosome), whose resolution would lead to reciprocal exchange of segments of the homologous chromosomes **(Figure 2.5C)**. While both mechanisms would lead to long stretches of LOH spanning the chromosome arm and depend on *Rad51*,

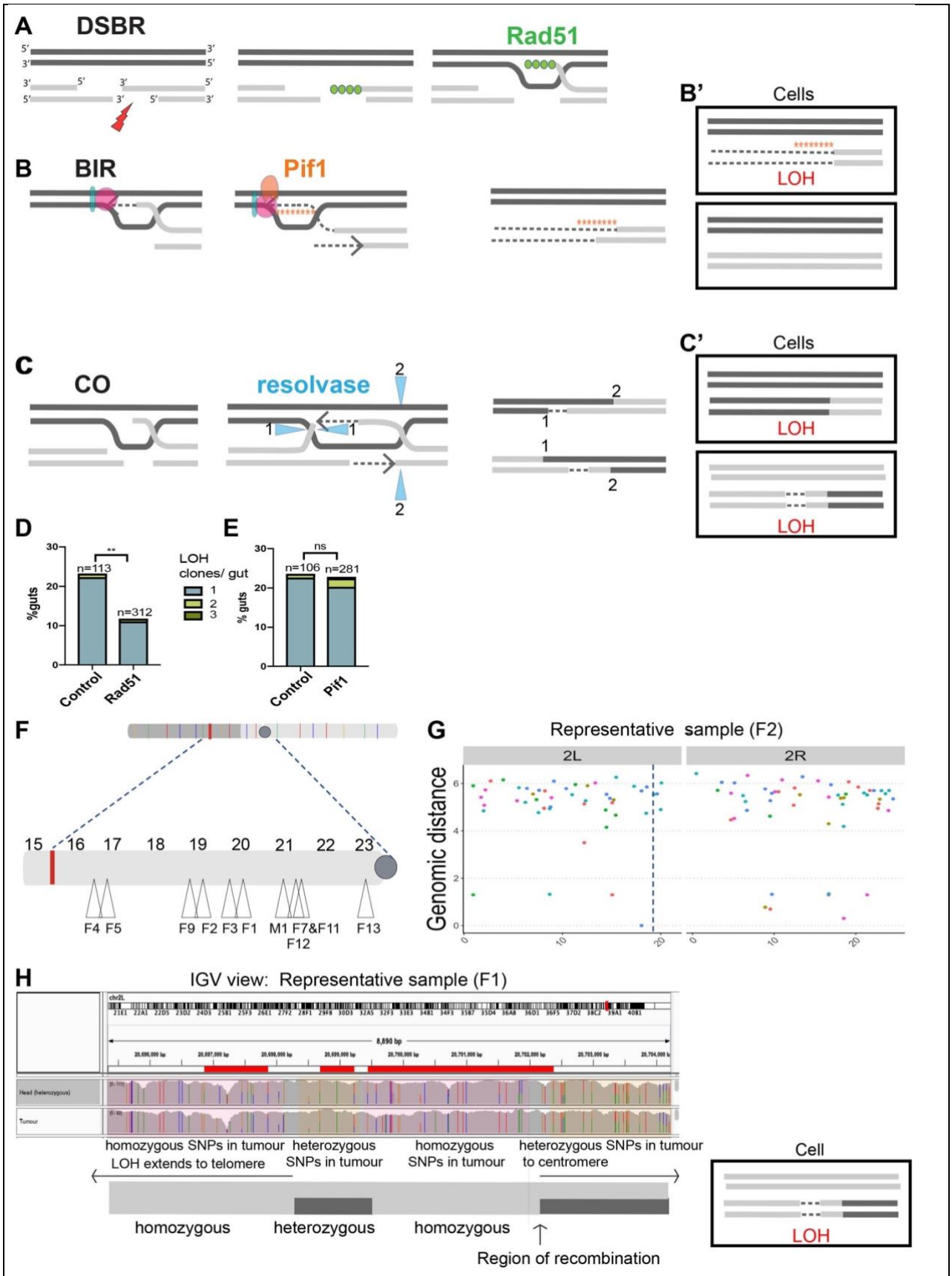


Figure 2.5: Evidence for a cross-over mediated mechanism of mitotic recombination via double-Holliday junction resolution

(A) DSBR model: DSBR is initiated by 5' end resection and Rad51 coating of a 3' end.

(B) 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 (red stars). The resulting chromosomes are shown, where one is unaltered and one has been repaired by the de novo synthesis. The sister chromatids are not shown for simplicity.
(B') Examples of possible progeny cells is shown, whereby the top cell has LOH and the bottom cell is wild-type.
(C) A CO mechanism is mediated by a double-Holliday junction 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.
(C') Examples of resulting progeny are shown. The top cell has undergone LOH without gene conversion tracts whereas the bottom cell represents LOH with short regions of intervening gene conversion (seen by a shift back to homo SNP state). An example of DNA sequencing data supporting events represented in the bottom cell is shown in (G) below.
(D) Frequency of LOH events in control compared to Rad51 knockdown in ISCs. Adult flies were shifted to 29°Cto induce RNAi for 2 weeks then dissected. Fisher's exact test p=0.0048.
(E) Frequency of LOH events in control compared to Pif1 knockdown in ISCs. RNAi was induced in adult ISCs for 2 weeks prior to dissection. Fisher's exact test (ns).
(F) Summary of all mapped recombination regions between *Su(H)* locus (red) and the centromere in female samples (F1-F12) and the male sample (M1). Numbers on the chromosome arm correspond to chromosome coordinates.
(G) Representative rainfall plot showing no mutation pileup by the mapped recombination site in sample F2. Mapped recombination site denoted by dotted line.
(H) 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 tumour sample shows a shift from het-> homo SNPs, then goes from homo->het, before going back from homo->het throughout the rest of the chromosome arm. A schematic of the cell of origin is shown on the right panel and is like the bottom panel of C'.

they differ in several respects. First, BIR, would only alter 1 of the 2 daughter chromosomes, on the other hand, the classic DSBR model involving double-Holliday junction resolution would result in 2 altered homologous chromosomes with reciprocal exchange events being inherited in the resulting daughter cells. Unfortunately, because the ISC divides asymmetrically, giving rise to one stem cell and one daughter that is lost through differentiation, we cannot assay the two resulting chromosome products. Nevertheless, these two mechanisms differ in their genetic requirements. For this reason, we first decided to test roles of DNA repair proteins specific to each mechanism.

BIR relies on the DNA helicase Pif1, important to unwind the double-strand DNA thereby allowing DNA polymerase delta to copy DNA from the homologous donor chromosome **(Figure 2.5B)**. We thus knocked-down *Pif1* in the ISCs of

Su(H)$^{\Delta 47}$/+ flies. The percentage of guts with LOH showed no difference between *Pif1* knockdown (22.8%, n=281) and the control (23.6%, n=106) **(Figure 2.5E)**, suggesting that *Pif1* does not play an integral role in the repair giving rise to the LOH events we see, hinting towards a non-BIR mechanism.

The resolution of double-Holliday DNA structures relies on nucleases/resolvases. We therefore tested a function of *mus81*, using stem cell-specific knockdown as previously described. While *mus81* knockdown in ISCs led to a significant decrease of LOH events compared to controls in one experiment, a second repeat experiment failed to show differences between knockdown and control contexts (**Figure 2.S4).** Further testing additional RNAi lines targeting *mus81* as well as for the Yen1/Gen1 resolvase, *gen*, will clarify whether resolvase activity is important for the LOH that we detect.

Whole-genome sequencing data supports cross over via a double-Holliday structure

In addition to distinct genetic dependencies, DSBR using BIR differs from double-Holliday junction resolution with respect to mutational signatures occurring in proximity to the initial region where LOH arises. We therefore returned to our whole-genome sequencing data to identify potential genomic signatures that may discriminate between these mechanisms. We first had to determine more precisely the regions where LOH was initiated.

For each of the *Su(H)* neoplastic LOH samples (12 females and 1 male), we used LOHcator (https://github.com/nriddiford/LOHcator) as well as IGV to inspect the regions harbouring the shift in zygosity from het->homo SNPs in the LOH

samples, but not in the respective head controls, where the parental chromosomes have distinct, heterozygous SNPs (**Figure 2.5F**, **Figure S5** and methods; recombination sites in samples F8 and F10 due to EC contamination, see **Table S1**). LOHcator assessed the shift in zygosity revealed through assaying informative SNPs along the chromosome arm and determined the centromeric-most informative SNP that is homozygous in the neoplastic sample. We then examined the identified recombination site for mutational features.

We reasoned that a BIR-based mechanism of LOH would result in *de novo* mutational hotspots near the LOH initiation regions (**Figure 2.5B'**). This is well established from previous studies in yeast that have shown that BIR gives rise to mutation hotspots, up to 36kb away from the DSB initiating the recombination sites with mutations increasing up to 2,800-fold compared to spontaneous events (Sakofsky et al. 2014). This is due to the error-prone copying by DNA polymerase delta coupled with inefficient proofreading activity, as well as accumulation of single-stranded DNA behind the replication bubble that is sensitive to nucleases (Sakofsky et al. 2014; Deem et al. 2011). We thus carried out de novo single-nucleotide variant (SNV) calling to assess mutations throughout the genome. None of our samples had evidence for mutation hotspots near the recombination sites (**Figure 2.5G**). Indeed, SNV densities did not significantly change throughout the entire chromosome arms (representative sample in **Figure 2.5G,** see **Figure 2.S6** for all samples). Therefore, these data argue against BIR as driving LOH in ISCs, at least in the 11 samples analysed.

Strikingly, a more in-depth analysis of the sequencing lent further support for the notion that mutant chromosomes were produced by a classic DSBR model via double-Holliday junction intermediates. Resolution of double-Holliday junction

intermediates produces 2 distinct types of crossover products (**Figure 2.5 C-C'**): (1) Those where a crossover occurs directly, swapping homologous chromosome arms, therefore leading to a shift from from het->homo SNPs. These types of events would also be seen in a BIR model. (2) A second class of crossover product (see **Figure 2.5 C-C'**, **lower panel**) would result in intermediate tracts of DNA where SNPs would alternate from het->homo and then homo->het again, followed by het->homo and crossover. These are due to strand invasion and the synthesis process as well as corrections by mismatch repair machinery. Importantly, these particular products should be specific to resolution of double-Holliday junctions. 4/11 samples had evidence for these types of DNA products, where short tracts of heterozygous SNPs were found in regions of LOH nearby the region of recombination (**Figure 2.5H, Figure 2.S7**). The samples showed good tumour purity based on our EC contamination calculation **(Table 2.S1)**, excluding contaminating wild-type cells as a possible reason for this. For example, sample F1 had a shift from het->homo SNP at position Chr 2L: 20701878-20702379 on the genome. 1.5kb more distally, a region of heterozgosity was again detected, clearly supported by 2 informative SNPs, followed by again a shift from het->homo throughout the chromosome arm **(Figure 2.5H)**. Similar shifts from het-> homo and homo->het could be detected on 3 other samples (**Figure 2.S7** and data not shown for sample F13). Importantly, these types of tracts correspond to the predicted outcomes of upon resolution of double-Holliday structures and are detected in mitosis in yeast (Lee and Petes 2010). In addition, the remaining 7/11 samples showed a simple shift from het-> homo, and appear remaining homozygous throughout the chromosome arm, which are also predicted outcomes of Holliday junction resolution (**Figure 2.5 C-C', upper panel**). Thus, altogether, our data strongly suggest that cross-over arises upon

double-Holliday junction resolution, and not BIR, leads to LOH. While this mechanism allows repair of double-strand breaks in stem cells, it leads to LOH of the tumour suppressor *Su(H)* and drives neoplasia initiation.

Mapping of LOH initiation regions provides insight into potential sequence drivers of MR

We next wanted to understand whether DNA sequence features may have contributed to the DNA damaging event driving MR. Interestingly, 2/11 samples had recombination sites that arose within the *Histone Locus Cluster* (**Figure 2.6A, B**). The *Histone Locus Cluster* is an array of 100 copies in tandem each containing 5 Histone genes (H2A, H2B, H3, H4 and H1) (**Figure 2.6C)**. The *Histone Locus Cluster* 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 exclusively transcribed in S phase when Histones are incorporating into newly synthesized DNA, which could

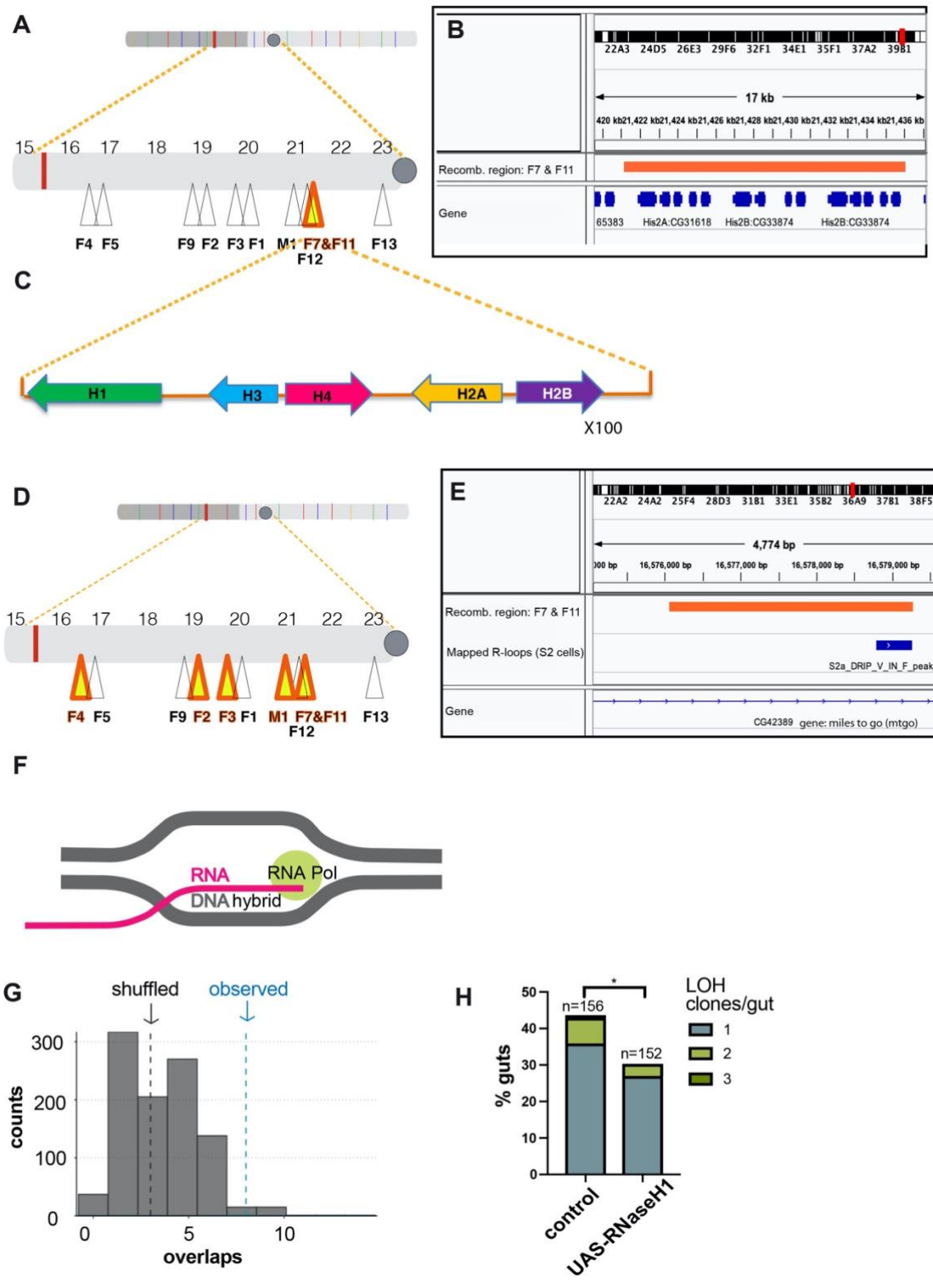


Figure 2.6: Mapping of breakpoint regions revealed potential genomic features that could drive MR

(A) 2 samples had recombination sites that occurred within the Histone Locus Cluster. Due to the repeat nature, the exact positions could not be mapped within this ~100kb region.
(B) IGV view with mapped recombination region of samples F7 and F11 shown in ORANGE.

(C) Schematic of Histone locus cluster- array of 100 copes of 5 Histone genes.
(D) 6 samples showed recombination region overlapping an R-loop (2 samples have the same region, the 2 samples overlapping the histone locus body.
(E) Example of R-loop and recombination region overlap. This region is in a gene that is moderately expressed in the stem cell with a rpkm value of 3.719 (http://flygutseq.buchonlab.com). Annotated symbol is CG42389.
(F) Structure of R-loops.
(G) Permutation test carried out using RegioneR. Test association between mapped region of recombination and mapped Rloop genomic feat. Shuffles: 10000 constricted on 2L. Breakpoint regions >10kb excluded (so the 2 histone locus breakpoints are excluded). Observed value is 7 because 3 mapped breakpoint regions had more than one mapped R-loop. P-value: 0.034.
(H) Frequency of LOH events upon RNaseH1 overexpression in the ISC compared to control with no Dl(ts) driver. Flies were shifted to 29°C after eclosion and dissected after 3 weeks. p-value= 0.018. Fisher's exact test.

possibly make it more prone to replication-fork collisions with the transcription machinery. Consistent incurrence of DNA damage at the *Histone Locus Cluster*, the *HIST1H* cluster in human B-cells is found to acquire marks of γH2AX (Boulianne et al. 2017). Further investigation of the *Histone Locus Cluster* could provide important insight into genomic features potentially driving MR from the homologous chromosome.

We then searched for additional sequence features adjacent to our mapped regions of recombination that could potentially lead to DNA damage. We specifically looked for sequences that could form non-B-form DNA such as G-quadruplexes, short inverted repeats (SIRs), cruciform DNA and R-loops. While we did not find significant overlap with non-B-form DNA such as G-quadruplexes, cruciform DNA (data not shown), we observed that 6/11 samples had recombination sites that were overlapping regions shown to form R-loops (**Figure 2.6D,E**). R-loops are RNA:DNA hybrids that usually form from transcripts of DNA that remain stably base-paired to the template from which they are derived resulting in a three-stranded structure **(Figure 2.6F)**. R-loops consequently cause severe blocks for replication fork progression and have been shown to pose challenges to genome stability, potentially causing a DSB leading to MR as a repair mechanism. We performed

permutation tests on the overlap between R-loops and our mapped regions of recombination and excluded the less resolved recombination regions that were >10kb. Although we did find significant enrichment of R-loops in our recombination sites (p=0.03; **Figure 2.6G**), the number of mapped breakpoint regions is quite low. However, we believe that the association could be of interest as 3/5 of mapped R-loops in our breakpoint regions are in genes shown to be expressed in ISCs http://flygutseq.buchonlab.com. An expanded dataset will be required to determine the potential significance of this association.

There has been increasing interest in R-loops recently as it is becoming clear that they play an important role in genome instability during replication. Despite the cell having ways in getting rid of these R-loops, R-loops often persist and have been shown to be significant sources of damage. One of the ways in which R-loops can be degraded is through RNaseH1 which is a conserved endonuclease that specifically hydrolises phosphodiester bonds of RNA bound to DNA. It has been shown in S2 cells that the expression of *RNaseH1* suppresses R loop formation and genome instability (Bayona-Feliu et al. 2017). We therefore tested the overexpression of RNaseH1 specifically in adult stem cells *in vivo* and found a reduction in LOH events. In the control expressing endogenous levels of *RNaseH1*, we found 43.6% guts with LOH clones (n=156) and a reduction to 30.3% guts with LOH clones (n=152) in guts overexpressing *RNaseH1* (**Figure 2.6H**). Our data, therefore, suggest a potential link with MR sites and R loops that merits further investigation.

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

The gut is an organ that responds rapidly to changes in the environment, allowing rapid stem cell proliferation in response to epithelial cell death (Gervais and Bardin 2017). We suspected that changes in the environmental and proliferative status could affect neoplasia formation due to the fact that we find more guts containing 2 or more LOH events than would be predicted. At 6 weeks of age, 33.7% of guts had 1 LOH event (**Figure 2.1E**). Based on this frequency, the chances of having 2 events, if they were to occur randomly would be 0.33X.033=0.1 (10%). Instead, we found that 26.5% had 2 events (**Figure 2.1E, Figure 2.7A**). The fact that the chances of having two LOH events is higher than expected, could be explained by environmental factors driving a higher proliferative gut status that would therefore result in an increased probability of DNA damage in multiple ISCs in those guts. We therefore wanted to explore potential environmental factors that might promote MR and LOH. It is known that X-ray irradiation causes chromosomal breaks that induce MR and can lead to LOH of the wild-type allele (Stern 1936a; Griffin et al. 2014). Therefore, as a positive control, we irradiated young 1-week old *Su(H)*$^{\Delta 47}$/+ flies with 40 Gray ionising radiation (IR) and compared them unirradiated flies at 3 and 6 weeks post-IR (**Figure 2.7B**). While non-IR treated flies contained 10.8% spontaneously arising LOH events at 3 weeks, IR treated flies had a significant increase in LOH events, with 78.4% of midguts containing mutant clones at 3 weeks (n=97; **Figure 2.7B**). In addition to that, most IR-treated midguts had multiple events (61.9%) and 9.3% midguts had more than 5. At 6 weeks, almost 100% of IR-treated

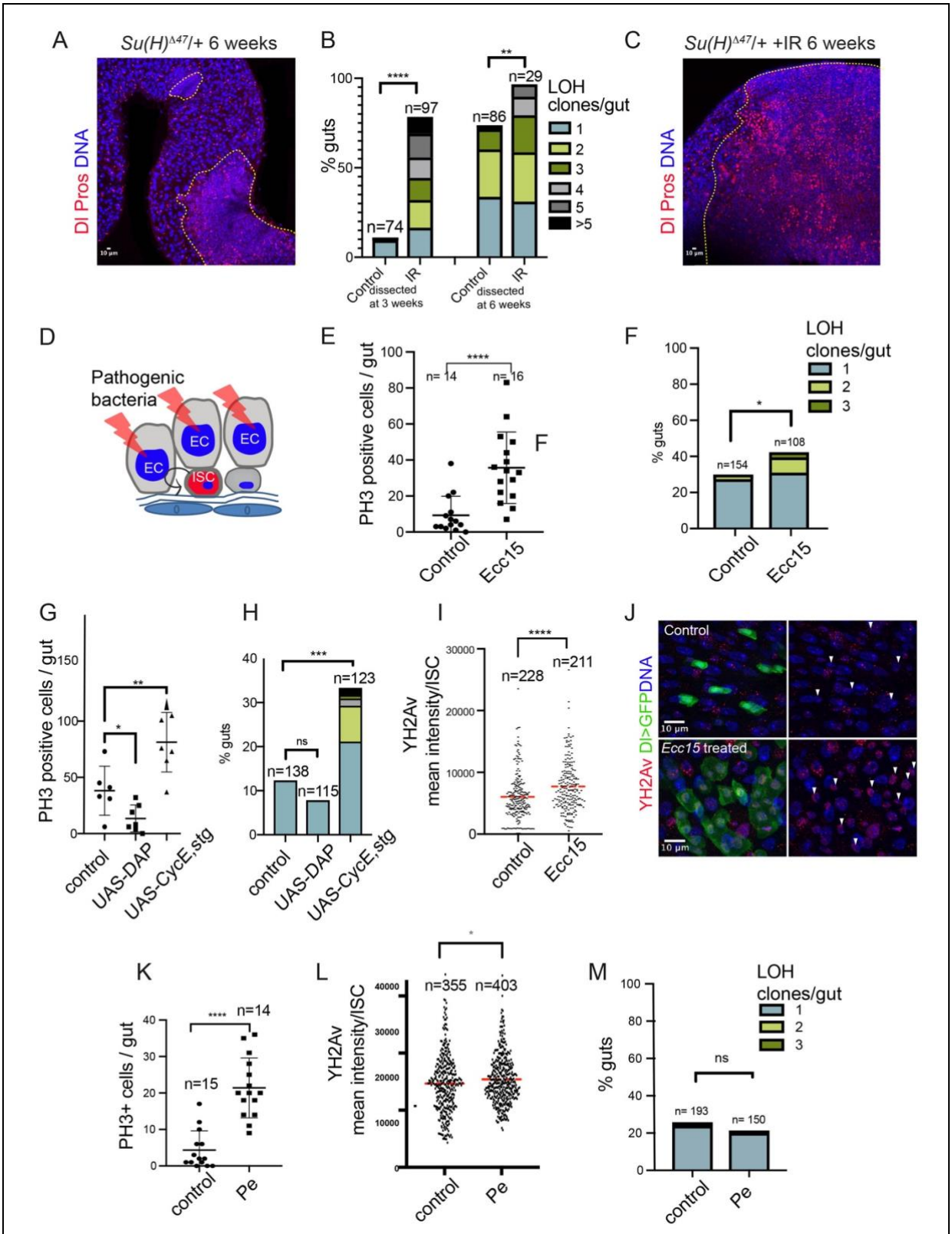


Figure 2.7: Environmental factors increasing loss of heterozygosity

(A) Example of 2 LOH events in aged *Su(H)*$^{\Delta 47}$/+ gut. Dl (ISCs, cytoplasmic RED), Pros (EEs, nuclear RED), DAPI (BLUE).
(B) Irradiated 1-week old *Su(H)*$^{\Delta 47}$/+ flies with 40 Gray (IR) compared to unirradiated flies at 3 and 6 weeks post-IR, showed significant increase in LOH frequency.
(C) Example of large LOH in IR treated aged *Su(H)*$^{\Delta 47}$/+ gut. . Dl (ISCs, cytoplasmic RED), Pros (EEs, nuclear RED), DAPI (BLUE).

(D) Infection with pathogenic bacteria kills EC cells and stimulates ISC proliferation.
(E) 3-5 day old flies dissected immediately after Ecc15 treatment to check for a proliferative response. Stat test: t-test with Welch's correction.
(F) Frequency of LOH clones in Ecc15 treated flies compared with control (Fisher's exact test, two-tailed).
(G) PH3 quantification of guts with UAS-DAP and UAS-CycE, Stg expressed in stem cells compared with control. Flies were shifted to 29°C immediately after eclosion and dissected after 1 week. T-test with Welch's correction.
(H) Frequency of LOH midguts overexpressing cell cycle genes compared with control (Fisher's exact test, two-tailed).
(I) Comparison of YH2Av mean intensity in ISC in Ecc15 treated and control flies. Treatment= 24 hours. Age of flies: 3-5 days old. T-test with Welch's correction.
(J) Representative images of YH2av in ISCs. Control, top panel. Ecc15 teated, bottom panel.
(K) 3-5 day old flies dissected immediately after Pe treatment to check for a proliferative response. Stat test: t-test with Welch's correction.
(L) Comparison of YH2Av mean intensity in ISC in Pe treated and control flies. Treatment= 24 hours. Age of flies: 3-5 days old. T-test with Welch's correction.
(M) Frequency of LOH clones in Pe treated flies compared with control (Fisher's exact test, two-tailed).

***p < 0.001; ****p < 0.0001; n.s., not significant (Fisher's exact test, two-tailed).

flies had very large LOH clones (shown in **Figure 2.7C**), with many occupying almost the entire midgut, strongly suggesting clone fusion. This leads to the inability to distinguish between separate clones due to clone fusion, which was likely reflected in the increased percentage of 1,2 and 3 clones/ gut and the decrease in >5 clones/ gut at 6 weeks compared to 3 weeks in the irradiated flies (**Figure 2.7B**). Importantly, these data illustrate that environmental changes to ISCs during adult life can increase the LOH frequency in the midgut.

We then assessed whether feeding with the pathogenic bacterial strain *Erwinia carotovora carotovora 15* (*Ecc15*), known to kill gut EC cells and promote ISC proliferation for (Jiang et al. 2009c) **(Figure 2.7D),** could alter the frequency of LOH. We treated $Su(H)^{\Delta 47}$/+ flies with punctual exposures to *Ecc15* during 24 hours, once per week in weeks 1, 2 and 3, providing time for recovery to minimise overall toxicity and avoid lifespan reduction. As previously reported, *Ecc15* treatment led to an increase in phospho-Histone 3+ cells after 24 hours of treatment (**Figure 2.7E**). We found that *Ecc15* treatment gave rise to significantly more LOH events at 5

weeks than controls, going from 29.9% of guts with at least one LOH event in controls to 42.3% upon *Ecc15* treatment (**Figure 2.7F**). In addition, an increase in the number of guts with more than 1 event was also found (**Figure 2.7F**).

We suspected that the effect of *Ecc15* might be due to the increased number of cell divisions due to stimulation of stem cell proliferation, thereby increasing the likelihood of replicative DNA damage. In order to alter cell division rates, we used combined overexpression of *Cyclin E* and *string*, previously shown to increase ISC proliferation (Kohlmaier et al. 2015) and the overexpression of the cyclin-dependent kinase inhibitor, Dacapo (Dap), known to slow cell division (Lane et al. 1996). Cell cycle genes were specifically expressed in the ISC using Dl^{GAL4} combined with *tub-GAL80*^{ts} (**see Methods**). Consistent with enhanced stem cell proliferation promoting LOH, we found that LOH events significantly increased from 12.3% (n=138) in controls, to 33.3% (n=123) in *UAS-CycE, Stg* conditions (**Figure 2.7G, H**). *Dap* expression did not significantly impact LOH frequency. Thus, enhanced proliferation, can increase LOH frequencies. This suggests that *Ecc15* treatment could impact LOH due to its stem cell proliferative effect. In addition, however, we found that *Ecc15* treatment significantly increased marks of DNA damage in the ISC detected through a stronger staining of γH2Av (fly γH2Ax) in the ISCs compared to the untreated controls (**Figure 2.7I, J**). Whether the impact of *Ecc15* on DNA damage is linked to increased stem cell proliferation, or whether this bacteria can damage DNA through an additional mechanism such as ROS production, is unclear. We conclude that *Ecc15* infection can induce stem cell DNA damage, drive ISC proliferation and promote LOH.

Interestingly, we found that another pathogenic bacteria, *Pseudomonas entomophila* (*Pe*), behaved differently. Treatment with *Pe* increased ISC

proliferation as previously reported (**Figure 2.7K**,(Jiang et al. 2009c), increased DNA damage marks (**Figure 2.7L**), though to a lesser extent than *Ecc15*, but did not cause an increase in LOH frequency (**Figure 2.7M**). Our data thus show that environmental changes, such as that of the gut microbiota, can influence the frequency of LOH events but that different bacteria can lead to different effects.

Discussion

Using whole-genome sequencing approaches, we demonstrated that a major contributor to genome alteration in *Drosophil*a ISCs is through MR-driven LOH. Our data suggest that MR depends, at least in part, on *Rad51* though is *Pif1* independent, suggesting a DSBR model whereby double-Holliday junction resolution leads to cross-over thereby promoting LOH. Our further analysis of genome sequencing using parental SNP information and *de novo* SNV profiling supports this notion: we failed to detect mutational pile-ups that commonly arise with BIR and, instead, found clear examples of chromosomes resulting from double-Holliday junction-based repair. Our sequencing data also point towards DNA features that might be prone to damage including the *Histone Locus Cluster* and regions that incur R-loops. Finally, we show that environmental factors, such as the pathogenetic bacteria, *Ecc15* promoted an increase in the frequency of LOH. Our findings reveal essential intrinsic and extrinsic factors acting on DNA damage and repair in adult stem cells to influence LOH, an important mechanism of tumour-suppressor inactivation.

Our study, for the first time, demonstrates that spontaneous LOH occurs in adult intestinal stem cells through MR and provides high resolution SNP-based mapping of recombination regions. We showed evidence of LOH via MR occurring

in 13 (12 females, 1 male) samples on chromosome 2L, 1 on chromosome 2R and 1 on chromosome 3R. Furthermore, we found no evidence supporting a BIR mechanism for the 17 samples that we assessed. Expansion of our dataset will allow us to determine whether BIR may be used in some instances and to what extent other types of mutation may promote LOH in this model. Our recent studies illustrated that, in addition to genome alteration by recombination, point mutations, deletions, complex structural variants, and transposable element mobility all alter the somatic genome of ISCs and could therefore contribute to LOH (Riddiford 2020; Siudeja and van den Beek 2020). However, given that 11/12 LOH samples were associated with cross-over and 4 showed clear evidence of being created through the resolution of double-Holliday junction intermediate, our data strong support MR being the primary mechanism by which LOH is generated in ISCs.

Our analysis of LOH in response to spontaneous DNA DSB-induced damage, relied on genome-wide approaches to map recombination sites occurring between homologous chromosome arms that led to cross-over events and phenotypic changes in the gut. It is possible that a much higher number of DSB are produced but repaired of the sister chromosome, as this is thought to be the predominant mechanism of DSBR repair. Events where repair occurs off of the sister chromosome would not result in LOH or give rise to phenotypic consequences and would, therefore, be undetectable in our assay.

Our study underscores the notion that adult organs are genetic patchworks, where somatic mutations drive divergent genomes within a common tissue. This is consistent with recent findings in human tissues where it is has been demonstrated that mutations arise and provide selective growth advantages of mutant lineages within normal or premalignant tissues, reviewed in (Al zouabi and Bardin 2020). For

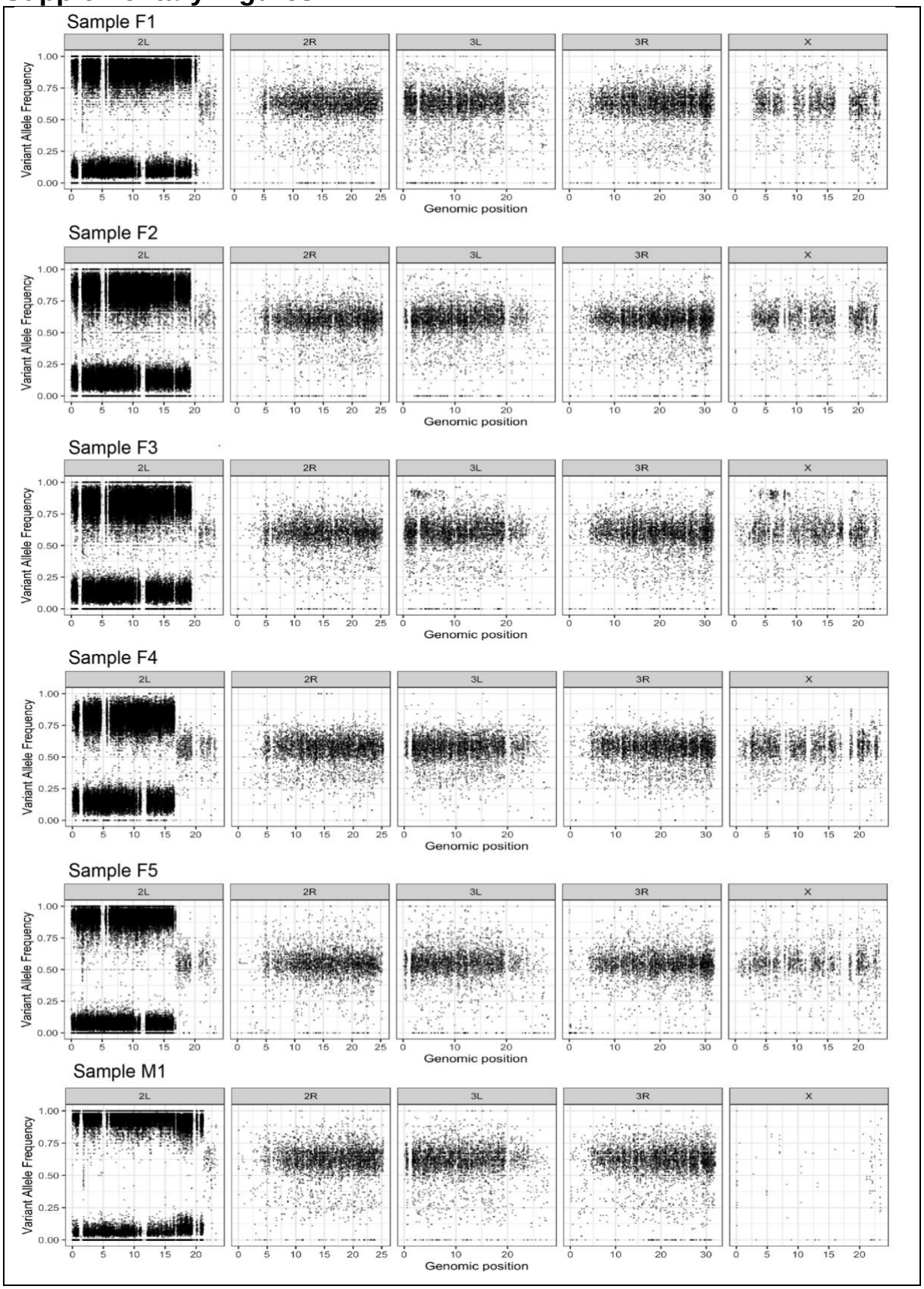
example, clonal hematopoiesis of the blood is driven by *de novo* mutations of the epigenetic regulators *TET2*, *DNMT3A*, *ASXL* (Busque et al. 2012; Genovese et al. 2014; Jaiswal et al. 2014; McKerrell et al. 2015; Coombs et al. 2017). It has also been recently shown using targeted sequencing that somatic inactivation of *NOTCH1*, *NOTCH2, NOTCH3* as well as *p53* occurs frequently in aged human skin as well as esophagus (Martincorena et al. 2015, 2018b; Yokoyama et al. 2019). Strikingly, clonal blood mutations have been found in 50% of individuals over 85 years of age (Zink et al. 2017a). Similarly, the mutations found in aged human skin and esophagus are prevalent in all individuals and affect large fractions of the tissue, for example with 18-32% of skin in adult individuals being mutant for cancer driver genes (Martincorena et al. 2015). Here, we found that LOH of the *Su(H)* locus, 7.5MB from the centromere chromosome 2L, occurred in 73% of aged individuals. In addition, LOH of the *Notch*, 20MB from the centromere of the X chromosome, was found even more frequently with ~80% of flies with at least one LOH event and a majority of guts with multiple independent LOH events, for an average of ~8 LOH clones per gut (Siudeja et al. 2015). Thus, if we apply this number and consider all 5 major chromosome arms in *Drosophila*, it is likely that 40 LOH 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 for at least part of one chromosome arm, raising the likelihood that this mechanism of somatic genetic diversity could alter tissue dynamics during the ageing process.

What are the underlying causes of such a high level of genome mutations? Our sequencing data point towards DNA features that might be prone to damage including the *Histone Locus Cluster* and regions that incur R-loops. A larger sequencing data set of spontaneous LOH events will be needed to lend support to

this hypothesis. The overexpression of *RNaseH1*, shown in S2 cells to reduce R-loop formation, showed reduced LOH frequency further suggesting that R-loops may be implicated in DNA damage in ISCs, though it remains to be verified that this treatment can reduce R-loop formation in ISCs. Nevertheless, our novel SNP mapping strategy of somatic LOH clones in the gut will be a powerful tool towards further investigating potential sequence features causing damage that is repaired through MR.

Interestingly, we could also show that DNA damage is induced through non-cell autonomous interaction with environmental factors, such as the pathogenic bacteria *Ecc15*. It is not clear whether the effect of *Ecc15* on DNA damage is linked to the increased amount of cell division occurring in response to treatment or whether it is independent of cell proliferation. Manipulating cell proliferation through overexpression of *CycE* and *String* to promote cell division, was sufficient to dramatically increase LOH events, suggesting that the number of replicative divisions a stem cell undergoes is an important variable, consistent with the proposal put forward by Tomasetti and Vogelstein (Tomasetti and Vogelstein 2015). Counterintuitively, however, treatment with the *Pe* bacteria, which could also promote cell proliferation and increase stem cell DNA damage, did not lead to increase LOH events. This could be due to the lower amount of DNA damage that *Pe* induces. Alternatively, Pe has been shown to block translation, which might, in turn, reduce the responsiveness of the cell to DNA damaging and impinge on DNA repair. Determining how additional pathogenic bacteria and other changes to environmental conditions may affect LOH events will be an important future goal.

Supplementary Figures



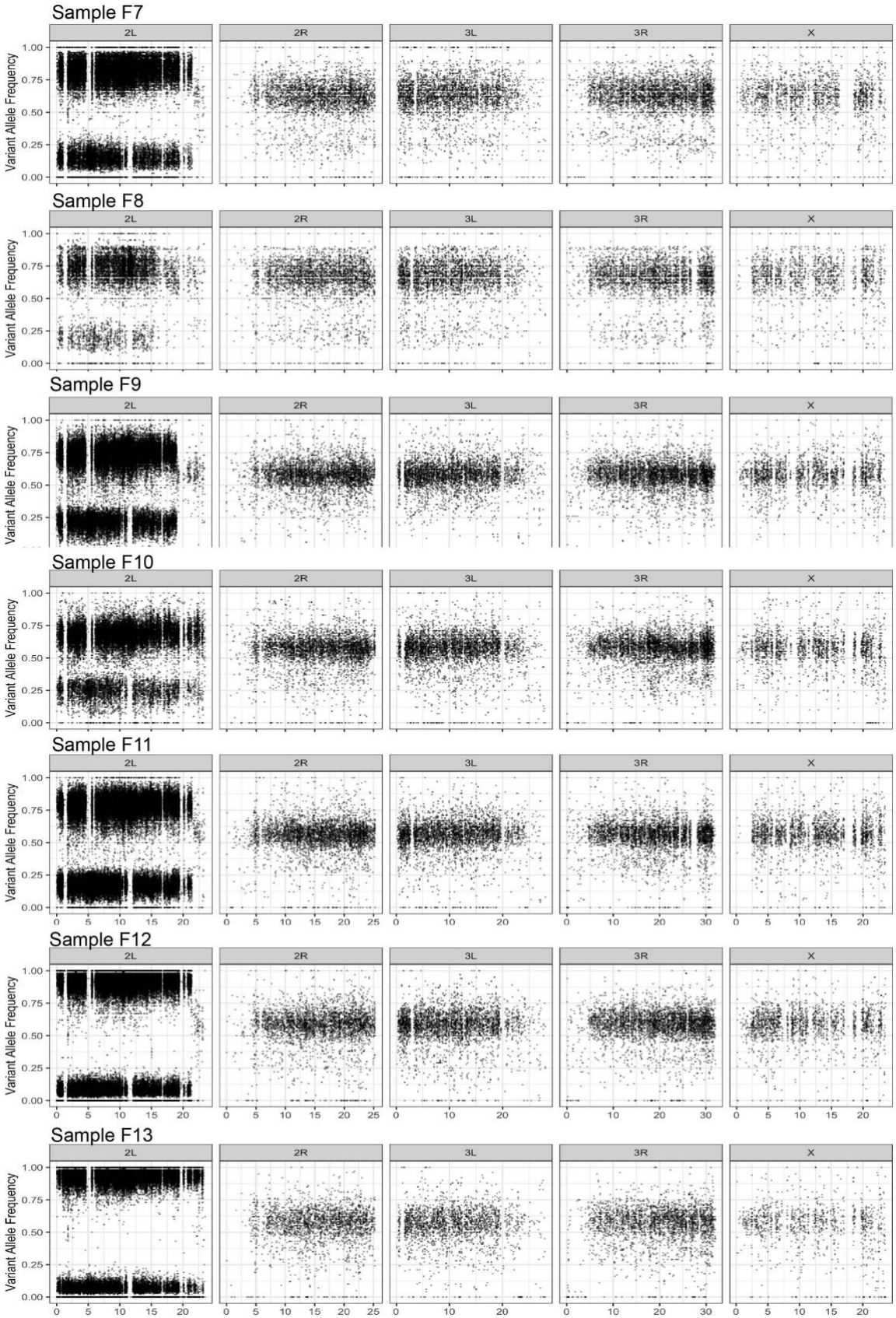


Figure 2.S 1: Variant allele frequency plots of samples supporting LOH of chromosome 2L

Plots of all female LOH samples, showing all chromosomes. Chromosome 2L has undergone an LOH event, with SNPs becoming homozygous (VAF ~1) shown in the left panels.

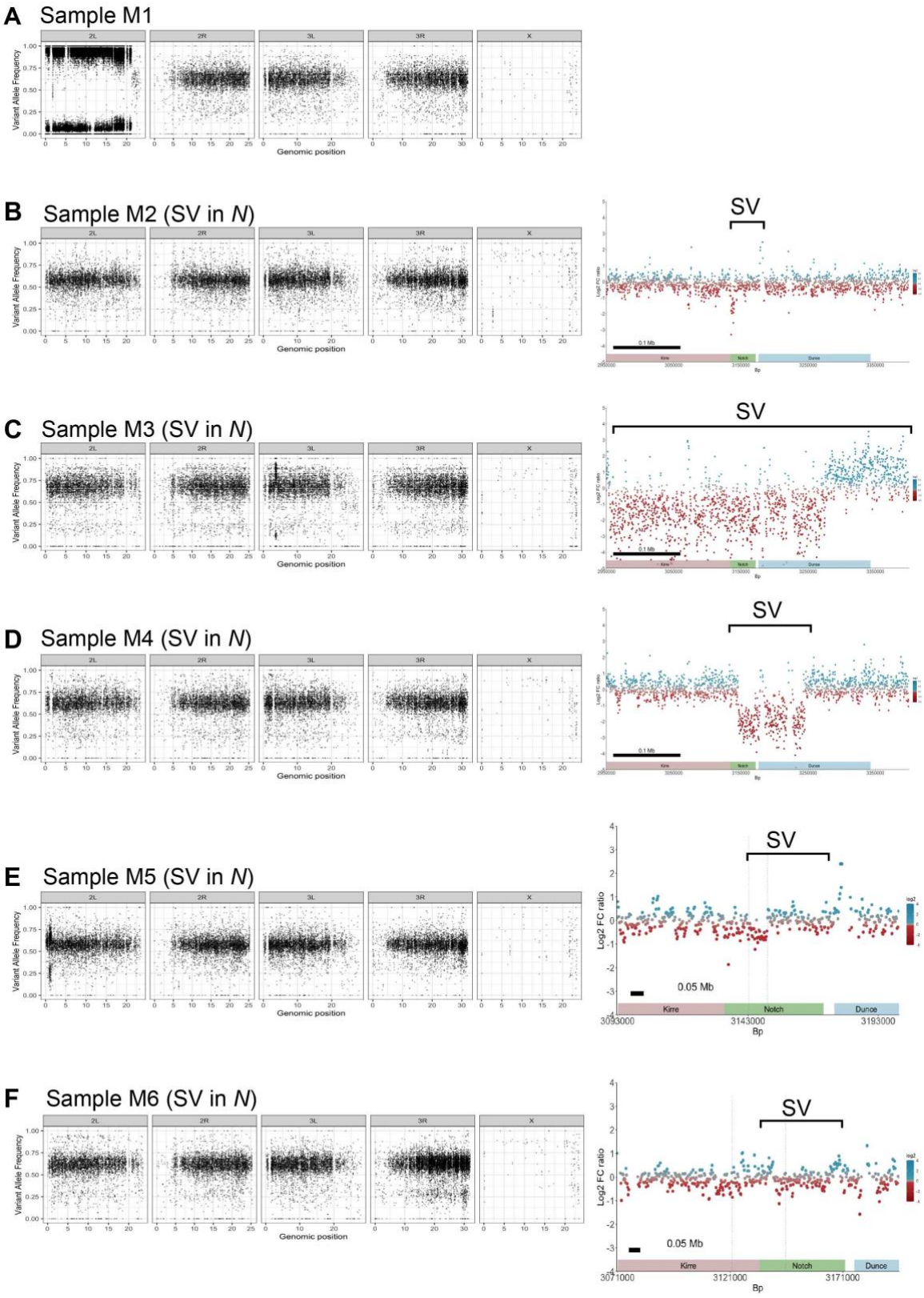


Figure 2.S2: Variant allele frequency plots of all male samples

Plots of all male samples, showing all chromosomes. Only sample M1 (A) shows that chromosome 2L has undergone an LOH event, with SNPs becoming homozygous (VAF ~1). The remaining samples (B-F) show no LOH on chromosome 2L (left panels) but rather show a structural variant (SV) panning the Notch locus on the X chromosome (right panels).

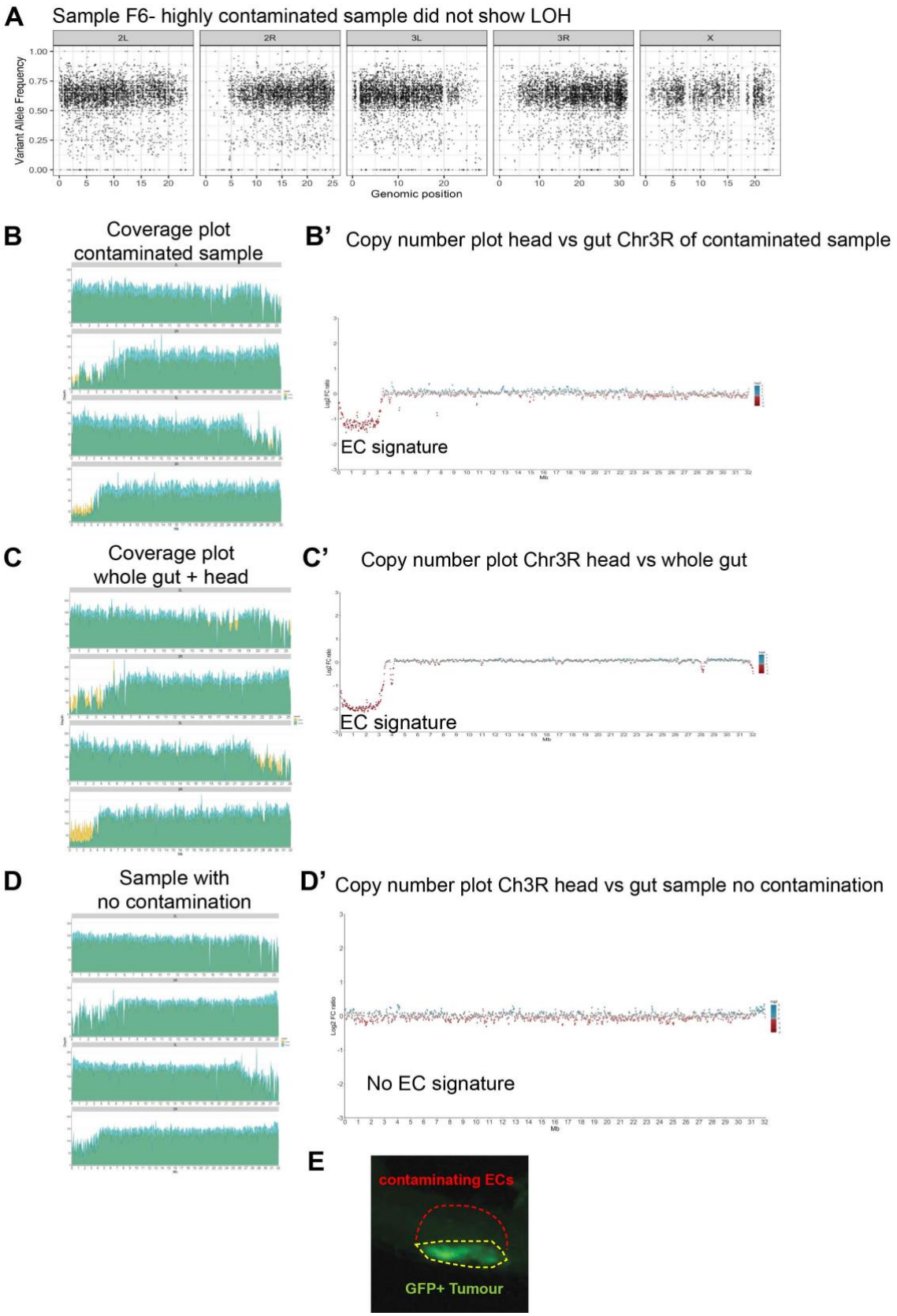


Figure 2.S 3: Enterocyte signature as a proxy for contamination

(A) Sample F6 had no detectable LOH based on VAF plots.
(B) Coverage plot of head (YELLOW) and neoplasia DNA (GREEN) showed evidence of genomic regions in the neoplastic sample that had low coverage (see YELLOW but no GREEN).
(B') Copy number plot of (B), where a large drop in coverage of the centromere-proximal part of chromosome 3R (Ch3R) is detected. This is suggestive of contaminating EC cells in the dissection as these polyploid cells do not fully replicate this part of the genome (see C)
(C) Coverage plot of sequencing data from a whole gut and head. Whole guts are primarily composed of EC cells, therefore, the under-replicated regions of EC cells are apparent (see YELLOW- head- but no GREEN-gut).
(C') Copy number plots of C, showing under-replicated region on Ch3R, "EC signature".
(D-D') Example of sample F4 showing no EC contamination.
(E)Image of dissected gut neoplasia. If surrounding area outlined in RED is included in the dissection, an EC signature can result due to contaminating DNA of ECs that do not fully endo-replicate all of their genome. We used this signature to determine which samples are not pure enough for analysis.

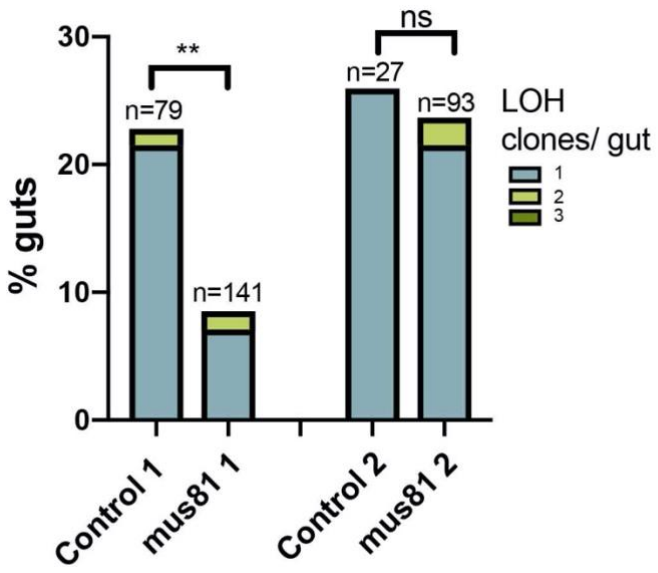


Figure 2.S 4: Variability of mus81 RNAi knockdown experiment.

Two replicate experiments showed different results, with the first one had a significant decrease upon RNAi expression compared to the control and the second replicate had no significant difference between the two experiments. The reason for the variability is unclear.

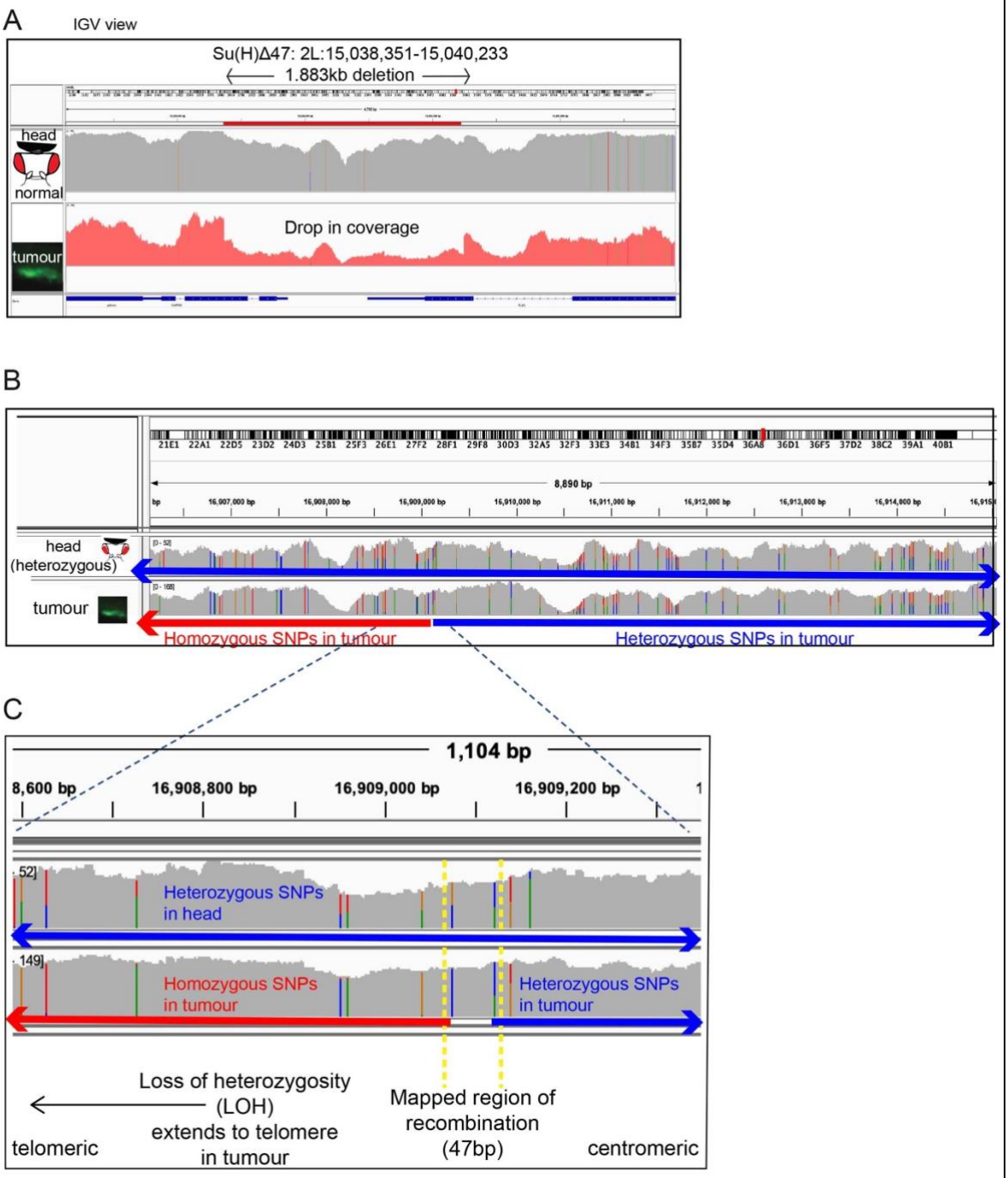
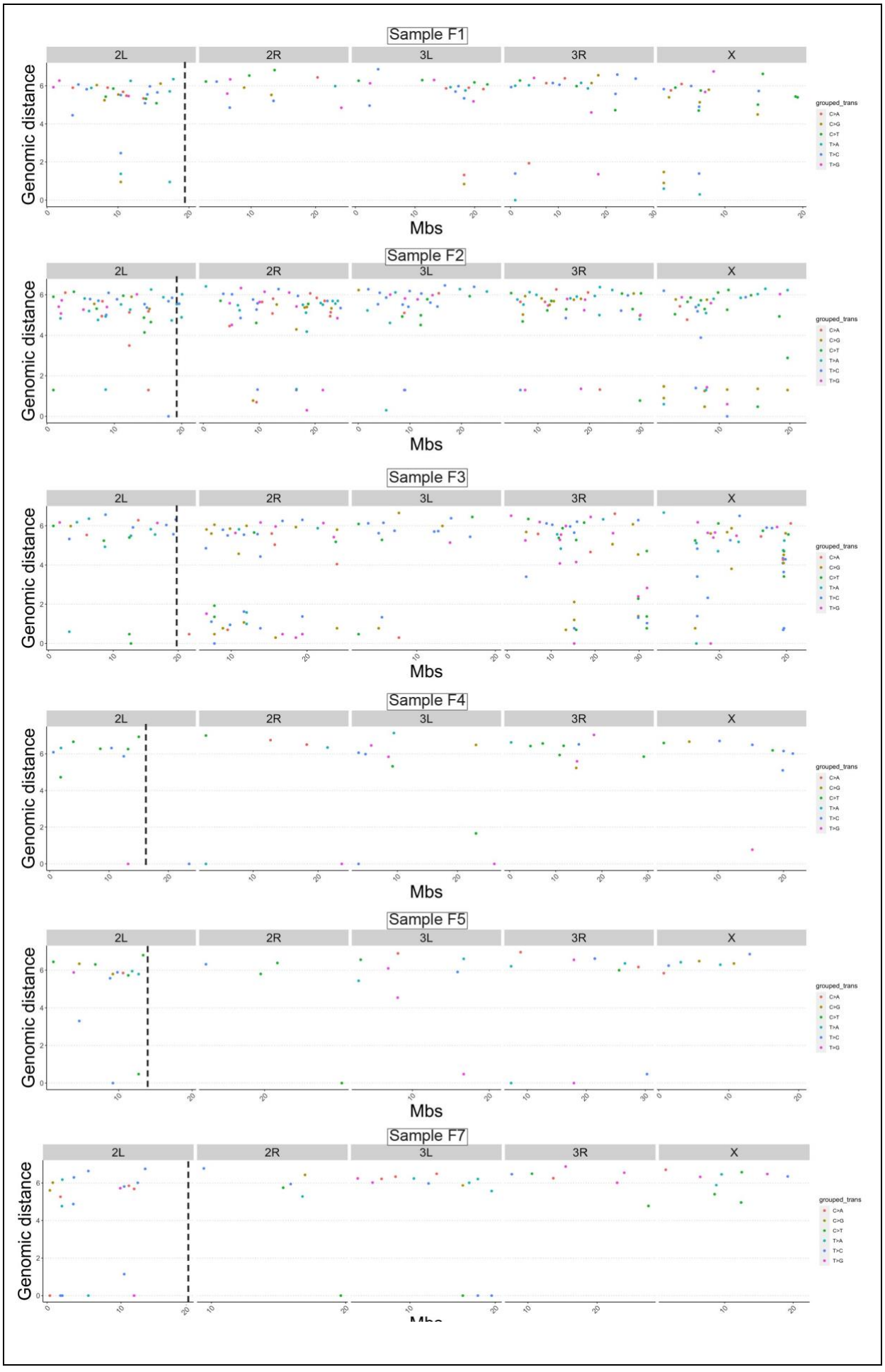


Figure 2.S 5: Mapping of recombination sites in LOH samples compared to heads

(A) An example of the drop in coverage detected at the *Su(H)* locus in LOH samples, sample F5 is shown.
(B) Sequencing data visualized on IGV of F5. SNPs relative to the reference genome are marked in colours. Both head and tumour (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, where they are still heterozygous in the head.
(C) A zoom of the region where recombination took place. We mapped within 47bp the last heterozygous SNP and the first homozygous SNP.


Sample F1
Sample F2
Sample F3
Sample F4
Sample F5
Sample F7
2L
2R
3L
3R
X
Genomic distance
Mbs
grouped_trans
C>A
C>G
C>T
T>A
T>C
T>G

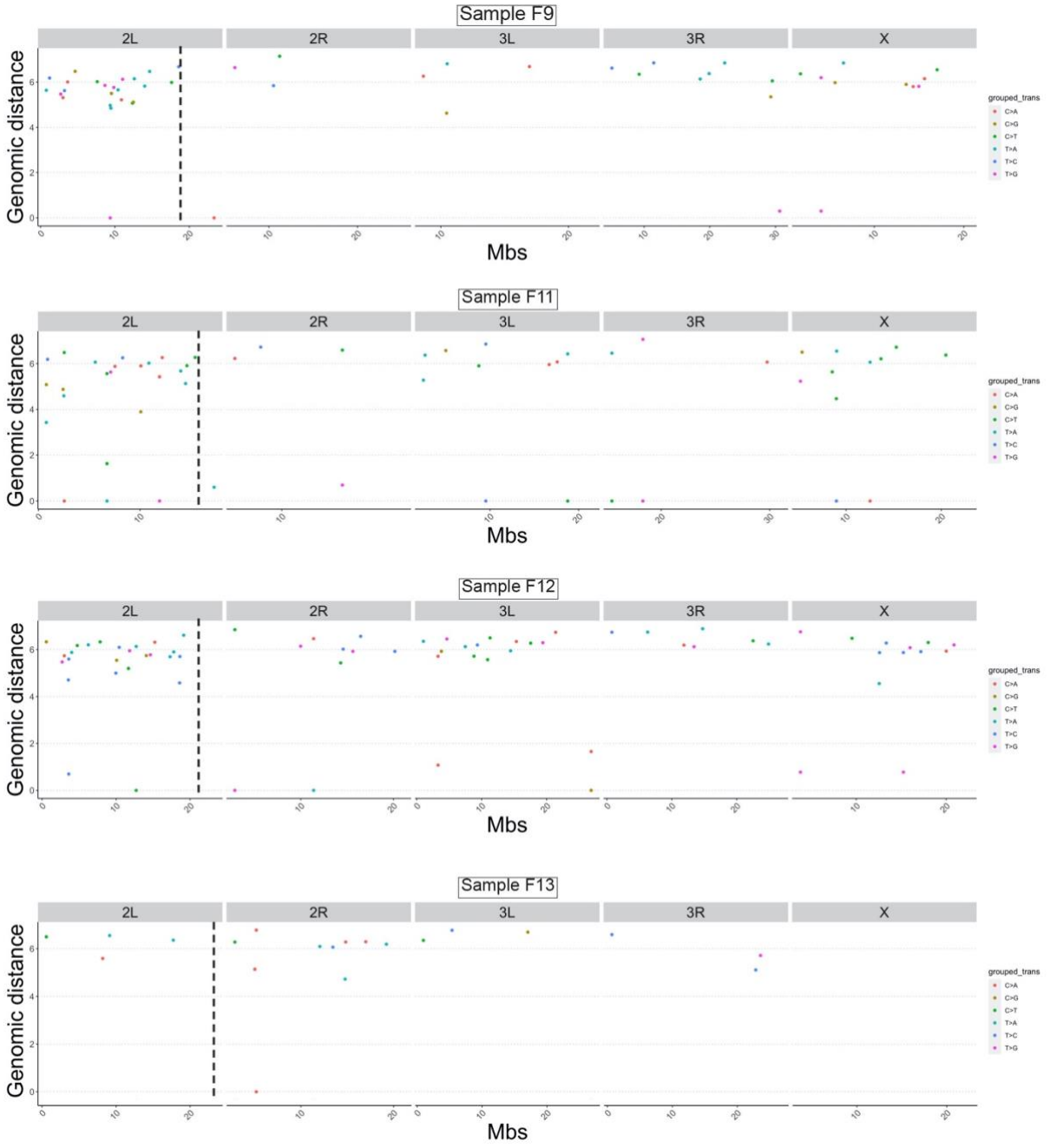


Figure 2.S 6: Rainfall plots of point mutations of all samples.

Each point represents the genomic distance between two SNVs (point mutations). A mutational pileup would be detected by an accumulation of points vertically. The dotted line represents the mapped recombination site. We observed no mutational hotspots on chromosome 2L, arguing against a BIR model.

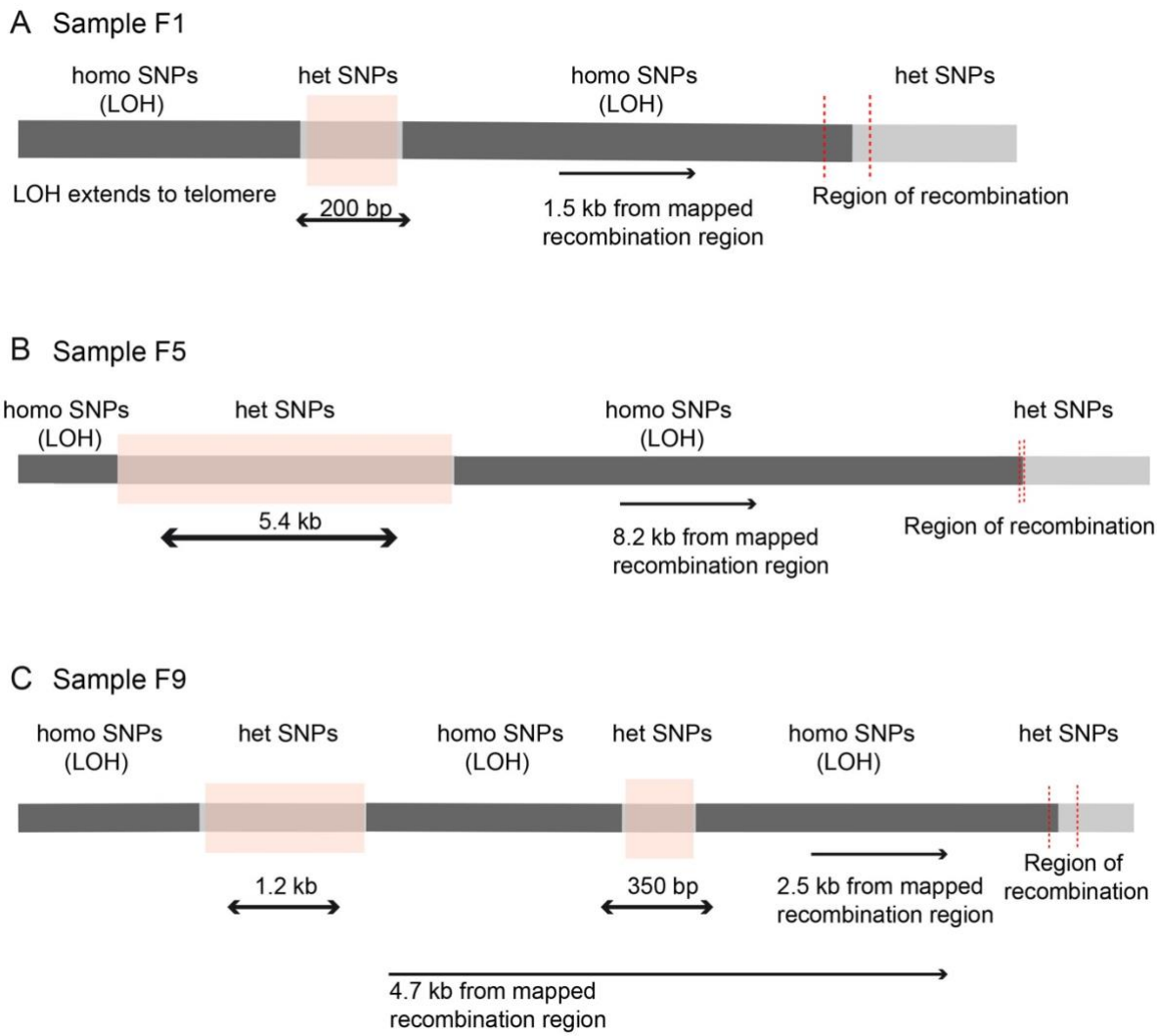


Figure 2.S 7: Schematic of sequencing data showing DNA tracts likely generated from resolution of double-Holliday junction structures

(A) Schematic of sample F1 region of recombination shown with red lines, as the centromere-proximal most region, where heterozygous SNPs were found to become homozygous in the tumour sample. A region approximately 1.5 kb away showed a conversion tract denoted by shift from homo-het 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 (RED) where there is found a shift from het->homo SNPs in the tumour, 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.

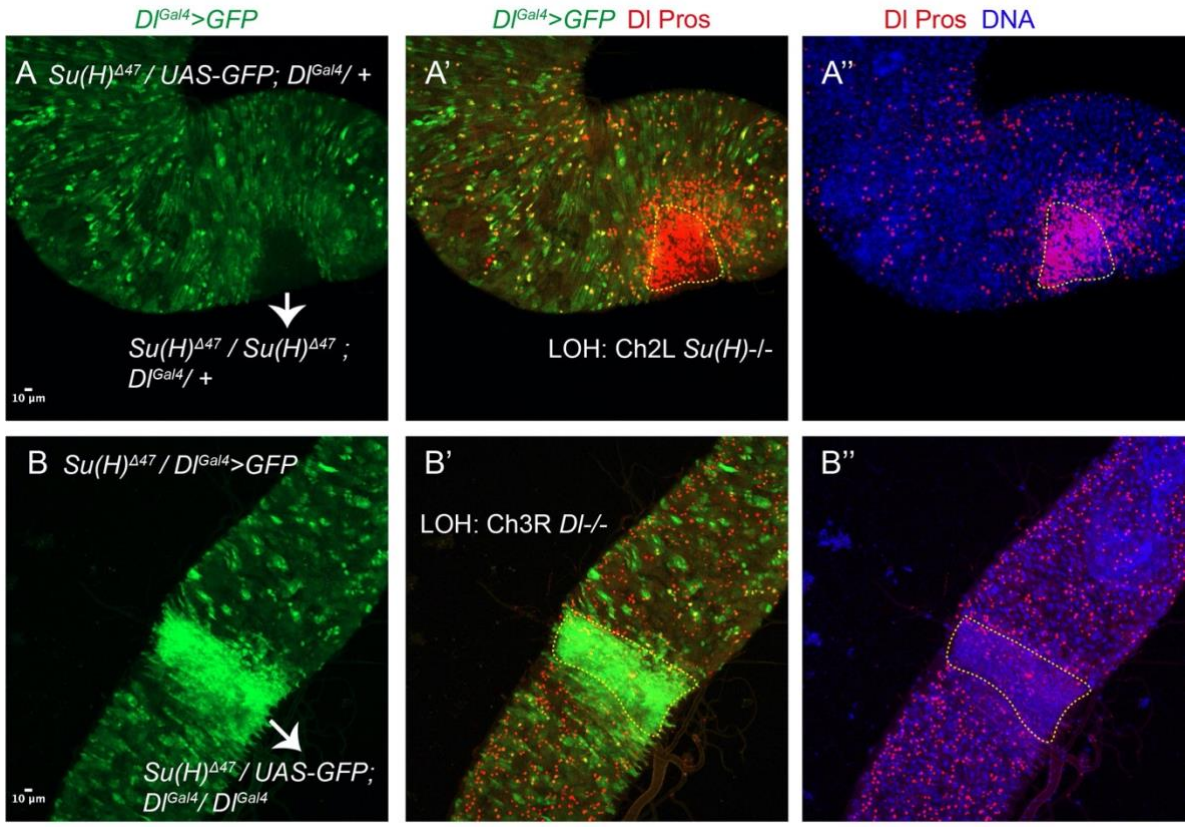


Figure 2.S 8: Quantification of LOH of Ch2L and 3R in the same flies

In experiments in Figures 5, 6, and 7 the genetic background *Su(H)*$^{\Delta 47}$*/ UAS-GFP*; *Dl*^{Gal4}*/ +* was used to drive stem cell specific expression in the ISC, due to the *Dl*^{Gal4} driver. However, as *Dl*^{Gal4} is a loss of function allele of *Dl*, also a *Notch* signaling component, two types of neoplastic LOH clone could arise: those where *Su(H)* is inactivated (A-A") and those where *Dl* 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 (A-A"). These events were scored in Figures 2.5-7. In contrast, LOH resulting from other events including mitotic recombination of 3R leading to *Dl* LOH, are GFP+ (B-B"). These events were scored, but not counted in the analysis of Figures 2.5-7.

Supplementary Table 1

Name of sample	Depth	Purity determined by EC contamination	2(purity) x depth	Able to LOH mechanism?	Able to map recomb. Region?
F1	41.62X	1.14	94.89	yes	yes
F2	50.61X	1.07	108.31	yes	yes
F3	47.30X	1.26	119.20	yes	yes
F4	63.40X	1.01	128.07	yes	yes
F5	79.34X	1.27	201.52	yes	yes
F6	49.42X	**0.44**	43.49	no	no
F7	30.80X	0.88	54.21	yes	yes
F8	23.62X	**0.66**	31.18	yes	no
F9	65.06X	0.76	98.89	yes	yes
F10	59.70X	**0.74**	88.36	yes	no
F11	72.50X	0.87	126.15	yes	yes
F12	49.87X	1.15	114.70	yes	yes
F13	48.95X	1.55	151.75	yes	yes
Excluded F14	**6.89X**	/	/	no	/
Excluded F15	**9.02X**	/	/	no	/
M1	39.39X	1.11	87.45	yes	yes
M2	65.09X	1.00	130.18	SV in *Notch* (not LOH)	/
M3	26.04X	0.68	35.41	SV in *Notch* (not LOH)	/
M4	30.00X	0.97	73.72	SV in *Notch* (not LOH)	/
M5	72.99X	0.68	99.27	SV in *Notch* (not LOH)	/
M6	45.85X	0.68	62.36	SV in *Notch* (not LOH)	/

Supplementary Table 2

Name of sample	Recombination region coordinates (chr2L:)	Recomb. Region esolution (bp)	Purity determined by EC contamination	Depth	2(purity) x depth	Detected heterozygous SNPs in LOH region?	Overlapping R-loop?	Other
F1	20701878-20702379	502	1.14	41.62X	94.89	yes (1)	no	
F2	19280917-19282776	1859	1.07	50.61X	108.31	no	yes	
F3	19695528-19725215	29687	1.26	47.30X	119.2	no	yes	Long region of homozygosity
F4	16576057-16579279	3222	1.01	63.40X	128.07	no	yes	
F5	16909074-16909121	47	1.27	79.34X	201.52	yes (1)	no	
F7	21421161-21436036	14875	0.88	30.80X	54.21	no	yes	Unmappable region (histone locus)
F9	18998061-18998705	644	0.76	65.06X	98.89	yes (2)	no	
F11	21421161-21436036	14875	0.87	72.50X	126.15	no	yes	Unmappable region (histone locus)
F12	21313913-21314794	881	1.15	49.87X	114.7	no	no	
F13	23213084-23218729	5645	1.55	48.95X	151.75	yes- complex (3+)	no	
M1	21156863-21163218	6355	1.11	39.39X	87.45	no	yes	

Acknowledgements
High-throughput sequencing has been performed by the ICGex NGS platform of the Institut Curie. We thank S. Lameiras and M. Bohec for the library preparations. This work was supported by grants from Fondation pour la Recherche Médicale (FRM) as well as funding from Worldwide cancer Research (WWCR). Benjamin Boumard collected samples for the *Dl*^{Gal4} sequencing background.

Author Contributions
L.A and A.J.B. designed the study. N.R developed the R scripts for bioinformatic analysis. These include scripts for generating coverage plots, copy number plots, rainfall plots, Variant allele Frequency plots and the permutation test for the genomic features enrichment analysis. L.A ran the scripts and generated the plots for the figures. L.A performed the analysis and performed the majority of the experiments. M.S. acquired images for the analysis of figures 2.7 I and L. M.S also helped in fly selections, dissections, staining and mounting of flies from figures 2.5 D and E, 2.6 H, 2.7 G, H, I, J and 2.S4. L.A and A.J.B wrote the manuscript.

2.2 Aneuploidy as a mechanism driving spontaneous loss of heterozygosity in *Drosophila* intestinal stem cells

Context

In section 2.1, we showed through whole genome sequencing that MR was a major mechanism driving the loss of heterozygosity in the fly intestine. MR accounted for LOH of heterozygous mutants for the Notch pathway component *Su(H),* on chromosome 2L. In addition, upon sequencing flies heterozygous for *O-fut1*^{4R6}*,* a null allele of a *Notch* pathway component on chromosome 2R, and *Dl*^{Gal4}, a null allele of the *Notch* ligand on 3R, we found evidence that MR can also take place on these chromosome arms. In addition, we found evidence for *another* LOH mechanism, aneuploidy, explained below.

Results

Sequencing evidence for aneuploidy-driven LOH

One out of the four *Dl*^{Gal4} samples sequenced showed no change at the *Dl* locus but, unexpectedly, showed clear LOH on the X chromosome instead, seen as a shift in the VAF plots indicating that parental SNPs of one genotype now become predominant (**Figure 2.8A**). In contrast to the VAF plots shown in **Figure 2.3 above** where there was no change in chromosome copy number; however, here we found a copy number loss across the entire X chromosome suggesting the loss of X chromosome (**Figure 2.8B**). This is known as “monosomy of the X” and is a form of aneuploidy. We reasoned that, for *Notch* inactivation to occur in females flies with 2 copies of *Notch*, perhaps the loss of the X chromosome was accompanied by the mutation of the remaining *Notch* gene present on the remaining X chromosome.

Consistent with this, additional evidence supporting inactivation of *Notch* due to a deletion was found (data not shown).

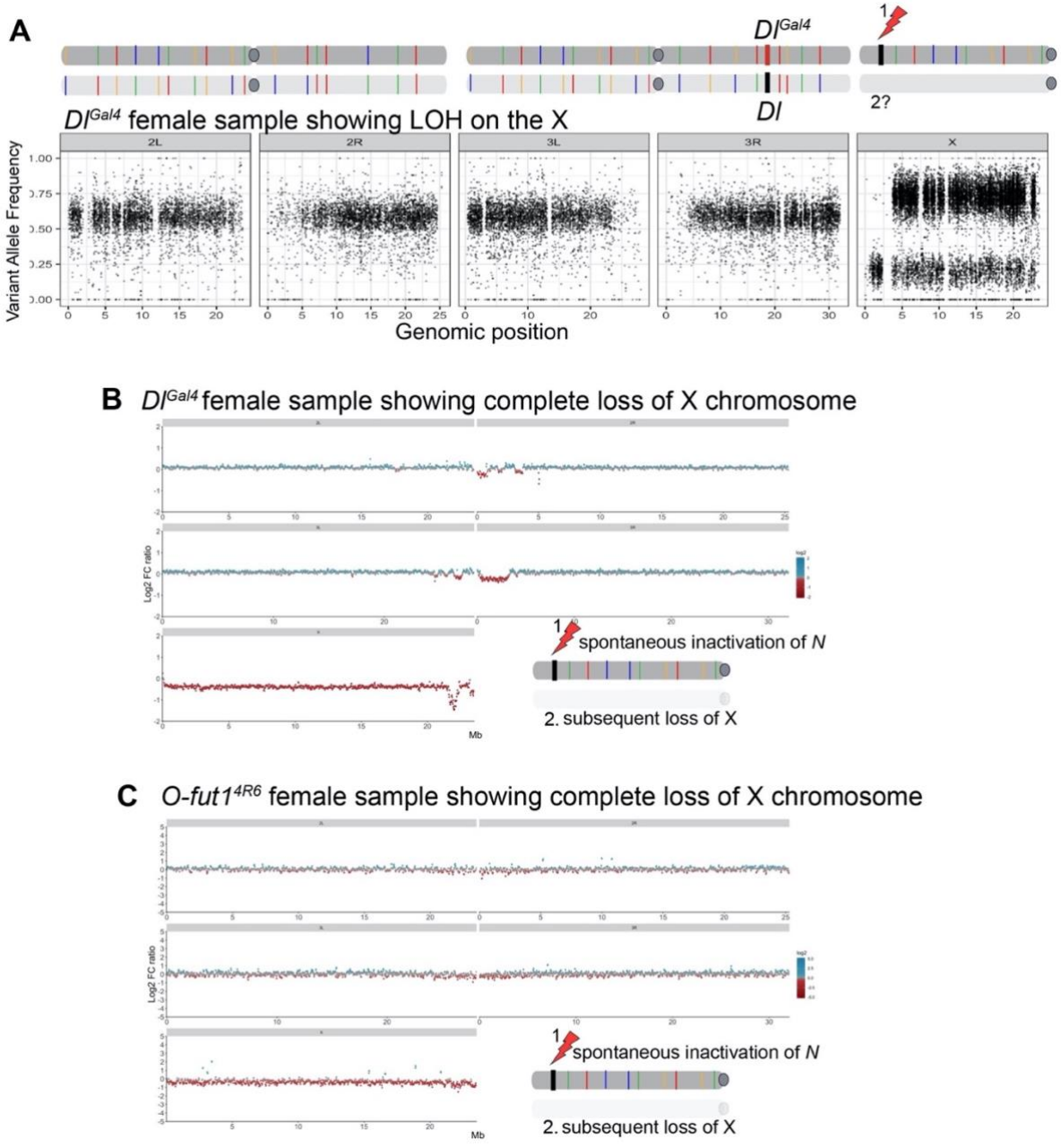


Figure 2.8: VAF plot showing no LOH on chr3R where *Dl* is located, but rather LOH on the X, with SNPs becoming homozygous. A potential explanation for the neoplastic phenotype was the biallelic inactivation of *Notch*.

(A) VAF plot showing no LOH on chr3R where *Dl* is located, but rather LOH on the X, with SNPs becoming homozygous. A potential explanation for the neoplastic phenotype was the biallelic inactivation of *Notch.*

(B) A copy number plot revealed the complete loss of copy number throughout the X chromosome (last panel). Chromosomes 2L, 2R and 3L and 3R are also shown with no whole-chromosomal change in copy number. Of note, the drop in copy number near the centromeres of chromosomes 2 and 3 are

attributed to EC contamination (explained in Chapter 2.1 figure 2.S3). This sequencing evidence shows the complete loss of the X chromosome. A schematic showing the two putative Notch inactivating events is shown.
(C) The complete loss of the X chromosome was also detected in an *O-fut1* sample. A schematic showing the two putative Notch inactivating events is shown.

Monosomy of the X was also detected in an *O-fut1*^{4R6} sample **(Figure 2.8C)**. Again, we suspected that biallelic inactivation of *Notch* on the X also took place, though we could not definitely determine the other putative *Notch*-inactivating event, likely due to low sequencing coverage. While it was surprising that in this instance, the neoplasia did not arise from LOH of the heterozygous mutant alleles that we were initially assaying (*Dl* and *O-fut1*), it nevertheless provided sequencing-based evidence for aneuploidy-driven LOH, a new mechanism of LOH in ISCs.

It seemed quite striking that a two-step process of biallelic inactivation was detected. Previously, no neoplasia were detected in wild-type females (n=519) (Siudeja et al. 2015) and (n=178) (Al zouabi et al. *in preparation*), which would argue against biallelic inactivation of Notch component. However this did not rule rare occurrences of biallelic inactivation. Indeed, we have recently reported the likely biallelic inactivation of a *Notch* pathway component *kuzbanian* (Riddiford 2020), suggesting that these events are possible, albeit rare. It could be, by chance, that we were able to detect a biallelic inactivation in the wild-type females and the other samples of *Dl* and *O-fut1* heterozygous flies. It is also possible that MR-driven LOH is selected against in the *Dl*^{Gal4}/+ and *O-fut1*^{4R6}/+ mutant backgrounds because it confers a reduced fitness, perhaps due to recessive cell lethal mutations on the chromosome arms that would lead to cell death upon MR. Further sequencing of this genetic background and further examination of wild-type female flies would allow us to address this. Nevertheless, the sequencing data indicated that aneuploidy could arise in the form of loss of the X chromosome. Therefore, we

wanted to further clarify the frequency with which aneuploidy and chromosome loss occurred.

The H4K16ac histone mark is a good readout for activation of dosage compensation in the *Drosophila* intestine

Thus far, our sequencing data only provided evidence for aneuploidy of the X chromosome. We reasoned that aneuploidy of the X chromosome might be tolerated because the gene dosage imbalance could be buffered by the “dosage compensation” pathway. In *Drosophila,* the dosage compensation pathway acts in males that have a solitary copy of the X. In order to ensure that males with only one X have equal gene products as females with two Xs, the transcription of genes on the single male X is increased, and ultimately the differences in the doses of X-linked genes between males and females are compensated, reviewed in (Lucchesi and Kuroda 2015).

In our previous study, *Notch* heterozygous mutant flies had numerous LOH events - (81% at 5 weeks of age, most of which were multiple events) (Siudeja et al. 2015). Since we detected loss of the X chromosome by sequencing, we postulated that a fraction of the LOH tumours arising in *Notch* heterozygous N^{55E11}/+ female flies might actually be attributed to loss of the wild-type X chromosome. Therefore, we next wanted to take advantage of activation of the dosage compensation pathway as a read-out to measure the frequency of aneuploidy events. The dosage compensation pathway was previously shown to be induced in genetic conditions that drive aneuploidy (Clemente-Ruiz et al. 2016), suggesting that *de novo* activation of the dosage compensation would occur in the context of X chromosome loss.

Dosage compensation is regulated by Male-sex-lethal (Msl2) that is specifically expressed in males but not in females (**Figure 2.9A, B**). Principally, Msl2 functions to upregulate transcription from the solitary X chromosome (Kelly 1995, Kelly 1997). It does so only in the males because the presence of a single X leads to the production of the inactive *Sex-Lethal-Male* (*sxl-M*) RNA transcript, which leaves Msl2 unspliced and active. In contrast, in females, the presence of 2 X chromosomes enables the production of Sex-Lethal-Female (*sxl-F*), which represses Msl2 through splicing. In males, the active Msl2 upregulates transcription by coating the X with a histone mark: acetylated Histone H4 Lysine 16 (H4K16ac) that modifies the chromatin structure to promote gene transcription (Gelbart et al. 2009). Thus, we hypothesized that the dosage compensation pathway is activated following loss of an X, which could be assessed by looking at Msl2 expression or H4K16ac deposition.

We aged N^{55E11}/+ females and dissected them at different ages (3-6 weeks of age). Initially, we wanted to look for LOH neoplasia that may be positive for Msl2. In wing imaginal discs from male larvae, anti-Msl2 antibody shows uniform specific positive subnuclear staining while it shows no signal in female wing discs (data not shown). However, in the gut, although we detected Msl2 staining in males, the staining seemed regional and not present in all the cells. It appeared to us that the Msl2 antibody was binding nonspecifically to the muscle layer in the gut, thus compromising our readout as we could be missing Msl2 signal in many cells (data not shown). We thus concluded that the Msl2 antibody works well in discs but is not suitable in the gut. Instead, we used antibody staining recognizing the H4K16ac, which gave a good signal in the gut (**Figure 2.9A-A", 2.10.A-A"'**).

In males, a subnuclear region coated by H4K16ac, presumably corresponding to the X chromosome, was detected in all the cells. As expected, female cells were largely devoid of this subnuclear signal (**Figure 2.9B, B”**), owing to the fact that Msl2 is not present in these cells to deposit H4K16ac. If a nucleus positive for H4K16ac is detected in a female, it likely corresponds to activation of dosage compensation due to the spontaneous loss of the X, thus providing a good readout for LOH via monosomy of the X.

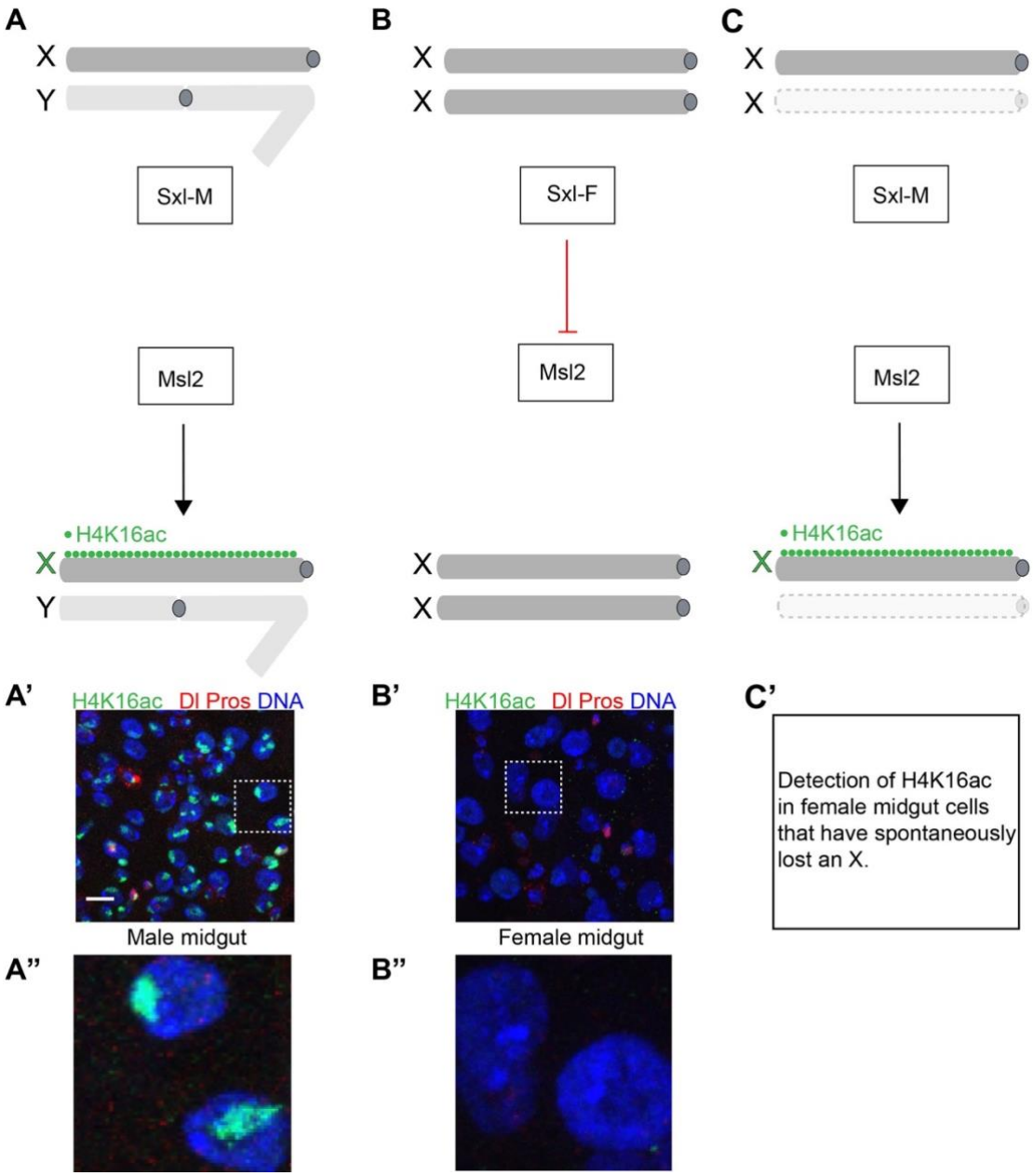


Figure 2.9: The activation of the dosage compensation pathway in *Drosophila* males can be visualised by staining for the H4K16ac mark

(A) The presence of one X chromosome in males leads to the production of Sex Lethal-Male (Sxl-M), that is unable to repress Msl2, which subsequently deposits H4K16ac along the X chromosome which promotes upregulated transcription of X-linked genes in males. (A'-A") Staining for H4K16ac reveals a distinct subnuclear region the nuclei of males, corresponding to the X chromosome that is covered by the histone mark. Zoom of outlined region in (A")

(B) The presence of two chromosomes in females leads to the production of SexLethal-Female (SxlF), that represses Msl2. Since Msl2 is repressed, there is no deposition of H4K16ac in females.

(B'-B'') Nuclei of females with 2 Xs showed no subnuclear signal for the H4K16ac mark. Zoom of outlined region shown in B".
(C) Loss of an X in a female (aneuploidy) would lead to the production of Sxl-M, activating the dosage compensation pathway via Msl2 covering the remaining X chromosome with H4K16ac.
(C'') We will use this assay to detect aneuploidies in female midguts.

Loss of X (aneuploidy) is detected in aging *N^{55E11}*/+ females

In order to determine whether LOH arose due to loss of an X chromosome, we dissected *N^{55E11}*/+ females at 3 weeks and staining for Pros, Dl and H4K16ac. Our data revealed LOH neoplasia positive for the H4K16ac mark with variable sizes and signal distribution (**Figure 2.10A-D'''**). We also detected some LOH neoplasia with the histone mark present in a portion of the neoplastic cells (**Figure 2.10D-D'''**). This could be attributed to either a subclonal aneuploid event that arose following a first LOH event that triggered neoplasia formation, or it could be attributed to clone fusion. The majority of LOH neoplasia, however, were negative for H4K16ac. At 3 weeks, out of 205 LOH neoplasia observed, 28 were positive for H4K16ac (13.7%) (**Figure 2.10E**), strongly suggesting that these are LOH neoplasia that arose by aneuploidy (**Figure 2.10G**). The remaining LOH neoplasia likely arose by MR, the mechanism covered in chapter 2.1 (**Figure 2.10F**). An increase in LOH neoplasia occurred with age, consistent with our previous findings (Siudeja et al. 2015) and (**Figure 2.1**). The proportion of neoplasia positive for the histone mark significantly increased between 3 and 4 weeks, but plateaued at around 25% (data not shown). Thus, our results show another novel mechanism for LOH: spontaneous aneuploidy. This represents another significant process leading to cancer initiation.

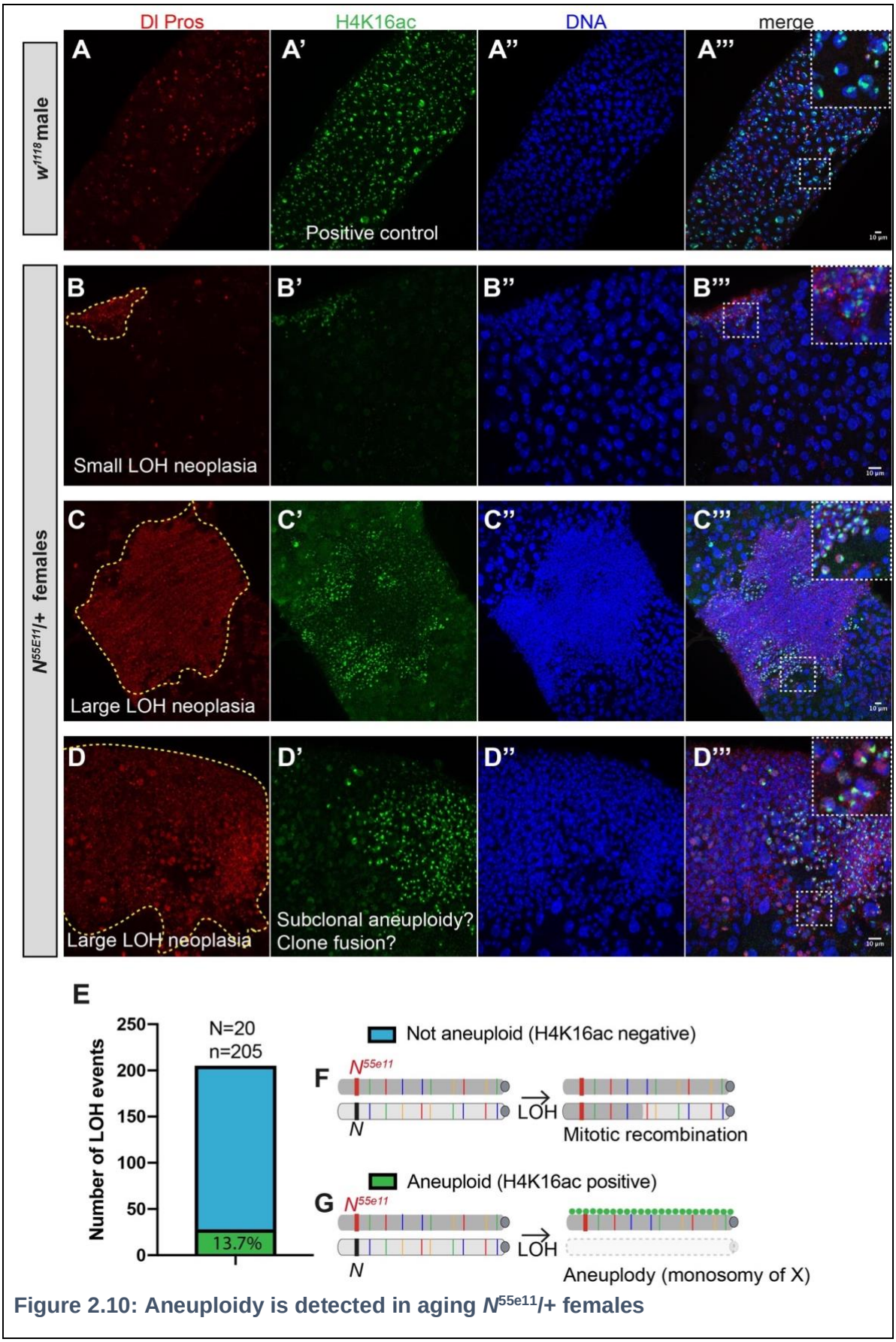


Figure 2.10: Aneuploidy is detected in aging N^{55e11}/+ females

(A-A’’’) A wild-type male midgut used as a positive control, showing a positive subnuclear signal of the H4K16ac mark in all cells. If this mark is detected in females, it denotes an aneuploidy of the X.

(B) A 3 week old N^{55e11}/+ female gut showing a small Notch loss of function neoplasia as shown by the accumulation of ISCs and EEs via Pros and Dl staining.

(B’) This LOH neoplasia is positive for the H4K16ac mark, suggesting that the LOH arose by aneuploidy.

(B’’) DAPI staining showing the LOH neoplasia as a cluster of diploid cells tightly packed together.
(B’’’) Merge of all channels.

(C-C’’’) A 3 week old N^{55e11}/+ female gut showing a large LOH neoplasia.

(D) This LOH neoplasia does not have the H4K16ac present in all the neoplastic cells. This could be due to the aneuploidy being a subclonal event, or it could be a result of clone fusion between a non-aneuploid event and an aneuploid event.

(E) Quantification of the proportion of aneuploid LOH neoplasia detected among the non-aneuploid LOH neoplasia detected.

(F and G) The likely mechanisms of LOH represented by the chromosomes.

Discussion

We have established a system to better understand spontaneous aneuploidies giving rise to LOH in ISCs. This will allow us to assess the potential buffering mechanisms that could equip the cell with tolerance for an abnormal cell complement. A question that remains however is regarding how much of the X is required to be lost for the H4K16ac mark to be deposited. Will the mark be deposited in cases of segmental aneuploidies whereby parts of the X are lost as opposed to the whole chromosome?

Moreover, while we only have evidence for aneuploidy of the X, we do not have any evidence for aneuploidy on autosomes, thus, I will discuss some future avenues to embark on in order to better understand both sex chromosome and autosomal aneuploidies.

Buffering mechanisms of sex chromosome aneuploidies

Abnormalities of sex chromosomes are not unusual, for example Klinefelter's Syndrome is a common disorder of sex chromosome aneuploidy in humans and is amongst other sex chromosomal aneuploidies, reviewed in (Visootsak and Graham 2006). This is because it can be tolerated as a consequence of "dosage compensation" coming into play (Raznahan et al. 2018). The *necessity* of dosage compensation buffering somatic aneuploidy happening in the stem cell in our system is not clear, and questions are yet to be resolved such as: is it possible for aneuploid cells to survive without the upregulation of the X? This can be tested by knocking down components of the dosage compensation pathway in cells losing an X chromosome. Annabelle Suisse, a postdoc in the lab will be addressing this.

Buffering mechanisms of autosomal aneuploidies

In contrast to sex chromosome aneuploidies, aneuploidies on autosomes on are more elusive with regards to buffering mechanisms. In *Drosophila,* there is evidence for a buffering mechanism on the fourth chromosome that is mediated by the *Painting of fourth* gene which increases transcription of the fourth chromosome upon the loss of one copy (Stenberg et al. 2009). Thus, it is possible that ISCs in the gut have a buffering system that allows them to maintain their fitness upon loss of an autosome. This will be a future direction in the lab to be tested through inducing loss of an autosome in heterozygous mutants of *Notch* pathway components on chromosomes 2 or 3.

Fitness of aneuploid cells

When do the spontaneous aneuploidies arise and do they have fitness defects? This can be further addressed through the quantification of the size LOH clones positive and negative for the H4K16ac mark in a time-course experiment for N^{55E11}/+ females. Comparing aneuploid-LOH clone size with non-aneuoploid LOH clone size, (presumably LOH driven by MR), will reveal if there are any differences in ISC division kinetics driven by both mechanisms.

Mechanisms of aneuploidy in aging ISCs

Finally, since we detected an increase of aneuploidy-driven LOH clones with age, we whether aberrant mitoses are more frequent in aged individuals. Staining for α-tubulin and other cell markers, we can identify cells in anaphase, and detect mis-segregated and lagging chromosomes in ISCs at different ages to see which mitotic defects (described in 1.2) become more apparent with age.

Overall, I have contributed towards establishing this new project in the lab that will be further explored by Annabelle Suisse.

Chapter 3
Discussion

Chapter 3: Discussion

Discussion Overview

In this chapter, I will begin with a critique of my experiments, highlighting some of the technical and experimental caveats whilst offering suggestions for improvement. I believe it is important to have these caveats in mind first in order to assess the results.

I will then discuss my results in light of previous findings to place them in context of what is known. I will finally end on providing some perspectives of where further research on LOH in stem cells could lead to and the implications it could have in the clinic.

3.1 Technical evaluation of the work and experimental caveats

Sample size of sequenced tumours

The sequencing of tumours detected in the *Drosophila* intestine established a novel SNP mapping strategy to interrogate underlying mechanisms of somatic LOH. From the 16 females and 6 males sequenced, we were able to gain information about how much coverage and tumour purity is required to confidently map recombination regions leading to LOH. We have mapped 11 recombination regions, which has provided further insight into whether the MR occurred via BIR or CO. The 11 recombination regions also pointed towards putative genomic features that could drive MR. Indeed, increasing the sample size from 11 breakpoint regions to 30 would make our analysis of underlying genome features more robust and overcome issues related to low statistical power and inflated false discovery rate. By increasing our sample size, we will be able to see if our findings about the histone locus cluster and R-loops hold up, and it will also allow for the discovery of new genomic features. It is important to recognise, however, that this work has laid significant groundwork for future sequencing to uncover genomic features underlying CO sites.

Tumour purity

The sequencing also allowed us to inquire how much read depth and tumour purity is needed for the mapping analysis. We realised through the course of mapping the breakpoint regions that contaminating ECs can impact the analysis. We thus deduced that tumour purity was a more important factor than read depth

as contaminating non-tumour SNPs can significantly compromise the confidence in mapping regions. We thus calculated EC contamination (see methods) and determined a threshold of an EC contamination value of 0.75 (see **supplementary Table 1**). Any value below 0.75 was very difficult to map using IGV as we were not able to confidently say if SNPs were heterozygous or homozygous in the tumour. For instance, in a contaminated sample, where LOH is expected, it is heterozygous instead as the contamination (likely from ECs) contributes to the reads. We will use these calculations and thresholds in future sequencing.

Controversy surrounding R-loops

The association of R-loops with the mapped recombination regions has been a point of controversial discussion over the past few years. In addition to the fact that our small sample size stops us from making *strong* suggestions of an association, consensual mechanistic and molecular uncertainties persist in the R-loop field. I will first provide some background on R-loops then state some of the issues concerning them.

R-loops form when the nascent complementary RNA binds to the template DNA strand, forming a DNA:RNA hybrid, leaving the template strand single-stranded. This forms a 3-stranded nucleic acid structure and usually forms during transcription (Thomas et al. 1976). An R-loop is different from a regular transcription bubble in that it is longer (has longer regions of DNA:RNA hybridisation than what is formed during transcription) and is more stable as a structure. What facilitates the stability and length of the hybridisation are a number of different factors. Firstly, C-rich DNA templates binding to the nascent RNA that is G-rich, leads to thermodynamically stable bonds that can be further stabilised by the stretch of C-

rich sequence (Roy and Lieber 2009; Roy et al. 2010). Another way in which stability can be established is by negative supercoiling generated behind the transcription bubble, reviewed in (Chedin et al. 2020; Kuzminov 2018). This is thought to promote R-loop formation by making this part of the DNA more underwound and thus more favourable to the stable annealing of the nascent transcript. It has been shown that high levels of transcription can lead to this negative underwound state known as “negative supercoiling”. The 3-stranded stable nature of R-loops make them a secondary structure in the DNA and there is growing evidence that these R-loops can cause topological stress for the incoming replication fork, nicely reviewed in (Rondón and Aguilera 2019). Thus, R-loops have attracted increasing interest recently as potent sources of DSBs, and so mapping and quantification methods of R-loops have been on the rise.

There are different ways to map and quantify R-loops, the most common way relies on the use of the S9.6 monoclonal antibody. This antibody however has been shown to also have an affinity to bind to dsRNA in addition to DNA:RNA hybrids (Hartono et al. 2018). Thus, important controls are required when studying R loops. Such controls involve using endoribonucleases (such as RNaseIII), to specifically cleave and remove dsRNA and are important for the acquisition of DNA:RNA specific signals. Additional controls include verifying the sensitivity of samples to the endoribonuclease that specifically hydrolises the phosphodiester bonds of RNA bound to DNA, called RNaseH1 (Smolka et al. 2020). Furthermore, another important aspect of R-loops is that they are dynamic in nature, mostly forming during replication, meaning that their detection is cell-cycle dependent (Hamperl et al. 2017). Thus, the specificity of the antibody combined with the dynamic nature of R-loops adds difficulty in the field to probe into their formation, presence and impact

on the genome. For a nice review on difficulties mapping R-loops, see (Vanoosthuyse 2018). Here, I highlight some of the things I have faced during my PhD regarding studying R-loops:

1.Discordant datasets of mapped R-loops in Drosophila

In *Drosophila,* there are only two available datasets of R-loops that have been mapped, both from S2 cell lines using a ChIP-seq method (Alecki et al. 2020; Bayona-Feliu et al. 2017). This ChIP-seq method relies on the S9.6 antibody that recognises DNA:RNA hybrid, and is known as DNA:RNA immunoprecipitation (DRIP-seq). We acquired DRIP-seq peaks from both datasets and found a discordance between the peaks. Further examination into this discordance revealed that one study carried out important controls after sonication, degrading dsRNA and checking for RNAseH1 sensitivity (Alecki et al. 2020), while the other did not (Bayona-Feliu et al. 2017). For our analysis, we used the dataset that applied more stringent controls but would like to highlight that the apparent lack of reproducibility between studies mapping R-loops is present. It is unclear why differences in the datasets exist and we cannot rule out that an R-loop mapped in one study and not in the other is false-positive in one study, or a false-negative in the other study. This suggests that further investigation could be fruitful.

2. The possibility of tissue-specific differences giving rise to different R-loops

There are no *Drosophila* gut-specific or ISC-specific data of mapped R-loops. Accordingly, we had to keep in mind that tissue-specific differences may exist between S2 cells and ISCs. To account for this, we assessed if the R-loops overlapping our recombination regions were in genes expressed in the ISC. 3 out 5 of the R-loops were in genes moderately/highly expressed in the ISC

(http://flygutseq.buchonlab.com), 1 was in a gene that was lowly expressed and 1 was intergenic. Despite the fact that most R-loops form over gene-bodies, there have been accounts mapping intergenic R-loops (Toriumi et al. 2013; Al-Hadid and Yang 2016). Still, we cannot be certain that tissue differences do not exist, especially because the transcriptome differs between S2 cells and ISCs. Altogether, when expanding the dataset, we should keep the possibility of tissue-specific differences giving rise to different R-loops in mind. A dataset mapping R-loops in the ISCs would be extremely helpful.

3. *Difficulty modulating R-loops in vivo*

Detecting an association between our mapped recombination regions and R-loops prompted us to want to increase/ decrease R-loops in vivo to see the impact of that on our LOH readout. Thus, we initially planned on taking a chemical route using drugs as well as a genetic route. Since there is an association with R-loops forming in regions of negatively supercoiled DNA, we wanted to use drugs that will increase the negative supercoiling and hence increase R-loops. Topoisomerase (Top1) is an enzyme that comes into play to relieve supercoiling and torsion. Using a Top1 inhibitor, such as camptothecin (CPT), would thus leave more supercoiled DNA and increase R-loops (Pommier 2006). This has been a way to increase R-loops in the literature (Chappidi et al. 2020). However, camptothecin also inhibits Topo1 activity that is unrelated to R-loops leading to DSBs that are R-loop independent. Thus, using camptothecin, on one hand, would increase R-loops but on the other hand, could be causing additional DSBs in the genome driving LOH. There would be difficulty in teasing apart the impact of increasing R-loops from the secondary impact of camptothecin.

Another way of modulating R-loop levels is by genetically increasing/decreasing the expression of the enzyme that resolves R-loops. I mentioned above that RNaseH is the enzyme specific for RNA:DNA hybrid dissolution in the nucleus (Stein and Hausen 1969; Amon et al. 2016; Cerritelli and Crouch 2009). However, in the mitochondria, RNaseH1 has other roles in mtDNA replication. We hypothesised that knocking-down *RNaseH1* in the ISCs would lead to secondary mitochondrial defects which could impact the cell and hence LOH readout. We thus resorted to testing the overexpression of *RNaseH1* in the ISCs, which would lead to a decrease in R-loops and less of an impact on functions in the mitochondria. Additional methods for altering R-loops *in vivo* would facilitate the study of their impact on cells in tissues.

4. *Visualising R-loops using immunofluorescence: a missing control for validating R-loop decrease*

Overexpressing *RNaseH1* decreases R-loops *in vitro*. Upon using a *UAS-RNaseH1* combined with the *Su(H)*$^{\Delta 47}$/+ genetic background, the *UAS-RNaseH1* was expressed using the temperature sensitive driver of ISCs Dl^{TS}. To validate whether the *RNaseH1* overexpression in ISCs indeed reduced R-loops, we would have to rely on the use of the S9.6 antibody to detect R-loops using immunofluorescence. In one experiment where I used the S9.6 antibody, a lot of signal was detected inside and outside the nucleus. Increasing accounts in the literature have been reporting that the nuclear signal is also not specific (see above). Thus, more than one control is needed to confirm levels of R-loops by immunofluorescence staining. Recently, a study used a GFP-tagged and catalytically inactive form of RNAseH1, shown to localize to R-loops (Chappidi et al.

2020). This would indicate the presence of R-loops and could be further confirmed with staining with the S9.6 antibody.

Despite the aforementioned difficulties, many teams are working on improving R-loop mapping protocols and refining methods of R-loop detection. For instance, MapR is a novel method that maps R-loops without using the S9.6 antibody and relies on using a catalytically dead version of RNAseH1 instead (Yan and Sarma 2020). Another recent improvement in the field is a new step preceding the immunoprecipitation step in DRIP-seq. This new step relies on converting cytosines to uracil in ssDNA. Since one of the strands in the R-loop in ssDNA, the R-loop can be precisely mapped because of the exclusive presence of uracils, and rules out the chance of detecting dsRNA as it will not have uracils (Malig et al. 2020). It is promising that our understanding of R-loops will improve in the coming years with these refined methods.

Additional biological repeats and RNAi lines

Due to the global health crisis that COVID19 has put us in, there was a delay with many of my experiments. Limited access to the lab during lockdown meant stopping many ongoing long-term experiments. I have repeated all experiments with at least 2 biological repeats, though my ability to carry out third biological repeats was limited for the following experiments: *Pif1* knockdown, *Mus81* knockdown, UAS-*RNaseH1* overexpression, UAS-*CycE,Stg* overexpression and UAS-*DAP* overexpression. Thus, all the data shown from the above experiments are data from 2 independent repeats. In addition, my ability to test more than one RNAi line for *Rad51*, *Pif1* and *Mus81* knockdowns was also limited. It is important to test multiple independent RNAi for different genes to rule out off target effects. I will test additional

RNAi for these genes or use genetic null contexts in addition to other candidate DNA repair genes in the near future prior to publication of my PhD work.

3.2 Discussion of results

Exploiting the clonal nature of LOH neoplasia in the *Drosophila* intestine: the novelty

My PhD took advantage of the LOH occurring in stem cells that have clonally expanded. This clonal nature of ISC-derived LOH is reminiscent of spontaneous tumours that arise as a result of LOH of tumour suppressor genes in humans. I interrogated the underlying mechanisms of these clones and how certain genetic and environmental factors can modulate their frequency. Using whole-genome sequencing approaches, I was able to find evidence for LOH driven by MR, and more indirectly, evidence for aneuploidy-driven LOH, whereby the loss of the entire X chromosome was detected. Thus my PhD, for the first time, established a novel system to study LOH in an *in vivo* stem cell model that gives rise to spontaneous tumours.

Indeed, the novelty does not lie within using *Drosophila* to gain insight into MR, as MR was first discovered in *Drosophila.* Classic studies from Curt Stern observed adjacent clones with bristles that were mutant for a marker (either *y* or *singed (sn)*). He reasoned that the two adjacent clones resulted from “somatic crossover and segregation”, populating two “twin-spots”, that originate from the same mitotic recombination event (Stern 1936a). Taking advantage of Stern’s findings, the development of powerful tools to perform routine genetic mosaic analysis was established, using induced mosaic techniques such as FLP/FRT and

MARCM (Brand and Perrimon 1993; Lee and Luo 2001). Therefore, the novelty of our system however, lies in our ability to apply modern genomics to understanding spontaneous MR events. Indeed, many questions remain unresolved regarding what causes uninduced spontaneous LOH that drives mosaicism in human tissues alluded to in 1.1. Of note, there have been studies mapping spontaneous MR sites in the male germline in *Drosophila*: Since meiotic recombination does not happen in male flies, the authors attributed any arising recombination in offspring to mitotic processes. In one study, the authors increased the rate of MR by using flies mutated for the *blm* helicase, whose mutation leads to elevated levels of HR. Our study is unique in the sense that it is examining MR in adult stem cells in a wild-type animals background with endogenous levels of DNA repair, which more suitably resembles tissue mosaicism and clonal expansions in human contexts.

There are less cell divisions in the male germline compared to that of the gut (Boyle et al. 2007). The male germline is maintained by 6-9 germline stem cells (GSCs) per testis (Matunis et al. 2012). In contrast, the *Drosophila* intestine is populated by around 1000 stem cells, that regenerate the gut every 1-2 weeks. Just based on these numbers, it seems likely that MR will be much more frequent per gut than per testes, making the gut an easier model in which to assess spontaneous MR. Consistent with this notion, in wild-type backgrounds, the rate of spontaneous LOH in the male germline was found to be extremely low though slightly increased in *blm* mutant flies to 2% for markers on opposite chromosome arms (LaFave et al. 2014). In contrast, we found that MR in the intestine occurs frequently, with more than 1 per gut for markers near the telomere. This high rate is due to the number of stem cells in the tissue, their proliferation rate, and the long aging time-span during which our assay was performed. With this, a number of insightful findings came

about from our study regarding both mechanism and putative intrinsic genomic drivers of MR as well as environmental extrinsic drivers.

Drosophila genomes are at least twice as polymorphic than human genomes (Langley et al. 2012; Wang et al. 2015). We were thus able to take advantage of the higher density of SNPs in *Drosophila* to map breakpoint regions down to a resolution of 47bp in one sample, overcoming some of the technical limits from human data. Of note, the 3 samples that had lowly resolved breakpoint regions 18kb-29kb, still had a better resolution that what could be achieved in human cells. To my knowledge, the most resolved region of recombination mapped was 2Mb from (Howarth et al. 2009). The reason why we had a sample with a region of recombination resolution of 29kb is because the recombination happened within a long stretch of non-informative SNPs composed of a region of homozygosity. Regions of homozygosity are present in all organisms, representing an ancestral haplotype inherited from both parents. By choosing parental genotypes with increased sequence divergence, we could get better resolution in these regions. Thus, we established a groundwork for future datasets that will provide more insight into other genomic features.

Mechanisms of LOH

Evidence for LOH driven by MR and aneuploidy

Our sequencing provided evidence that LOH predominantly occurs via MR on chromosome 2L. 12/13 *Su(H)$^{\Delta 47}$*/+ female samples showed clear MR (and one sample was inconclusive due to EC contamination). The sequencing also supported that MR takes place on chromosomes 2R and 3R. Therefore, the major mechanism driving LOH in our system is MR. We also unexpectedly detected LOH via

aneuploidy in 2 samples likely via an initial somatic SV in *Notch* and a second inactivation via aneuploidy- (monosomy) of the X. Our finding thus provided an additional mechanism of LOH in ISCs in which we can test. Since my PhD work mostly focused on MR, I was able to examine more closely the mechanisms in MR. Mechanisms of aneuploidy are being further studied by Annabelle Suisse, a postdoc in the lab, with whom I collaborated. A small discussion on aneuploidy is in chapter 2.2.

No evidence for LOH driven by SNVs or amitosis

Our data argues against frequent LOH through point mutation or amitosis. This suggests that an inactivation point mutation is far less common than a recombination event. We also did not detect any evidence for amitosis under our conditions. Amitosis is a mechanism reported by Luchetta and colleagues, whereby a cell that has started to differentiated and is 4n, would undergo a process that would lead to two parental chromosomes randomly segregating into the same cell, leading to LOH. This is a spindle-independent reduction of ploidy in cells that the authors claim happens in response to ISC dysfunction during periods of proliferative demand. In their study, the authors deplete ISCs following starvation and observe amitosis giving rise to new functional yet potentially problematic ISCs upon refeeding, since these ISCs can undergo LOH (Lucchetta and Ohlstein 2018). It might be of interest replicating the starvation conditions of the authors' experiments to see if LOH driven by amitosis would be detected by sequencing, or whether chromosome loss could explain their findings.

Detection of MR in males provides an opportunity compare the impact of sexual dimorphism on LOH

In males, we showed that MR takes place in 1 out of 6 samples. The detection of LOH events in males is confounded by SV inactivation of *Notch* using our *Su(H)/+;ProsV1Gal4; UAS-nlsGFP* assay to visually identify midguts containing neoplasias. For this reason, the remaining 5 samples sequenced showed SVs in *Notch,* located on the X and therefore hemizygous in males.

The fact that we see one male showing MR is of interest. In *Drosophila,* males and females have highly sexually dimorphic intestines, with the female intestine being larger and more proliferative (Reiff et al. 2015; Hudry et al. 2016). This proliferative nature of female intestines makes females undergo age-related hyperplasia, a decrease in intestinal barrier function and an increase in immune activation (Rera et al. 2012; Regan et al. 2016). Thus, males and females age differently. Knowing that MR also happens in males provides us with an opportunity to compare the same LOH events in sexually dimorphic organs that age differently. By using the method described in Figure S8, describing the GFP negative clones corresponding to LOH on the second chromosome specifically, we can compare MR-driven LOH frequency between males and females, which we expect might differ in frequency.

DSBR pathways that result in LOH

As for delineating the underlying mechanisms of MR in our system, we show that the stem-cell specific knockdown of the HR protein *Rad51* significantly reduced LOH events. Rad51 initiates the invasion into the homologous chromosome or the sister for the exchange of DNA material. The 10% guts detected with at least 1 LOH

clone after *Rad51* knockdown, could attributed to residual *Rad51* activity from the RNAi. Further testing by using a Rad51 antibody to check for any residual protein in addition to seeing the impact of using a *Rad51* null mutant in heterozygous and homozygous contexts is a possible follow-up. The prediction is that the homozygous mutant will bring LOH events to zero.

Whether the reduction of LOH upon *Rad51* knockdown is due to cell death because of failed DNA repair, or another DSBR pathway coming into effect, is of great significance. Previous work has demonstrated that different repair mechanisms compete for the same DSB substrates, where competition and choice of DSBR pathway mostly relies on cell cycle, availability of homologous repair templates, protein availability at the DSB site, phosphorylation of repair proteins and chromatin modulation of repair factor accessibility (Shrivastav et al. 2008). Essentially, when the choice is made, the other DSBR pathways are repressed (Mateos-Gomez et al. 2015). Could a suppression in HR activity lead to an increase of the other DSBR pathways such as NHEJ or alt-EJ that will not give rise to detrimental LOH? I discuss the implications of HR suppression in potential drug targeting in 3.4. Intriguingly, studies in male *Drosophila* germ cells have shown an age-dependent shift in DSBR pathway. Preston et al have shown that with age, there is a switch from NHEJ to HR when a DSB was induced (Preston et al. 2006). Such a shift led to an increase in LOH with age, underscoring the importance of DSBR pathway choice and a potential link with aging.

DSBR pathways that do not result in LOH provide promising alternative strategies for repair

Many possible DSBR pathways will give rise to MR-mediated repair of DSBs, but not all will result in LOH. For instance, repair via SDSA and repair via dHJ that result in non-crossovers (NCO), in addition to DNA repair events where repair occurs off of the sister. These would not be detected in our assay as they do not result in large enough changes in the DNA. In the case of repairing off the sister, there will be no change at all to the repaired cell the invading strand is identical to the donor strand. Questions relating to the choice of homologous chromosome instead of the sister remain unresolved, though some insight has been provided by Lee et al. Through microarray analysis of sectored colonies in yeast, Lee et al observed that sometimes two sister chromatids were broken at the same position. One explanation for this is timing of the DSB. They proposed that to explain the coincident breaks on the sisters, the DSB happens in G1. This way, when the broken chromosome is copied during S phase, it results in two sisters with DSBs. Consequently, in such a scenario, the intact homologous chromosome is the better choice for repair, as both sisters need to be repaired independently (Lee and Petes 2010). To what extent this holds up in *Drosophila* is uncertain and difficult to test. An idea of the timing of DNA damage may be illuminated through the use of a Fluorescent Ubiquitination-based Cell Cycle Indicator (FUCCI) tool (Zielke et al. 2014), but not the choice between sister and homologue for the repair. Though it may not be uncommon for repair to happen off the homologue in *Drosophila* as chromosomes are paired during the cell cycle, which may facilitate the use of the homologue during DNA repair (Stack and Brown 1969).

Next, I will discuss the putative genomic drivers we have detected, that may have caused the DSB driving the repair mechanisms discussed above.

Genomic Drivers of MR

The *Histone Locus Cluster* and R-loops were two putative genomic features in this study that could be driving MR. Both encompass genomic aspects that make them prone to DNA damage. 2 recombination regions coincided within the same 20kb within the *Histone Locus Cluster* while recombination can occur anywhere between the centromere and *Su(H)* over a distance of 7.5Mb. The *Histone Locus Cluster* is an array of 5 histone genes repeated 100 times in *Drosophila*. The tandem repeat nature of the *Histone Locus Cluster* is reminiscent of the tandem repeat nature of low copy repeats identified in mapping MR sites in human cell lines and Ty retrotransposons in yeast (Howarth et al. 2009; St. Charles and Petes 2013). Repetitive DNA has been shown to be able to form secondary structures causing an impediment to the incoming replication fork during S-phase, reviewed in (Polleys et al. 2017). Moreover, the histone genes in the *Histone Locus Cluster* are expressed exclusively during S phase in order to incorporate histones to the newly synthesised DNA during replication, needed for the packaging. S-phase dependent transcription has the potential to cause transcription-replication conflicts potentially leading to replication fork collapse. In humans, the clustering of the arrays is conserved with evidence of the *HIST1H* cluster in human B-cells acquiring marks of γH2AX (Boulianne et al. 2017). Further investigation of the *Histone Locus Cluster* could provide important insight into understanding whether S-phase dependent transcription of genes increases LOH.

Our finding that recombination regions are enriched for R-loops provides insight into another obstacle resulting from transcription. I have highlighted some important features of R-loops in 3.1. Importantly, R-loops are also secondary structures that pose a threat to the incoming replication fork. Irrespective of our sequencing association, R-loops remain a notorious source of DSB formation and have been shown to induce DNA damage and drive LOH as well as whole chromosome losses in yeast (Amon et al. 2016; Wahba et al. 2016; O'connell et al. 2015). We thus attempted to decrease R-loops *in vivo* in our system, by overexpressing RNaseH1, which is an enzyme specific for RNA:DNA hybrid dissolution (Stein and Hausen 1969; Amon et al. 2016; Cerritelli and Crouch 2009). We found a significant decrease in LOH from 2 experiments when overexpressing RNAseH1, hinting towards a role for R-loop driving MR in our system. Current efforts in the R-loop field will overcome the caveats mentioned in 3.1 and I am hopeful that in the future we will be able to probe into their possible contribution to LOH in our system.

Impact of the stem cell niche and environment on LOH

Each individual fly gut may be impacted by numerous sources of individual variability. We thus propose that the proliferative status of each gut is a possible explanation for the presence of multiple events in some guts, more than that would be predicted by chance. One way in which this can occur is via the cell-autonomous increase in proliferation driven by a tumour in the tissue. The tumour produces mitogenic signals in the niche, stimulating further events (Patel et al. 2015). A second way in which an increase in proliferation can arise in the gut is from pathogenic bacteria, which also causes an increase in mitogenic signals. We show

an increase in LOH after weekly treatments with *Ecc15*, demonstrating an example of how an environmental factor can drive LOH. This resembles the impact of an *E.coli* strain, NC101, which induces cancer initiation in humans (see intro 1.3). *Ecc15* illicits an immune response in host gut cells, particularly ECs, by ROS production and compensatory proliferation. We show both a proliferative response and an increase in DNA damage marks in ISCs, marked by γ-H2Av, but it is difficult to untangle whether the increase in DNA damage marks detected is due to ROS or proliferation or both. *Ecc15* showed more DNA damage than *Pe,* which might explain the increase in LOH upon *Ecc15* but not *Pe* treatment, leading to more LOH. What remains unresolved is the contribution of bacteria to non-replication (S-phase) dependent DSBs. Since ROS damages proteins, lipids and nucleic acids of a cell, could that contribute towards more LOH per mitoses? It would be insightful to see the impact of more environmental damaging agents, such as an additional bacteria that may illicit a different immune response, or paraquat that is known to increase oxidative damage through ROS. Separating downstream effects of these environmental factors could be a future endeavor.

A very interesting avenue of research that future studies in the lab could address, is the impact of non-autonomous tissue on the progression of LOH clonal expansion. Cancer research has classically focused on identifying tumour-autonomous processes and previous studies in the fly intestine have shown that tumour-induced cell competition in the tissue promote neoplastic transformation (Suijkerbuijk et al. 2016). Seeking experimental evidence for the other direction where *normal* tissue-induced cell competition can affect neoplastic growth will be of interest. We hypothesise that the homeostatic properties such as proliferation status and cell turnover can impact the expression of clonal expansion. This can be tested

using precise genetic tools to manipulate tissue properties in a non-autonomous manner, which is very difficult to do in a mammalian model.

LOH with age

In our study, we show that aged heterozygous mutant flies acquire patches of somatically arising mutant clones. These clones are a consequence of stem cells acquiring spontaneous LOH driving positive selection due to increased fitness. With time, clones expand, leading to genetically mosaic tissue and contributing more LOH to the tissue, potentially impacting homeostasis and functional integrity. We show that by aging *Su(H)* /+ flies and dissecting at 1, 3 and 6 weeks, the number of detected spontaneous clones increases dramatically between 3 weeks and 6 weeks of age with an approximate 62% increase in guts harbouring at least 1 LOH clone. This is reminiscent of data in humans showing mutant clonal expansions increasing with age. Clonal hematopoiesis for instance exponentially increases in aged individuals (Zink et al. 2017b). Additionally, the trend of a dramatic increase in clonal events matches the sudden increase in cancer risk during the second half of human life (Depinho 2000; Cancer Registration Statistics, England 2019). What is contributing to this age-related increase?

In chapter 1.3 (introduction) I discussed the age-related changes that occur in the *Drosophila* gut. These include the accumulation of DNA damage, the increase in proliferation, the accumulation of ROS through gut dysbiosis and the gradual decline in gut barrier permeability, all factors that could drive DSBs. However, I would like to emphasise here a point not been previously considered: In addition to driving DSBs could the aforementioned age-related factors *also* influence the *choice* of repair? I mentioned earlier that there is evidence in *Drosophila* that aging male

flies show a switch in DSBR pathway from NHEJ to HR and an increase of detected LOH, at least in the germline (Preston et al. 2006). It is possible, therefore, that the choice of repair mechanism employed by the cell has a larger impact on LOH than the amount of DNA damage. Since the different DSBR pathways NHEJ, HR and alt-EJ all compete for the same DSB substrate, the age-associated oxidation impacting a DNA repair component, could alter the choice of pathway employed (Stadtman 1992) reviewed in (White and Vijg 2016; Reeg and Grune 2015).

3.3 Implications of the research and conclusions

Implications of the research and perspectives

The scientifically exploitable results that have arisen from my PhD project are that we have developed the very first system for systematically studying LOH using a high-resolution technique in a living animal that leads to neoplastic growth. Our *in vivo* stem cell model system of the *Drosophila* intestine will provide an important link in the evolutionary conservation of MR mechanisms between unicellular yeast and complex tissue of higher eukaryotes (in which studying the molecular mechanisms of LOH has proven to be technically difficult). We also identified both cell-intrinsic and extrinsic factors that can drive LOH and thus provide important insight into cancer initiation that can be further explored for therapy.

Individuals who are particularly vulnerable to LOH-driven cancers, such as those who are born with a germline mutation in a tumour suppressor gene, can potentially benefit from therapy promoting an alternative DSBR pathway. Development of a strategy where the choice of repair is skewed towards NHEJ and alt-EJ will reduce the chance of detrimental LOH. Many therapies are being developed targeting DNA repair pathways using specific DNA repair pathway

inhibitors (Kelley et al. 2014; Gavande et al. 2016; Lange et al. 2011; Chernikova et al. 2012). Although unanticipated issues may arise when directly inhibiting pathways, a better understanding of the genetic interaction between pathways (NHEJ, alt-EJ and HR) and their proteins can lead to alternative routes of inhibition. Modulating protein function of certain HR proteins by binding to regions outside of the catalytic site could be a potential way to suppress detrimental HR. A future goal would also be to decrease the likelihood of repairing off of the homologue and promote repair off the sister.

On the other hand, individuals who have an autosomal dominant disease caused by autosomal dominant haploinsufficient genetic variants can potentially benefit from MR-driven LOH through restoration of the wild-type. Somatic genetic rescue by MR has been observed in diseases such as Icthyosis and Blackfan anemia (Jongmans et al. 2018; Venugopal and et al 2017; Choate et al. 2010; Pan et al. 2020). If natural gene therapy is achievable, targeting the HR pathway using drugs may also be achievable to correct the functionality of the mutated copy through recombining the mutant copy out. This would be done in a tissue where the corrected cell can clonally expand and populate most of the tissue, and would be particularly useful in clonal tissues such as the blood. One could further imagine that cell competition strategies involving non-autonomous tissue (mentioned in 3.2) could be put into place in order to provoke the clonal expansion of the corrected cell. A potential tissue where this could be achieved is the skin.

Our research also sheds light into the relationship between pathogenic bacteria and LOH. Bacteria that make up the microbiota can also be important initiators of many cancers and are poorly understood. With advances in microbiome

therapies (MBT) that are currently under development, patients at risk can benefit from counteracting and correcting the negative effects of dysbiosis (Wong 2019).

Conclusions

The main objectives of my thesis were:

I. To gain an understanding of the molecular mechanism of LOH in a complex tissue of a higher eukaryote using an *in vivo* stem cell model system in *Drosophila*.

II. To identify environmental factors that can increase LOH events in stem cells, such as the microbiota.

The main results demonstrated that:

1. Though whole-genome sequencing of LOH tumours caused by somatic LOH events and profiling copy number changes and changes in heterozygosity of single-nucleotide polymorphisms, we established that LOH arises through mitotic recombination.
2. We found *Rad51* to be implicated in mitotic recombination-driven LOH.
3. Fine mapping of recombination sites did not reveal mutational pile-ups that commonly arise with a break-induced replication mechanism and instead showed clear examples of chromosomes resulting from cross-over resulting from double-Holliday junction-based repair.
4. The mapped recombination regions also provided insight into potential genomic sequence features that may promote mitotic recombination,

including an association with the repeated region of the *Histone Locus Cluster* and regions previously mapped to form R loops.

5. Infection with the enteric pathogenic bacteria, *Ecc15*, increased LOH frequency.
6. We also found evidence for aneuploidy-driven LOH, where loss of the X chromosome in females gives rise to LOH tumours. This led us to develop a way to visualise loss of X events through immunofluorescence and differentiate them from mitotic-recombination-driven events.

Materials & Methods

Materials and Methods

Drosophila stocks and aging

The following fly stocks and alleles were used in this study: From the Bloomington stock center:;;P{UAS-GFP.nls}; (BL 4776). From the Vienna *Drosophila* Resource Centre (VDRC): *Rad51(spnA) RNAi;;;* (VDRC 13362), *;;Mus81 RNAi;* (VDRC 33688), *;;Pif1 RNAi;* (VDRC 34533). The following stocks were generous gifts: W^{1118} (M. McVey), *UAS-RNAseH1GFP* (M. Uhlirova), *ProsV1Gal4* (J. de Navascués), *Su(H)*Δ^{47}; (F. Schweizguth), *O-fut1*^{4R6}(K. Irvine), Dl^{Gal4} (S. Hou), N^{55e11}(Couturier et al., 2012) (F. Schweisguth), *UAS-DAP, UAS-CycE,Stg, tubGAL80ts ; Dl-GAL4* (B. Edgar).

For standard aging of experiments in Figure 1 flies were maintained at 25° on a standard medium composition. For aging experiments, flies were crossed in standard vials (10-15 females per vial) and newly eclosed progeny were collected over 3-4 days. Females are 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 ever 1-2 days. Every 7 days flies were transferred without CO_2 anaesthesia to clean cages. Dead flies were scored upon each food change to assess survival rates. Age of flies at dissection: 6 weeks.

ISC-specific expression of Gal4 whose activity is controlled by the temperature sensitive, ubiquitously expressed GAL80

In experiments in Figures 2.5, 2.6, and 2.7 the genetic background *Su(H)* Δ^{47}*/ UAS-GFP*; *Dl*^{Gal4}*/* + was used to drive stem cell specific expression in the ISC, due to the *Dl*^{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° thereafter.

The shift to 29°C induced ISC specific expression of *Rad51RNAi, Mus81 RNAi, Pif1 RNAi* (Figures 2.5 and S4). Flies were dissected after 2 weeks at 29°C. The shift to 29°C induced ISC specific expression of *UAS-RNaseH1 GFP* (Figures 2.6). Flies were dissected after 3 week at 29°C. The shift to 29°C induced ISC specific expression of *UAS-DAP, UAS-CycE* (Figures 2.7). Flies were dissected after 1 week at 29°C.

Bacteria Treatments

Adult *Su(H)*$^{\Delta 47}$/+ 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 OD200 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 and was assayed by phospho-histone 3 staining 25 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 (Bardin et al. 2010). 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 antibodies were used: mouse anti-Delta extra-cellular domain (1/1000; DSHB), Mouse anti-Pros MR1A-c (1/1000; DSHB); chicken anti-GFP (1/2000, Abcam), anti-PH3 rabbit, (1:1000; Millipore), anti-γH2Av (1:1000; Millipore), anti-H4K16Ac (Rabbit, 1:500; MERCK Millipore).

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 to 1.5-mm intervals (512X512 or 1024X1024) using 20X or 40X oil objectives through the whole-mounted posterior midguts. Representative images are shown in all panels)

Quantification

All quantification was carried out blind.

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.

GFP- assay

Experiments in Figures 2.5, 2.6, and 2.7 the genetic background *Su(H)*$^{\Delta 47}$*/ UAS-GFP*; *Dl*^{Gal4}*/* + was used to drive stem cell specific expression in the ISC, due to the *Dl*^{Gal4} driver. However, as *Dl*^{Gal4} is a loss of function allele of *Dl*, also a *Notch* signaling component, two types of neoplastic LOH clone could arise: those where *Su(H)* is inactivated and those where Dl 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 Figures 2.5-7. In contrast, LOH resulting from other events including mitotic recombination of 3R leading to Dl LOH, are GFP+. These events were scored, but not counted in the analysis of Figures 2.5-7. See Figure 2.S8.

Statistical analysis for LOH clones was carried out 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.

YH2Av quantification – image J

Images were acquired on the LSM900 confocal microscope. A maximum Z-projection was generated for all images on Image J. Only nuclear YH2Av intensity in the ISCs was measured.

Su(H) $^{\Delta 47}$*, O-fut1*^{4R6} *and Dl*^{Gal4} sequencing

on the Novaseq sequencer

Su(H)$^{\Delta 47}$*/+ ;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 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 150 bp mode.

Bioinformatics

Reads were aligned to the *Drosophila* genome release 6.12 using bwa mem v0.7.15. For SNV analysis generated for rainfall plots in figures 2.5 and 2.S6, structural variant calling and filtering: see (Riddiford 2020 and

https://github.com/bardin-lab). Regions of LOH were assigned using https://github.com/nriddiford/nf-lohcator

Calculating coverage

Mosdepth was used to calculate genome-wide sequencing coverage. CNVPlotteR was used to generate coverage and copy number plots.

https://github.com/nriddiford/cnvPlotteR.git

Calculating EC contamination

To determine the likely contamination of tumour samples with EC cells, we compared tumour/normal read counts in a genomic region that is clearly underendoreplicated in ECs (3R:0-3000000; see **Figure 2.S3**). Here, normalised read counts were sampled from 100 kb windows, shuffled 100 times, within this locus. Read depth ratios were calculated for each shuffled window by dividing the read count in the tumour sample with that in the normal sample.

Mapping of regions of recombination

First, on IGV, we confirmed the LOH at the *Su(H)* locus in the tumour. *Su(H)*$^{\Delta 47}$ has a 1.883kb deletion at 2L: 15,038,351 – 15,040,233 removing *Su(H)* and CIAPIN1 transcribed sequences. A drop in the sequence coverage is shown in the tumour at this region (**Figure 2.5A**). A complete loss of coverage is not detected in at the *Su(H)*$^{\Delta 47}$ locus in the tumour is because the *Su(H)*$^{\Delta 47}$allele has a rescue construct. We also 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 (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 relative to that for 11 out of the 13 MR 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 thus assigned a value for each mapped region considering both coverage and purity (see **supplementary Table 1**). We were unable to map the region of recombination for 2 samples because of EC contamination. These samples however allowed us to benchmark what is deemed “too impure” for the mapping of regions of recombination. We assigned a purity value (determined by EC contamination) of **0.75** to be our cutoff, anything below that is too contaminated to map region of recombination.

Bibliography

Bibliography

Chapter 1 references

1. Aguilera A, Gómez-González B. 2008. Genome instability: A mechanistic view of its causes and consequences. *Nat Rev Genet* **9**: 204–217.
2. Al zouabi L, Bardin AJ. 2020. Stem Cell DNA Damage and Genome Mutation in the Context of Aging and Cancer Initiation. *Cold Spring Harb Perspect Biol* a036210.
3. Amcheslavsky A, Jiang J, Ip YT. 2009. Tissue Damage-Induced Intestinal Stem Cell Division in Drosophila. *Cell Stem Cell* **4**: 49–61.
4. Amoyel M, Bach EA. 2014. Cell competition: how to eliminate your neighbours. *Development* **141**: 988–1000. http://dev.biologists.org/cgi/doi/10.1242/dev.079129.
5. Arrojo e Drigo R, Lev-Ram V, Tyagi S, Ramachandra R, Deerinck T, Bushong E, Phan S, Orphan V, Lechene C, Ellisman MH, et al. 2019. Age Mosaicism across Multiple Scales in Adult Tissues. *Cell Metab*. https://linkinghub.elsevier.com/retrieve/pii/S1550413119302505.
6. Arthur JC, Perez-Chanona E, Mühlbauer M, Tomkovich S, Uronis JM, Fan T-J, Campbell BJ, Abujamel T, Dogan B, Rogers AB, et al. 2012. *Intestinal Inflammation Targets Cancer-Inducing Activity of the Microbiota*.
7. Ayyaz A, Jasper H. 2013. Intestinal inflammation and stem cell homeostasis in aging Drosophila melanogaster. *Front Cell Infect Microbiol* **3**.
8. Baker DJ, Jin F, Jeganathan KB, van Deursen JM. 2009. Whole Chromosome Instability Caused by Bub1 Insufficiency Drives Tumorigenesis through Tumor Suppressor Gene Loss of Heterozygosity. *Cancer Cell* **16**: 475–486. http://dx.doi.org/10.1016/j.ccr.2009.10.023.
9. Baker DJ, Van Deursen JM. 2010. Chromosome missegregation causes colon cancer by APC loss of heterozygosity. *Cell Cycle* **9**: 1711–1716.
10. Bakker ML, Boukens BJ, Mommersteeg MTM, Brons JF, Wakker V, Moorman AFM, Christoffels VM. 2008. Transcription factor Tbx3 is required for the specification of the atrioventricular conduction system. *Circ Res* **102**: 1340–1349.
11. Baldwin KM, Hakim RS. 1991. Growth and differentiation of the larval midgut epithelium during molting in the moth, Manduca sexta. *Tissue Cell* **23**: 411–422.
12. Baumann O. 2001. Posterior midgut epithelial cells differ in their organization of the membrane skeleton from other Drosophila epithelia. *Exp Cell Res* **270**: 176–187.
13. Baumgartner C, Toifl S, Farlik M, Halbritter F, Scheicher R, Fischer I, Sexl V, Bock C, Baccarini M. 2018a. An ERK-Dependent Feedback Mechanism Prevents Hematopoietic Stem Cell Exhaustion. *Cell Stem Cell* **22**: 879-892.e6.
14. Baumgartner C, Toifl S, Farlik M, Halbritter F, Scheicher R, Fischer I, Sexl V, Bock C, Baccarini M. 2018b. An ERK-Dependent Feedback Mechanism

Prevents Hematopoietic Stem Cell Exhaustion. *Cell Stem Cell*.
15. Beebe K, Lee WC, Micchelli CA. 2010. JAK/STAT signaling coordinates stem cell proliferation and multilineage differentiation in the Drosophila intestinal stem cell lineage. *Dev Biol* **338**: 28–37.
16. Beerman I, Seita J, Inlay MA, Weissman IL, Rossi DJ. 2014. Quiescent hematopoietic stem cells accumulate DNA damage during aging that is repaired upon entry into cell cycle. *Cell Stem Cell* **15**: 37–50. http://dx.doi.org/10.1016/j.stem.2014.04.016.
17. Behjati S, Huch M, Roerink S, Gundem G, van de Wetering M, Robinson B, van Boxtel R, Blokker J, Cuppen E, Campbell P, et al. 2014. Genome sequencing of normal cells reveals developmental lineages and mutational processes. *Nature* **513**: 422–425.
18. Bignell GR, Greenman CD, Davies H, Butler AP, Edkins S, Andrews JM, Buck G, Chen L, Beare D, Latimer C, et al. 2010. Signatures of mutation and selection in the cancer genome. *Nature* **463**: 893–898.
19. Biteau B, Hochmuth CE, Jasper H. 2008. JNK Activity in Somatic Stem Cells Causes Loss of Tissue Homeostasis in the Aging Drosophila Gut. *Cell Stem Cell* **3**: 442–455.
20. Biteau B, Jasper H. 2014. Slit/Robo Signaling Regulates Cell Fate Decisions in the Intestinal Stem Cell Lineage of Drosophila. *Cell Rep*.
21. Bitner K, Shahrestani P, Pardue E, Mueller LD. 2020. Predicting death by the loss of intestinal function. *PLoS One* **15**: 1–11. http://dx.doi.org/10.1371/journal.pone.0230970.
22. Blokzijl F, De Ligt J, Jager M, Sasselli V, Roerink S, Sasaki N, Huch M, Boymans S, Kuijk E, Prins P, et al. 2016. Tissue-specific mutation accumulation in human adult stem cells during life. *Nature* **538**: 260–264.
23. Boehm A, Khalturin K, Erxleben FA, Hemmrich G, Ulrich C, Lopez-quintero JA, Oberg H, Puchert M, Rosenstiel P, Wittlieb J, et al. 2013. FoxO is a critical regulator of stem cell maintenance in immortal Hydra. *Ann Neurosci* **109**: 19702.
24. Bondar T, Medzhitov R. 2010. p53-mediated hematopoietic stem and progenitor cell competition. *Cell Stem Cell* **6**: 309–322.
25. Booth LN, Brunet A. 2016. The Aging Epigenome. *Mol Cell* **62**: 728–744.
26. Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvák Z, Levin HL, Macfarlan TS, et al. 2018. Ten things you should know about transposable elements. *Genome Biol* **19**: 199.
27. Boveri T. 1914. *Henry Harris's new translation of Theodor Boveri's 1914 monograph Zur Frage der Entstehung maligner Tumoren Concerning the Origin of Malignant Tumours*. The Company of Biologists and Cold Spring Harbor Laboratory Press.
28. Bozcuk AN. 1972. DNA synthesis in the absence of somatic cell division associated with ageing in Drosophila subobscura. *Exp Gerontol* **7**: 147–150.
29. Brown S, Pineda CM, Xin T, Boucher J, Suozzi KC, Park S, Matte-Martone C, Gonzalez DG, Rytlewski J, Beronja S, et al. 2017. Correction of aberrant growth preserves tissue homeostasis. *Nature* **548**: 334–337. http://dx.doi.org/10.1038/nature23304.
30. Brumby AM, Richardson HE. 2003. scribble mutants cooperate with oncogenic Ras or Notch to cause neoplastic overgrowth in Drosophila. - Brumby, Richardson - The EMBO jou.pdf. **22**: 5769–5779.

31. Brunet A, Rando TA. 2017. Interaction between epigenetic and metabolism in aging stem cells. *Curr Opin Cell Biol* **45**: 1–7.
32. Buchon N, Broderick NA, Chakrabarti S, Lemaitre B. 2009a. Invasive and indigenous microbiota impact intestinal stem cell activity through multiple pathways in Drosophila. *Genes Dev* **23**: 2333–2344.
33. Buchon N, Broderick NA, Kuraishi T, Lemaitre B. 2010. Drosophila EGFR pathway coordinates stem cell proliferation and gut remodeling following infection. *BMC Biol* **8**: 152. http://www.biomedcentral.com/1741-7007/8/152.
34. Buchon N, Broderick NA, Poidevin M, Pradervand S, Lemaitre B. 2009b. Drosophila Intestinal Response to Bacterial Infection: Activation of Host Defense and Stem Cell Proliferation. *Cell Host Microbe* **5**: 200–211. http://dx.doi.org/10.1016/j.chom.2009.01.003.
35. Buchon N, Osman D, David FPA, Yu Fang H, Boquete JP, Deplancke B, Lemaitre B. 2013. Morphological and Molecular Characterization of Adult Midgut Compartmentalization in Drosophila. *Cell Rep*.
36. Buscarlet M, Provost S, Zada YF, Bourgoin V, Mollica L, Dubé MP, Busque L. 2018. Lineage restriction analyses in CHIP indicate myeloid bias for TET2 and multipotent stem cell origin for DNMT3A. *Blood* **132**: 277–280.
37. Busque BL, Mio R, Mattioli J, Brais E, Blais N, Lalonde Y, Maragh M, Gilliland DG. 1990. Nonrandom X-Inactivation Patterns in Normal Females: Lyonization Ratios Vary With Age. *Blood* 59–66.
38. Busque L, Patel JP, Figueroa ME, Vasanthakumar A, Provost S, Hamilou Z, Mollica L, Li J, Viale A, Heguy A, et al. 2012. Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis. *Nat Genet* **44**: 1179–1181. http://dx.doi.org/10.1038/ng.2413.
39. Busuttil RA, Garcia AM, Reddick RL, Dollé MET, Calder RB, Nelson JE, Vijg J. 2007. Intra-organ variation in age-related mutation accumulation in the mouse. *PLoS One* **2**: 1–10.
40. Cai C, He HH, Gao S, Chen S, Yu Z, Gao Y, Chen S, Chen MW, Zhang J, Ahmed M, et al. 2014. Lysine-Specific Demethylase 1 Has Dual Functions as a Major Regulator of Androgen Receptor Transcriptional Activity. *Cell Rep* **9**: 1618–1627.
41. Carrero D, Soria-Valles C, López-Otín C. 2016. Hallmarks of progeroid syndromes: lessons from mice and reprogrammed cells. *Dis Model Mech* **9**: 719–735.
42. Carvalho CMB, Lupski JR. 2016. Mechanisms underlying structural variant formation in genomic disorders. *Nat Rev Genet* **17**: 224–238.
43. Cavallucci V, Fidaleo M, Pani G. 2016. Neural Stem Cells and Nutrients: Poised Between Quiescence and Exhaustion. *Trends Endocrinol Metab* **27**: 756–769.
44. Cavenee WK, Dryja t TP, Phillips RA, Benedict WF, Godbout R, Gallie BL, Murphree AL, Strong LC, White RL. 1983. Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. *Nature* **305**: 9–14.
45. Chakrabarti S, Liehl P, Buchon N, Lemaitre B. 2012. Infection-induced host translational blockage inhibits immune responses and epithelial renewal in the Drosophila gut. *Cell Host Microbe* **12**: 60–70.
46. Chandel NS, Jasper H, Ho TT, Passegué E. 2016. Metabolic regulation of stem cell function in tissue homeostasis and organismal ageing. *Nat Cell Biol* **18**: 823–832.
47. Chen J, Astle CM, Harrison DE. 2000. Genetic regulation of primitive

hematopoietic stem cell senescence. *Exp Hematol* **28**: 442–450.
48. Chen J, Xu N, Wang C, Huang P, Huang H, Jin Z, Yu Z, Cai T, Jiao R, Xi R. 2018. Transient Scute activation via a self-stimulatory loop directs enteroendocrine cell pair specification from self-renewing intestinal stem cells. *Nat Cell Biol* **20**: 152–161.
49. Cho IJ, Lui PPW, Obajdin J, Riccio F, Stroukov W, Willis TL, Spagnoli F, Watt FM. 2019. Mechanisms, Hallmarks, and Implications of Stem Cell Quiescence. *Stem Cell Reports* **12**: 1190–1200. https://doi.org/10.1016/j.stemcr.2019.05.012.
50. Choate KA, Lu Y, Zhou J, Choi M, Elias PM, Farhi A, Nelson-Williams C, Crumrine D, Williams ML, Nopper AJ, et al. 2010. Mitotic Recombination in Patients with Ichthyosis Causes Reversion of Dominant Mutations in KRT10. *Science (80-)* **60**: 94–97.
51. Choi NH, Kim JG, Yang DJ, Kim YS, Yoo MA. 2008. Age-related changes in Drosophila midgut are associated with PVF2, a PDGF/VEGF-like growth factor. *Aging Cell* **7**: 318–334.
52. Chunduri NK, Storchová Z. 2019. The diverse consequences of aneuploidy. *Nat Cell Biol* **21**: 54–62. http://dx.doi.org/10.1038/s41556-018-0243-8.
53. Cimini D, Moree B, Canman JC, Salmon ED. 2003. Merotelic kinetochore orientation occurs frequently during early mitosis in mammalian tissue cells and error correction is achieved by two different mechanisms. *J Cell Sci* **116**: 4213–4225.
54. Clark RI, Salazar A, Yamada R, Fitz-Gibbon S, Morselli M, Alcaraz J, Rana A, Rera M, Pellegrini M, Ja WW, et al. 2015. Distinct Shifts in Microbiota Composition during Drosophila Aging Impair Intestinal Function and Drive Mortality. *Cell Rep* **12**: 1656–1667.
55. Conboy IH, Conboy MJ, Smythe GM, Rando TA. 2003. Notch-Mediated Restoration of Regenerative Potential to Aged Muscle. *Science (80-)* **302**: 1575–1577.
56. Coombs CC, Zehir A, Devlin SM, Kishtagari A, Syed A, Jonsson P, Hyman DM, Solit DB, Robson ME, Baselga J, et al. 2017. Therapy-Related Clonal Hematopoiesis in Patients with Non-hematologic Cancers Is Common and Associated with Adverse Clinical Outcomes. *Cell Stem Cell* **21**: 374-382.e4.
57. Cottrell S, Bodmer WF, Bicknell D, Kaklamanis L. 1992. Molecular analysis of APC mutations in familial adenomatous polyposis and sporadic colon carcinomas. *Lancet* **340**: 626–630.
58. Coufal NG, Garcia-perez JL, Peng GE, Yeo GW, Mu Y, Lovci MT, Morell M, Shea KSO, Moran J V, Gage FH. 2009. L1 Retrotransposition in Human Neural Progenitor Cells Nicole. *Nature* **460**: 1127–1131.
59. Cromie GA, Connelly JC, Leach DRF. 2001. Recombination at double-strand breaks and DNA ends: Conserved mechanisms from phage to humans. *Mol Cell* **8**: 1163–1174.
60. Cronin RM, Field JR, Bradford Y, Shaffer CM, Carroll RJ, Mosley JD, Bastarache L, Edwards TL, Hebbring SJ, Lin S, et al. 2014. Phenome-wide association studies demonstrating pleiotropy of genetic variants within FTO with and without adjustment for body mass index. *Appl Genet Epidemiol* **5**: 250. http://journal.frontiersin.org/Journal/10.3389/fgene.2014.00250/full%5Cnhttp://journal.frontiersin.org/Journal/10.3389/fgene.2014.00250/pdf.
61. Crow JF pd., Kimura M. 1965. EVOLUTION IN SEXUAL AND ASEXUAL

POPULATIONS * is advantageous because It has often been said that sexual reproduction of genotypes that can be produced by a recombina- number of the enormous. *Am Nat* **XCIX**: 439–450.
62. Dé F, Paˆques R, Paˆques P, Haber JE. 1999. *Multiple Pathways of Recombination Induced by Double-Strand Breaks in Saccharomyces cerevisiae*.
63. De Navascués J, Perdigoto CN, Bian Y, Schneider MH, Bardin AJ, Martínez-Arias A, Simons BD. 2012. Drosophila midgut homeostasis involves neutral competition between symmetrically dividing intestinal stem cells. *EMBO J* **31**: 2473–2485.
64. Deem A, Keszthelyi A, Blackgrove T, Vayl A, Coffey B, Mathur R, Chabes A, Malkova A. 2011. Break-induced replication is highly inaccurate. *PLoS Biol* **9**: 1000594. /pmc/articles/PMC3039667/?report=abstract (Accessed August 25, 2020).
65. Dolle MET. 2002. Mutational fingerprints of aging. *Nucleic Acids Res* **30**: 545–549.
66. Dolle MET, Snyder WK, Gossen JA, Lohman PHM, Vijg J. 2000. Distinct spectra of somatic mutations accumulated with age in mouse heart and small intestine. *PNAS* **97**.
67. Duijf PHG, Schultz N, Benezra R. 2013. Cancer cells preferentially lose small chromosomes. *Int J Cancer* **132**: 2316–2326.
68. Edifizi D, Nolte H, Babu V, Castells-Roca L, Mueller MM, Brodesser S, Krüger M, Schumacher B. 2017. Multilayered Reprogramming in Response to Persistent DNA Damage in C. elegans. *Cell Rep* **20**: 2026–2043.
69. Erwin JA, Paquola ACM, Singer T, Gallina I, Novotny M, Quayle C, Bedrosian TA, Alves FIA, Butcher CR, Herdy JR, et al. 2016. L1-associated genomic regions are deleted in somatic cells of the healthy human brain. *Nat Neurosci* **19**: 1583–1591.
70. Evrony GD, Cai X, Lee E, Hills LB, Elhosary PC, Lehmann HS, Parker JJ, Atabay KD, Gilmore EC, Poduri A, et al. 2012. Single-neuron sequencing analysis of l1 retrotransposition and somatic mutation in the human brain. *Cell* **151**: 483–496. http://dx.doi.org/10.1016/j.cell.2012.09.035.
71. Faulkner GJ, Garcia-Perez JL. 2017. L1 Mosaicism in Mammals: Extent, Effects, and Evolution. *Trends Genet* **33**: 802–816. http://dx.doi.org/10.1016/j.tig.2017.07.004.
72. Fey MF, Liechti-Gallati S, Von Rohr A, Borisch B, Theilkas L, Schneider V, Oestreicher M, Nagel S, Ziemiecki A, Tobler A. 1994. Clonality and X-Inactivation Patterns in Hematopoietic Cell Populations Detected by the Highly Informative M27@ DNA Probe. *Blood*. www.bloodjournal.org.
73. Flach J, Bakker ST, Mohrin M, Conroy PC, Pietras EM, Reynaud D, Alvarez S, Diolaiti ME, Ugarte F, Forsberg EC, et al. 2014. Replication stress is a potent driver of functional decline in ageing haematopoietic stem cells. *Nature* **512**: 198–202.
74. Franco I, Johansson A, Olsson K, Vrtačnik P, Lundin P, Helgadottir HT, Larsson M, Revêchon G, Bosia C, Pagnani A, et al. 2018. Somatic mutagenesis in satellite cells associates with human skeletal muscle aging. *Nat Commun* **9**.
75. Freeman HJ. 2008. Colorectal cancer risk in Crohn's disease. *World J Gastroenterol* **14**: 1810–1811.
76. Funk LC, Zasadil LM, Weaver BA. 2016. Living in CIN: Mitotic Infidelity and

Its Consequences for Tumor Promotion and Suppression. *Dev Cell* **39**: 638–652.
77. Gan B, Hu J, Jiang S, Liu Y, Sahin E, Zhuang L, Fletcher-Sananikone E, Colla S, Wang YA, Chin L, et al. 2010. Lkb1 regulates quiescence and metabolic homeostasis of haematopoietic stem cells. *Nature* **468**: 701–704.
78. Ganuza M, Hall T, Finkelstein D, Wang Y-D, Chabot A, Kang G, Bi W, Wu G, McKinney-Freeman S. 2019. The global clonal complexity of the murine blood system declines throughout life and after serial transplantation. *Blood* **133**: blood-2018-09-873059.
79. Garcia AM, Calder RB, Dollé MET, Lundell M, Kapahi P, Vijg J. 2010. Age- and temperature-dependent somatic mutation accumulation in Drosophila melanogaster. *PLoS Genet* **6**: 27.
80. Garinis GA, Uittenboogaard LM, Stachelscheid H, Fousteri M, van Ijcken W, Breit TM, van Steeg H, Mullenders LHF, van der Horst GTJ, Brüning JC, et al. 2009. Persistent transcription-blocking DNA lesions trigger somatic growth attenuation associated with longevity. *Nat Cell Biol* **11**: 604–615.
81. Genovese G, Kähler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF, Chambert K, Mick E, Neale BM, Fromer M, et al. 2014. Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence. *N Engl J Med* **371**: 2477–2487. http://www.nejm.org/doi/10.1056/NEJMoa1409405.
82. Gervais L, Bardin AJ. 2017. Tissue homeostasis and aging: new insight from the fly intestine. *Curr Opin Cell Biol*.
83. Giese H, Snyder WK, Van Oostrom C, Van Steeg H, Dollé MET, Vijg J. 2002. Age-related mutation accumulation at a lacZ reporter locus in normal and tumor tissues of Trp53-deficient mice. *Mutat Res - Genet Toxicol Environ Mutagen* **514**: 153–163.
84. Guo L, Karpac J, Tran SL, Jasper H. 2014. PGRP-SC2 promotes gut immune homeostasis to limit commensal dysbiosis and extend lifespan. *Cell* **156**: 109–122.
85. Gutierrez-martinez P, Hogdal L, Nagai M, Kruta M, Singh R, Sarosiek K, Nussenzweig A, Beerman I, Letai A, Rossi DJ. 2018. Diminished apoptotic priming and ATM signalling confer a survival advantage onto aged haematopoietic stem cells in response to DNA damage. *Nat Cell Biol*. http://dx.doi.org/10.1038/s41556-018-0054-y.
86. Ha EM, Oh CT, Ryu JH, Bae YS, Kang SW, Jang I hwan, Brey PT, Lee WJ. 2005. An antioxidant system required for host protection against gut infection in Drosophila. *Dev Cell* **8**: 125–132.
87. Haber JE. 2007. *Genome Dynamics and Stability*. http://www.springerlink.com/index/4467604188l31721.pdf.
88. Haigis KM, Caya JG, Reichelderfer M, Dove WF. 2002. Intestinal adenomas can develop with a stable karyotype and stable microsatellites. *Proc Natl Acad Sci U S A* **99**: 8927–8931.
89. Harrison DE. 1978. Loss of proliferative capacity in immunohemopoietic stem cells caused by serial transplantation rather than aging. *J Exp Med* **147**: 1526–1531.
90. Hartenstein V, Martinez P. 2019. Structure, development and evolution of the digestive system. *Cell Tissue Res* **377**: 289–292.
91. He L, Si G, Huang J, Samuel ADT, Perrimon N. 2018. Mechanical regulation of stem-cell differentiation by the stretch-activated Piezo channel.

Nature **555**: 103–106.
92. Hicks WM, Yamaguchi M, Haber JE. 2011. Real-time analysis of double-strand DNA break repair by homologous recombination. *Proc Natl Acad Sci U S A* **108**: 3108–3115.
93. Hirschhorn R, Yang DR, Puck JM, Huie ML, Jiang C, Kurlandsky LE. 1996. Spontaneous in vivo reversion to normal of an inherited mutation in a patient with adenoside deaminase deficiency. *Nat Genet* **13**.
94. Holstege H, Pfeiffer W, Sie D, Hulsman M, Nicholas TJ, Lee CC, Ross T, Lin J, Miller MA, Ylstra B, et al. 2014. Somatic mutations found in the healthy blood compartment of a 115-yr-old woman demonstrate oligoclonal hematopoiesis. *Genome Res* **24**: 733–742.
95. Howarth K, Ranta S, Winter E, Teixeira A, Schaschl H, Harvey JJ, Rowan A, Jones A, Spain S, Clark S, et al. 2009. A mitotic recombination map proximal to the APC locus on chromosome 5q and assessment of influences on colorectal cancer risk. *BMC Med Genet* **10**: 1–11.
96. Hsieh JCF, Van Den Berg D, Kang H, Hsieh CL, Lieber MR. 2013. Large chromosome deletions, duplications, and gene conversion events accumulate with age in normal human colon crypts. *Aging Cell* **12**: 269–279.
97. Hu DJK, Jasper H. 2019. Control of Intestinal Cell Fate by Dynamic Mitotic Spindle Repositioning Influences Epithelial Homeostasis and Longevity. *Cell Rep* **28**: 2807-2823.e5.
98. Huang N, Lee I, Marcotte EM, Hurles ME. 2010. Characterising and predicting haploinsufficiency in the human genome. *PLoS Genet* **6**: 1–11.
99. Inomata K, Aoto T, Binh NT, Okamoto N, Tanimura S, Wakayama T, Iseki S, Hara E, Masunaga T, Shimizu H, et al. 2009. Genotoxic Stress Abrogates Renewal of Melanocyte Stem Cells by Triggering Their Differentiation. *Cell* **137**: 1088–1099. http://dx.doi.org/10.1016/j.cell.2009.03.037.
100. Issigonis M, Tulina N, De Cuevas M, Brawley C, Sandler L, Matunist E. 2009. JAK-STAT signal inhibition regulates competition in the drosophila testis stem cell niche. *Science (80-)* **326**: 153–156.
101. Jacobs KB, Yeager M, Zhou W, Wacholder S, Rodriguez-santiago B, Hutchinson A, Deng X, Horner M, Cullen M, Epstein CG, et al. 2012. Detectable clonal mosaicism and its relationship to aging and cancer. *Nat Genet* **44**: 651–658.
102. Jaiswal S, Ebert BL. 2019. Clonal hematopoiesis in human aging and disease. *Science* **366**. http://www.ncbi.nlm.nih.gov/pubmed/31672865.
103. Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman P V, Mar BG, Coleman Lindsley R, Mermel CH, Burtt N, Chavez A, et al. 2014. Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes NIH Public Access. *N Engl J Med* **371**: 2488–2498. http://cancer.sanger.ac.uk/cancergenome/projects/cosmic.
104. Janzen V, Forkert R, Fleming HE, Saito Y, Waring MT, Dombkowski DM, Cheng T, DePinho RA, Sharpless NE, Scadden DT. 2006. Stem-cell ageing modified by the cyclin-dependent kinase inhibitor p16 INK4a. *Nature* **443**: 421–426.
105. Jiang H, Edgar BA. Intestinal stem cells in the adult Drosophila midgut.
106. Jiang H, Patel PH, Kohlmaier A, Grenley MO, McEwen DG, Edgar BA. 2009a. Cytokine/Jak/Stat Signaling Mediates Regeneration and

Homeostasis in the Drosophila Midgut. *Cell* **137**: 1343–1355.
107. Jiang H, Patel PH, Kohlmaier A, Grenley MO, McEwen DG, Edgar BA. 2009b. Cytokine/Jak/Stat Signaling Mediates Regeneration and Homeostasis in the Drosophila Midgut. *Cell* **137**: 1343–1355.
108. Jin Z, Kirilly D, Weng C, Kawase E, Song X, Smith S, Schwartz J, Xie T. 2008. Differentiation-Defective Stem Cells Outcompete Normal Stem Cells for Niche Occupancy in the Drosophila Ovary. *Cell Stem Cell* **2**: 39–49.
109. Jonason A S, Kunala S, Price G J, Restifo R J, Spinelli H M, Persing J A, Leffell D J, Tarone R E, Brash D E. 1996. Frequent clones of p53-mutated keratinocytes in normal human skin. *Proc Natl Acad Sci U S A* **93**: 14025–14029.
110. Jongmans MCJ, Diets I, Quarello P, Garelli E, Kuiper R, Pfundt R. 2018. Somatic reversion events point towards RPL4 as a novel disease gene in a condition resembling Diamond-Blackfan anemia. *Haematologica* **103**: 607–609.
111. Jongmans MCJ, Verwiel ETP, Heijdra Y, Vulliamy T, Kamping EJ, Hehir-Kwa JY, Bongers EMHF, Pfundt R, Van Emst L, Van Leeuwen FN, et al. 2012. Revertant somatic mosaicism by mitotic recombination in dyskeratosis congenita. *Am J Hum Genet* **90**: 426–433. http://dx.doi.org/10.1016/j.ajhg.2012.01.004.
112. Jonkman MF, Scheffer H, Stulp R, Pas HH, Nijenhuis M, Heeres K, Owaribe K, Pulkkinen L, Uitto J. 1997. Revertant mosaicism in epidermolysis bullosa caused by mitotic gene conversion. *Cell* **88**: 543–551.
113. Kamminga LM, Bystrykh L V, De Boer A, Houwer S, Douma J, Weersing E, Dontje B, De Haan G. 2006. The Polycomb group gene Ezh2 prevents hematopoietic stem cell exhaustion. *Blood* **107**. http://genereg.ornl.gov/webgestalt/,.
114. Karin M, Clevers H. 2016. Reparative inflammation takes charge of tissue regeneration. *Nat Rev* **529**.
115. Kiritsi D, Jo N, Tljo BI, He Y, Pasmooij AMG, Onder M, Happle R, Jonkman MF, Bruckner-tuderman L, Has C. 2012. Revertant mosaicism in a human skin fragility disorder results from slipped mispairing and mitotic recombination. *J Clin Invest* **122**: 1742–1746.
116. Knudson AG. 1995. Mutation and cancer: A personal odyssey. *Adv Cancer Res* **67**: 1–23.
117. Kolahgar G, Suijkerbuijk SJE, Kucinski I, Poirier EZ, Mansour S, Simons BD, Piddini E. 2015. Cell Competition Modifies Adult Stem Cell and Tissue Population Dynamics in a JAK-STAT-Dependent Manner. *Dev Cell* **34**: 297–309. http://dx.doi.org/10.1016/j.devcel.2015.06.010.
118. Kozar S, Morrissey E, Nicholson AM, van der Heijden M, Zecchini HI, Kemp R, Tavaré S, Vermeulen L, Winton DJ. 2013. Continuous Clonal Labeling Reveals Small Numbers of Functional Stem Cells in Intestinal Crypts and Adenomas. *Cell Stem Cell* **13**: 626–633.
119. Krishnamurthy J, Ramsey MR, Ligon KL, Torrice C, Koh A, Bonner-Weir S, Sharpless NE. 2006. $p16^{INK4a}$ induces an age-dependent decline in islet regenerative potential. *Nature* **443**: 453–457.
120. Kuhn HG, Dickinson-Anson H, Gage FH. 1996. Neurogenesis in the dentate gyrus of the adult rat: age-related decrease of neuronal progenitor proliferation. *J Neurosci* **16**: 2027–33. http://www.ncbi.nlm.nih.gov/pubmed/8604047.

121. Lahue RS, Au KG, Modrich P. 1989. Mismatch Correction in a Defined System. *Science (80-)* **245**: 1–6.
122. Lapunzina P, Monk D. 2011. The consequences of uniparental disomy and copy number neutral loss-of-heterozygosity during human development and cancer. *Biol Cell* **103**: 303–317.
123. Laurie CC, Laurie CA, Rice K, Doheny KF, Zelnick LR, Mchugh CP, Ling H, Hetrick KN, Pugh EW, Amos C, et al. 2012. Detectable clonal mosaicism from birth to old age and its relationship to cancer A. *Nat Genet* **44**: 642–650.
124. Leblond C, Walker B. 1956. Renewal of Cell Populations. *From Dep Anatomy, McGill Univ Montr Canada* 1–9.
125. Lee-Six H, Olafsson S, Ellis P, Osborne R, Mathijs S, Moore L, Georgakopoulos N, Torrente F, Noorani A, Goddard M, et al. 2019. The landscape of somatic mutation in normal colorectal epithelial cells. *Nature* **574**.
126. Lee JH, Huynh M, Silhavy JL, Kim S, Dixon-Salazar T, Heiberg A, Scott E, Bafna V, Hill KJ, Collazo A, et al. 2012. De novo somatic mutations in components of the PI3K-AKT3-mTOR pathway cause hemimegalencephaly. *Nat Genet* **44**: 941–5. http://www.ncbi.nlm.nih.gov/pubmed/22729223.
127. Lee PS, Greenwell PW, Dominska M, Gawel M, Hamilton M, Petes TD. 2009. A Fine-Structure Map of Spontaneous Mitotic Crossovers in the Yeast Saccharomyces cerevisiae. **5**.
128. Lehane M, Billingsley P. 1996. *Biology of the Insect Midgut*. Chapman and Hal, London.
129. Lemaitre B, Hoffmann J. 2007. The Host Defense of Drosophila melanogaster . *Annu Rev Immunol* **25**: 697–743.
130. Lemaitre B, Miguel-Aliaga I. 2013. The Digestive Tract of Drosophila melanogaster . *Annu Rev Genet* **47**: 377–404.
131. Levayer R. 2019. Solid stress, competition for space and cancer: the opposing roles of mechanical cell competition in tumour initiation and growth. *Semin Cancer Biol*. https://linkinghub.elsevier.com/retrieve/pii/S1044579X1930015X.
132. Li H, Qi Y, Jasper H. 2013. Dpp Signaling Determines Regional Stem Cell Identity in the Regenerating Adult Drosophila Gastrointestinal Tract. *Cell Rep*.
133. Li H, Qi Y, Jasper H. 2016. Preventing Age-Related Decline of Gut Compartmentalization Limits Microbiota Dysbiosis and Extends Lifespan. *Cell Host Microbe*.
134. Li Z, Cai X, Cai C, Wang J, Zhang W, Petersen BE, Yang F. 2011. Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. **118**: 4509–4519.
135. Li Z, Woo CJ, Iglesias-Ussel MD, Ronai D, Scharff MD. 2004. The generation of antibody diversity through somatic hypermutation and class switch recombination. *Genes Dev* **18**: 1–11.
136. Lin JK, Chang SC, Yang YC, Li AFY. 2003. Loss of heterozygosity and DNA aneuploidy in colorectal adenocarcinoma. *Ann Surg Oncol* **10**: 1086–1094.
137. Lindahl T. 1974. An N-glycosidase from Escherichia coli that releases free uracil from DNA containing deaminated cytosine residues. *Proc Natl Acad*

Sci U S A **71**: 3649–53. http://www.ncbi.nlm.nih.gov/pubmed/4610583%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC433833.
138. Lindahl T, Barnes DE. 2000. Repair of Endogenous DNA Damage. *Cold Spring Harb Symp Quant Biol* **65**: 127–134. http://symposium.cshlp.org/cgi/doi/10.1101/sqb.2000.65.127.
139. Lindahl T, Nyberg B. 1972. Rate of Depurination of Native Deoxyribonucleic Acid. *Biochemistry* **11**: 3610–3618.
140. Lindhurst MJ, Parker VER, Payne F, Sapp JC, Rudge S, Harris J, Witkowski AM, Zhang Q, Groeneveld MP, Scott CE, et al. 2012. Mosaic overgrowth with fibroadipose hyperplasia is caused by somatic activating mutations in PIK3CA. *Nat Genet* **44**: 928–933.
141. Liu N, Matsumura H, Kato T, Ichinose S, Takada A, Namiki T, Asakawa K, Morinaga H, Mohri Y, De Arcangelis A, et al. 2019. Stem cell competition orchestrates skin homeostasis and ageing. *Nature*. http://www.nature.com/articles/s41586-019-1085-7.
142. Lodato MA, Barton AR, Yandava CN, Woodworth MB, Yang P, Walsh CA, Park PJ, Sherman MA, Hatem NE, Kwon M, et al. 2017. Aging and neurodegeneration are associated with increased mutations in single human neurons. *Science (80-)* **359**: 555–559.
143. Lodato MA, Woodworth MB, Lee S, Evrony GD, Mehta BK, Karger A, Chittenden TW, D'Gama AM, Cai X, Luquette LJ, et al. 2015. Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science (80-)* **350**: 94–98. http://www.ncbi.nlm.nih.gov/pubmed/26430121.
144. Lopez-Garcia C, Klein AM, Simons BD, Douglas †, Winton J. 2010. Intestinal Stem Cell Replacement Follows a Pattern of Neutral Drift. *Science (80-)* **330**. http://science.sciencemag.org/.
145. Mable BK, Otto SP. 2001. Masking and purging mutations following EMS treatment in haploid, diploid and tetraploid yeast (Saccharomyces cerevisiae). *Genet Res* **77**: 9–26.
146. MacRae SL, Croken MM, Calder RB, Aliper A, Milholland B, White RR, Zhavoronkov A, Gladyshev VN, Seluanov A, Gorbunova V, et al. 2015. DNA repair in species with extreme lifespan differences. *Aging (Albany NY)* **7**: 1171–84. http://www.ncbi.nlm.nih.gov/pubmed/26729707%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4712340.
147. Marianes A, Spradling AC. 2013. Physiological and stem cell compartmentalization within the Drosophila midgut. *Elife* **2013**: 1–19.
148. Martin N, Beach D. 2014. Ageing as developmental decay : *Trends Mol Med* **20**: 1–8.
149. Martincorena I, Fowler JC, Wabik A, Lawson ARJ, Abascal F, Hall MWJ, Cagan A, Murai K, Mahbubani K, Stratton MR, et al. 2018a. Somatic mutant clones colonize the human esophagus with age. **3879**: 1–14.
150. Martincorena I, Fowler JC, Wabik A, Lawson ARJ, Abascal F, Hall MWJ, Cagan A, Murai K, Mahbubani K, Stratton MR, et al. 2018b. Somatic mutant clones colonize the human esophagus with age. *Science*.
151. Martincorena I, Fowler JC, Wabik A, Lawson ARJ, Abascal F, Hall MWJ, Cagan A, Murai K, Mahbubani K, Stratton MR, et al. 2018c. Supplementary Materials for. 1–39.

152. Martincorena I, McLaren S, Tubio JM, Campbell PJ, Van Loo P, Alexandrov LB, Jones PH, Fullam A, Roshan A, Widaa S, et al. 2015. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science (80-)* **348**: 880–886.
153. Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, Van Loo P, Davies H, Stratton MR, Campbell PJ. 2017. Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell*. http://linkinghub.elsevier.com/retrieve/pii/S0092867417311364.
154. Matsumura H, Mohri Y, Thanh Binh N, Morinaga H, Fukuda M, Ito M, Kurata S, Hoeijmakers J, Nishimura EK. 2016. Stem cells: Hair follicle aging is driven by transepidermal elimination of stem cells via COL17A1 proteolysis. *Science (80-)* **351**.
155. McClintock B. 1950. The Origin and Behaviour of Mutable Loci in Maize. *Proc Natl Acad Sci* **36**: 345–355.
156. McKerrell T, Park N, Moreno T, Grove CS, Ponstingl H, Stephens J, Crawley C, Craig J, Scott MA, Hodkinson C, et al. 2015. Leukemia-Associated Somatic Mutations Drive Distinct Patterns of Age-Related Clonal Hemopoiesis. *Cell Rep* **10**: 1239–1245.
157. McVey M, Lee SE. 2008. MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genet* **24**: 529–538.
158. Merritt AJ, Potten CS, Watson AJM, Loh DY, Nakayama K, Hickman JA. 1995. Differential expression of bcl-2 in intestinal epithelia Correlation with attenuation of apoptosis in colonic crypts and the incidence of colonic. **2271**: 2261–2271.
159. Micchelli CA, Perrimon N. 2006a. Evidence that stem cells reside in the adult Drosophila midgut epithelium. *Nature* **439**: 475–479.
160. Micchelli CA, Perrimon N. 2006b. Evidence that stem cells reside in the adult Drosophila midgut epithelium. *Nature*.
161. Miguel-Aliaga I, Jasper H, Lemaitre B. 2018. Anatomy and physiology of the digestive tract of drosophila melanogaster. *Genetics*.
162. Miller A. 1994. *The Biology of Drosophila*. Cold Spring Harbour Laboratory Press.
163. Miller A. 1965. *The internal anatomy and histology of the imago of Drosophila melanogaster.*
164. Mohrin M, Bourke E, Alexander D, Warr MR, Barry-holson K, Beau MM Le, Morrison CG, Passegué E. 2011. Repair and Mutagenesis. *Cell Stem Cell* **7**: 174–185.
165. Molofsky A V., Slutsky SG, Joseph NM, He S, Pardal R, Krishnamurthy J, Sharpless NE, Morrison SJ. 2006. Increasing $p16^{INK4a}$ expression decreases forebrain progenitors and neurogenesis during ageing. *Nature* **443**: 448–452.
166. Morata G, Ripoll P. 1975. *Minutes: Mutants of Drosophila Autonomously Affecting Cell Division Rate*.
167. Moreno E, Basler K. 2004. dMyc Transforms Cells into Super-Competitors developing tissue remains unchanged and no morpho. *Cell* **117**: 117–129.
168. Morrison SJ, Wandycz AM, Akashi K, Globerson A, Weissman IL. 1996. The aging of hematopoietic stem cells. *Nat Med* **2**: 1011–1016. https://doi.org/10.1038/nm0996-1011.
169. Moskalev AA, Shaposhnikov M V., Plyusnina EN, Zhavoronkov A,

Budovsky A, Yanai H, Fraifeld VE. 2013. The role of DNA damage and repair in aging through the prism of Koch-like criteria. *Ageing Res Rev* **12**: 661–684. http://dx.doi.org/10.1016/j.arr.2012.02.001.

170. Mueller MM, Castells-roca L, Babu V, Ermolaeva MA. 2015. Europe PMC Funders Group DAF-16 / FoxO and EGL-27 / GATA promote developmental growth in response to persistent somatic DNA damage. *Nat Cell Biol* **16**: 1168–1179.
171. Muotri AR, Chu VT, Marchetto MCN, Deng W, Moran J V., Gage FH. 2005. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* **435**: 903–910.
172. Nakazawa H, English D, Randell PL, Nakazawa K, Martel N, Armstrong BK, Yamasaki H. 1993. UV and skin cancer: specific p53 gene mutation in normal skin as a biologically relevant exposure measurement. *Proc Natl Acad Sci* **91**: 360–364.
173. Nicholson AM, Olpe C, Hoyle A, Thorsen AS, Rus T, Colombé M, Brunton-Sim R, Kemp R, Marks K, Quirke P, et al. 2018. Fixation and Spread of Somatic Mutations in Adult Human Colonic Epithelium. *Cell Stem Cell* **22**: 909-918.e8.
174. Niedernhofer LJ, Gurkar AU, Wang Y, Vijg J, Hoeijmakers JHJ, Robbins PD. 2018. Annual Review of Biochemistry Nuclear Genomic Instability and Aging. *Annu Rev Biochem* 295–322.
175. Nishimura EK, Granter SR, Fisher DE. 2005. Mechanisms of hair graying: Incomplete melanocyte stem cell maintenance in the niche. *Science (80-)* **307**: 720–724.
176. Nystul T, Spradling A. 2007. An Epithelial Niche in the Drosophila Ovary Undergoes Long-Range Stem Cell Replacement. *Cell Stem Cell* **1**: 277–285.
177. O'Brien LE. 2013. Regional specificity in the drosophila midgut: Setting boundaries with stem cells. *Cell Stem Cell* **13**: 375–376.
178. O'Huallachain M, Karczewski KJ, Weissman SM, Urban AE, Snyder MP. 2012. Extensive genetic variation in somatic human tissues. *Proc Natl Acad Sci U S A* **109**: 18018–23. http://www.ncbi.nlm.nih.gov/pubmed/23043118%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3497787.
179. Oberdoerffer P, Michan S, McVay M, Mostoslavsky R, Vann J, Park SK, Hartlerode A, Stegmuller J, Hafner A, Loerch P, et al. 2008. SIRT1 Redistribution on Chromatin Promotes Genomic Stability but Alters Gene Expression during Aging. *Cell* **135**: 907–918.
180. Ohlstein B, Spradling A. 2007. Multipotent Drosophila intestinal stem cells specify daughter cell fates by differential notch signaling. *Science (80-)* **315**: 988–993.
181. Ohlstein B, Spradling A. 2006. The adult Drosophila posterior midgut is maintained by pluripotent stem cells. *Nature*.
182. Otto SP, Goldstein DB. 1992. Recombination and the evolution of diploidy. *Genetics* **131**: 745–751.
183. Park JS, Lee SH, Na HJ, Pyo JH, Kim YS, Yoo MA. 2012. Age- and oxidative stress-induced DNA damage in Drosophila intestinal stem cells as marked by Gamma-H2AX. *Exp Gerontol* **47**: 401–405.
184. Perdigoto CN, Schweisguth F, Bardin AJ. 2011. Distinct levels of Notch activity for commitment and terminal differentiation of stem cells in the adult

fly intestine. *Development* **138**: 4585–4595.
185. Perrot V, Richard S, Valero M. 1991. Transition from haploidy to diploidy. *Nature* **351**.
186. Petruseva IO, Evdokimov AN, Lavrik OI. 2017. Genome Stability Maintenance in Naked. *Acta Naturae* **9**: 31–41.
187. Poduri A, Evrony GD, Xuyu C, Elhosary PC, Beroukhim R, Lehtinen MK, Hills LB, Heinzen EL, Hill A, Hill RS, et al. 2012. Somatic Activation ofAKT3 Causes Hemispheric Developmental Brain Malformations. *Neuron* **74**: 41–48.
188. Potten CS, Grant HK. 1998. The relationship between ionizing radiation-induced apoptosis and stem cells in the small and large intestine. *Br J Cancer* **78**: 993–1003. http://www.ncbi.nlm.nih.gov/pubmed/9792141.
189. Razin A, Riggs A. 1980. DNA Methymation and Gene Function. *Science (80-)* **210**.
190. Rehen SK, McConnell MJ, Kaushal D, Kingsbury MA, Yang AH, Chun J. 2002. Chromosomal variation in neurons of the developing and adult mammalian nervous system. *Proc Natl Acad Sci* **98**: 13361–13366.
191. Rera M, Bahadorani S, Cho J, Koehler CL, Ulgherait M, Hur JH, Ansari WS, Lo T, Jones DL, Walker DW. 2011. Modulation of longevity and tissue homeostasis by the drosophila PGC-1 homolog. *Cell Metab* **14**: 623–634.
192. Rera M, Clark RI, Walker DW. 2012. Intestinal barrier dysfunction links metabolic and inflammatory markers of aging to death in Drosophila. *Proc Natl Acad Sci U S A* **109**: 21528–21533.
193. Resnik-Docampo M, Koehler CL, Clark RI, Schinaman JM, Sauer V, Wong DM, Lewis S, D'Alterio C, Walker DW, Jones DL. 2017. Tricellular junctions regulate intestinal stem cell behaviour to maintain homeostasis. *Nat Cell Biol* **19**: 52–59.
194. Revy P, Kannengiesser C, Fischer A. 2019. Somatic genetic rescue in Mendelian haematopoietic diseases. *Nat Rev Genet* **20**: 582–598. http://dx.doi.org/10.1038/s41576-019-0139-x.
195. Rivière JB, Mirzaa GM, O'Roak BJ, Beddaoui M, Alcantara D, Conway RL, St-Onge J, Schwartzentruber JA, Gripp KW, Nikkel SM, et al. 2012. De novo germline and postzygotic mutations in AKT3, PIK3R2 and PIK3CA cause a spectrum of related megalencephaly syndromes. *Nat Genet* **44**: 934–940.
196. Rogakou EP, Pilch DR, Orr AH, Ivanova VS, Bonner WM. 1998. *DNA Double-stranded Breaks Induce Histone H2AX Phosphorylation on Serine 139* Downloaded from*. http://www.jbc.org/.
197. Rossi DJ, Bryder D, Seita J, Nussenzweig A, Hoeijmakers J, Weissman IL. 2007. Deficiencies in DNA damage repair limit the function of haematopoietic stem cells with age. *Nature* **447**: 725–729.
198. Rübe CE, Fricke A, Widmann TA, Fürst T, Madry H, Pfreundschuh M, Rübe C. 2011. Accumulation of DNA damage in hematopoietic stem and progenitor cells during human aging. *PLoS One* **6**.
199. Sacco A, Mourkioti F, Tran R, Choi J, Llewellyn M, Kraft P, Shkreli M, Delp S, Pomerantz JH, Artandi SE, et al. 2010. Short telomeres and stem cell exhaustion model duchenne muscular dystrophy in mdx/mTR mice. *Cell* **143**: 1059–1071.
200. Saini N, Ramakrishnan S, Elango R, Ayyar S, Zhang Y, Deem A, Ira G, Haber JE, Lobachev KS, Malkova A. 2013. Migrating bubble during break-

induced replication drives conservative DNA synthesis. *Nature* **502**: 389–392.
201. Sakofsky CJ, Roberts SA, Malc E, Mieczkowski PA, Resnick MA, Gordenin DA, Malkova A. 2014. Break-induced replication is a source of mutation clusters underlying kataegis. *Cell Rep* **7**: 1640–1648.
202. Sancar A. 1993. Photobiology school nucleotide excision. **57**: 905–921.
203. Saus E, Iraola-Guzmán S, Willis JR, Brunet-Vega A, Gabaldón T. 2019. Microbiome and colorectal cancer: Roles in carcinogenesis and clinical potential. *Mol Aspects Med*.
204. Schaniel C, Sirabella D, Qiu J, Niu X, Lemischka IR, Moore KA. 2011. Wnt-inhibitory factor 1 dysregulation of the bone marrow niche exhausts hematopoietic stem cells. *Blood* **118**: 2420–2429.
205. Scopelliti A, Cordero JB, Diao F, Strathdee K, White BH, Sansom OJ, Vidal M. 2014. Local control of intestinal stem cell homeostasis by enteroendocrine cells in the adult Drosophila midgut. *Curr Biol* **24**: 1199–1211.
206. Serra E, Rosenbaum T, Nadal M, Winner U, Ars E, Estivill X, Lázaro C. 2001. Erratum: Mitotic recombination effects homozygosity for NF1 germline mutations in neurofibromas (Nature Genetics (2001) 28 (294-296)). *Nat Genet* **29**: 100.
207. Sharpless NE, DePinho RA. 2007. How stem cells age and why this makes us grow old. *Nat Rev Mol Cell Biol* **8**: 703–713.
208. Sheltzer JM, Amon A. 2011. The aneuploidy paradox: Costs and benefits of an incorrect karyotype. *Trends Genet* **27**: 446–453. http://dx.doi.org/10.1016/j.tig.2011.07.003.
209. Shima N, Hartford SA, Duffy T, Wilson LA, Schimenti KJ, Schimenti JC. 2003. Phenotype-based identification of mouse chromosome instability mutants. *Genetics* **163**: 1031–1040.
210. Singh SK, Singh S, Gadomski S, Sun L, Pfannenstein A, Magidson V, Chen X, Kozlov S, Tessarollo L, Klarmann KD, et al. 2018. Id1 Ablation Protects Hematopoietic Stem Cells from Stress-Induced Exhaustion and Aging. *Cell Stem Cell* **23**: 252-265.e8.
211. Siudeja K, Bardin AJ. 2017a. Somatic recombination in adult tissues: What is there to learn? *Fly (Austin)* **11**: 121–128. http://dx.doi.org/10.1080/19336934.2016.1249073.
212. Siudeja K, Bardin AJ. 2017b. Somatic recombination in adult tissues: What is there to learn? *Fly (Austin)* **11**: 121–128. http://dx.doi.org/10.1080/19336934.2016.1249073.
213. Siudeja K, Nassari S, Gervais L, Skorski P, Lameiras S, Stolfa D, Zande M, Bernard V, Frio TR, Bardin AJ. 2015. Frequent Somatic Mutation in Adult Intestinal Stem Cells Drives Neoplasia and Genetic Mosaicism during Aging. *Cell Stem Cell* **17**: 663–674. http://dx.doi.org/10.1016/j.stem.2015.09.016.
214. Slaughter P, Southwick W. 1953. “Field cancerization” in oral stratified squamous epithelium. *Cancer* **6**: 1963–1968.
215. Smith JJ, Timoshevskaya N, Ye C, Holt C, Keinath MC, Parker HJ, Cook ME, Hess JE, Narum SR, Lamanna F, et al. 2018. The sea lamprey germline genome provides insights into programmed genome rearrangement and vertebrate evolution. *Nat Genet* **50**: 270–277.
216. Snippert HJ, van der Flier LG, Sato T, van Es JH, van den Born M, Kroon-

Veenboer C, Barker N, Klein AM, van Rheenen J, Simons BD, et al. 2010. Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. *Cell* **143**: 134–144. http://dx.doi.org/10.1016/j.cell.2010.09.016.

217. Snodgrass R. 1935. *Principles of Insect Morphology*. Annals of the Entomological Society of America https://doi.org/10.1093/aesa/28.3.408.
218. Sotiropoulou PA, Candi A, Mascré G, De Clercq S, Youssef KK, Lapouge G, Dahl E, Semeraro C, Denecker G, Marine JC, et al. 2010. Bcl-2 and accelerated DNA repair mediates resistance of hair follicle bulge stem cells to DNA-damage-induced cell death. *Nat Cell Biol* **12**: 572–582. http://dx.doi.org/10.1038/ncb2059.
219. Sousa-Victor P, Gutarra S, García-Prat L, Rodriguez-Ubreva J, Ortet L, Ruiz-Bonilla V, Jardí M, Ballestar E, González S, Serrano AL, et al. 2014. Geriatric muscle stem cells switch reversible quiescence into senescence. *Nature* **506**: 316–321.
220. Spalding KL, Bergmann O, Alkass K, Bernard S, Salehpour M, Huttner HB, Boström E, Westerlund I, Vial C, Buchholz BA, et al. 2013. XDynamics of hippocampal neurogenesis in adult humans. *Cell* **153**: 1219.
221. St. Charles J, Petes TD. 2013. High-Resolution Mapping of Spontaneous Mitotic Recombination Hotspots on the 1.1 Mb Arm of Yeast Chromosome IV. *PLoS Genet* **9**.
222. Staber PB, Zhang P, Ye M, Welner RS, Nombela-Arrieta C, Bach C, Kerenyi M, Bartholdy BA, Zhang H, Alberich-Jordà M, et al. 2013. Sustained PU.1 Levels Balance Cell-Cycle Regulators to Prevent Exhaustion of Adult Hematopoietic Stem Cells. *Mol Cell* **49**: 934–946.
223. Stern C. 1936. Somatic crossing over and segregation in Drosophila Melanogaster. *Genetics* **21**: 625–730.
224. Sudo K, Ema H, Morita Y, Nakauchi H. 2000. Age-associated characteristics of murine hematopoietic stem cells. *J Exp Med* **192**: 1273–80. http://www.ncbi.nlm.nih.gov/pubmed/11067876%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2193349.
225. Sweetser DB, Hough H, Whelden JF, Arbuckle M, Nickoloff JA. 1994. Fine-resolution mapping of spontaneous and double-strand break-induced gene conversion tracts in Saccharomyces cerevisiae reveals reversible mitotic conversion polarity. *Mol Cell Biol* **14**: 3863–3875.
226. Thiagalingam S, Laken S, Willson JKV, Markowitz SD, Kinzler KW, Vogelstein B, Lengauer C. 2001. Mechanisms underlying losses of heterozygosity in human colorectal cancers. *Proc Natl Acad Sci U S A* **98**: 2698–2702.
227. Tian D, Zumei F, Nancy H, R. Woodrow S, Judy M, David D. 1998. Multifocal accumulation of p53 protein in esophageam carcinoma: Evidence for field cancerization. *Int J Cancer* **575**: 568–575.
228. Tierney MT, Stec MJ, Rulands S, Simons BD, Sacco A, Jolla L, Program R, Burnham S, Medical P, Jolla L, et al. 2018. to tissue repair and homeostatic aging. *Cell Stem Cell* **22**: 119–127.
229. Tighe A, Johnson VL, Albertella M, Taylor SS. 2001. Aneuploid colon cancer cells have a robust spindle checkpoint. *EMBO Rep* **2**: 609–614.
230. Traulsen A, Lenaerts T, Pacheco JM, Dingli D. 2013. On the dynamics of neutral mutations in a mathematical model for a homogeneous stem cell

population. *J R Soc Interface* **10**.
231. Tuna M, Knuutila S, Mills GB. 2009. Uniparental disomy in cancer. *Trends Mol Med* **15**: 120–128.
232. Upton KR, Gerhardt DJ, Jesuadian JS, Richardson SR, Sánchez-Luque FJ, Bodea GO, Ewing AD, Salvador-Palomeque C, Van Der Knaap MS, Brennan PM, et al. 2015. Ubiquitous L1 mosaicism in hippocampal neurons. *Cell* **161**: 228–239.
233. van Velthoven CTJ, Rando TA. 2019. Stem Cell Quiescence: Dynamism, Restraint, and Cellular Idling. *Cell Stem Cell* **24**: 213–225. https://doi.org/10.1016/j.stem.2019.01.001.
234. Vaughen J, Igaki T. 2016. Slit-Robo Repulsive Signaling Extrudes Tumorigenic Cells from Epithelia. *Dev Cell* **39**: 683–695. http://dx.doi.org/10.1016/j.devcel.2016.11.015.
235. Venugopal P, et al. 2017. Self-reverting mutations partially correct the blood phenotype in a Diamonf Blackfan anemia patient. *Haematologica* **102**: 117–120.
236. Vodovar N, Vinals M, Liehl P, Basset A, Degrouard J, Spellman P, Boccard F, Lemaitre B. 2005. Drosophila host defense after oral infection by an entomopathogenic Pseudomonas species. *Proc Natl Acad Sci* **102**: 11414–11419. http://www.pnas.org/cgi/doi/10.1073/pnas.0502240102.
237. Walter D, Lier A, Geiselhart A, Thalheimer FB, Huntscha S, Sobotta MC, Moehrle B, Brocks D, Bayindir I, Kaschutnig P, et al. 2015. Exit from dormancy provokes DNA-damage-induced attrition in haematopoietic stem cells. *Nature* **520**: 549–552.
238. Wang J, Davis RE. 2014. Programmed DNA elimination in multicellular organisms. *Curr Opin Genet Dev*.
239. Welch JS, Ley TJ, Link DC, Miller CA, Larson DE, Koboldt DC, Wartman LD, Lamprecht TL, Liu F, Xia J, et al. 2012. The origin and evolution of mutations in acute myeloid leukemia. *Cell* **150**: 264–278.
240. Williams AB, Schumacher B. 2017. DNA damage responses and stress resistance: Concepts from bacterial SOS to metazoan immunity. *Mech Ageing Dev* **165**: 27–32.
241. Yamamoto M, Ohsawa S, Kunimasa K, Igaki T. 2017. The ligand Sas and its receptor PTP10D drive tumour-suppressive cell competition. *Nature* **542**: 246–250. http://dx.doi.org/10.1038/nature21033.
242. Yin Y, Dominska M, Yim E, Petes TD. 2017. High-resolution mapping of heteroduplex DNA formed during UV-induced and spontaneous mitotic recombination events in yeast. *Elife* **6**: 1–34.
243. Yizhak K, Aguet F, Kim J, Hess J, Kubler K, Grimsby J, Frazer R, Zhang H, Haradhvala N, Rosebrock D, et al. 2019. A comprehensive analysis of RNA sequences reveals macroscopic somatic clonal expansion across normal tissues. *Science (80-)* **364**: 416339. https://www.biorxiv.org/content/early/2018/09/13/416339.
244. Yokoyama A, Kakiuchi N, Yoshizato tetsuichi, Nannya Y, Suzuki H, takeuchi Y, Shiozawa Y, Sato Y, Aoki K, Ki Kim S, et al. 2019. Age-related remodelling of oesophageal epithelia by mutated cancer drivers. *Nature*. https://doi.org/10.1038/s41586-018-0811-x.
245. Yue F, Bi P, Wang C, Li J, Liu X, Kuang S. 2016. Conditional Loss of Pten in Myogenic Progenitors Leads to Postnatal Skeletal Muscle Hypertrophy but Age-Dependent Exhaustion of Satellite Cells. *Cell Rep* **17**: 2340–2353.

246. Zhang J, Walter JC. 2014. Mechanism and regulation of incisions during DNA interstrand cross-link repair. *DNA Repair (Amst)* **19**: 135–142.
247. Zhang W, Remenyik E, Zelterman D, Brash DE, Wikonkal NM. *Escaping the stem cell compartment: Sustained UVB exposure allows p53-mutant keratinocytes to colonize adjacent epidermal proliferating units without incurring additional mutations*. www.pnas.orgcgidoi10.1073pnas.241353198.
248. Zhang Y, Kalderon D. 2001. Hedgehog acts as a somatic stem cell factor in the Drosophila ovary. *Nature* **410**: 599–604. http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L32267206%0Ahttp://dx.doi.org/10.1038/35069099.
249. Zhu M, Lu T, Jia Y, Luo X, Gopal P, Li L, Odewole M, Renteria V, Singal AG, Jang Y, et al. 2019. Somatic Mutations Increase Hepatic Clonal Fitness and Regeneration in Chronic Liver Disease. *Cell* 1–14. https://linkinghub.elsevier.com/retrieve/pii/S0092867419302880.
250. Zink F, Stacey SN, Norddahl GL, Frigge ML, Magnusson OT, Jonsdottir I, Thorgeirsson TE, Sigurdsson A, Gudjonsson SA, Gudmundsson J, et al. 2017. Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. *Blood*.

Chapter 2 references

1. Al zouabi L, Bardin AJ. 2020. Stem Cell DNA Damage and Genome Mutation in the Context of Aging and Cancer Initiation. *Cold Spring Harb Perspect Biol* a036210.
2. Barbera MA, Petes TD. 2006. Selection and analysis of spontaneous reciprocal mitotic cross-overs in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **103**: 12819–12824.
3. Bardin AJ, Perdogoto CN, Southall TD, Brand AH, Schweisguth F. 2010. Transcriptional control of stem cell maintenance in the Drosophila intestine. *Development* **137**: 705–714.
4. Bayona-Feliu A, Casas-Lamesa A, Reina O, Bernués J, Azorín F. 2017. Linker histone H1 prevents R-loop accumulation and genome instability in heterochromatin. *Nat Commun* **8**. http://dx.doi.org/10.1038/s41467-017-00338-5.
5. Boulianne B, Robinson ME, May PC, Castellano L, Blighe K, Thomas J, Reid A, Müschen M, Apperley JF, Stebbing J, et al. 2017. Lineage-Specific Genes Are Prominent DNA Damage Hotspots during Leukemic Transformation of B Cell Precursors. *Cell Rep* **18**: 1687–1698.
6. Busque L, Patel JP, Figueroa ME, Vasanthakumar A, Provost S, Hamilou Z, Mollica L, Li J, Viale A, Heguy A, et al. 2012. Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis. *Nat Genet* **44**: 1179–1181. http://dx.doi.org/10.1038/ng.2413.
7. Cavenee WK, Dryja t TP, Phillips RA, Benedict WF, Godbout R, Gallie BL, Murphree AL, Strong LC, White RL. 1983. Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. *Nature* **305**: 9–14.
8. Coombs CC, Zehir A, Devlin SM, Kishtagari A, Syed A, Jonsson P, Hyman DM, Solit DB, Robson ME, Baselga J, et al. 2017. Therapy-Related Clonal Hematopoiesis in Patients with Non-hematologic Cancers Is Common and Associated with Adverse Clinical Outcomes. *Cell Stem Cell* **21**: 374-382.e4.
9. Cottrell S, Bodmer WF, Bicknell D, Kaklamanis L. 1992. Molecular analysis of APC mutations in familial adenomatous polyposis and sporadic colon carcinomas. *Lancet* **340**: 626–630.
10. Couto SS. 2011. The pathologist's slide reveals more than meets the eye: Loss of heterozygosity and cancer biology. *Vet Pathol* **48**: 236–244.
11. Crabtree M, Sieber OM, Lipton L, Hodgson S V., Lamlum H, Thomas HJW, Neale K, Phillips RKS, Heinimann K, Tomlinson IPM. 2003. Refining the relation between "first hits" and "second hits" at the APC locus: The "loose fit" model and evidence for differences in somatic mutation spectra among patients. *Oncogene* **22**: 4257–4265.
12. Crow JF pd., Kimura M. 1965. EVOLUTION IN SEXUAL AND ASEXUAL POPULATIONS * is advantageous because It has often been said that sexual reproduction of genotypes that can be produced by a recombina- number of the enormous. *Am Nat* **XCIX**: 439–450.
13. Deem A, Keszthelyi A, Blackgrove T, Vayl A, Coffey B, Mathur R, Chabes A, Malkova A. 2011. Break-induced replication is highly inaccurate. *PLoS*

Biol **9**: 1000594. /pmc/articles/PMC3039667/?report=abstract (Accessed August 25, 2020).
14. Ferguson M, Petkau K, Shin M, Galeneza A, Foley E. 2020. Symbiotic Lactobacillus brevis promote stem cell expansion and tumorigenesis in the Drosophila intestine. *bioRxiv* 1–45.
15. Genovese G, Kähler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF, Chambert K, Mick E, Neale BM, Fromer M, et al. 2014. Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence. *N Engl J Med* **371**: 2477–2487. http://www.nejm.org/doi/10.1056/NEJMoa1409405.
16. Gerlach SU, Herranz H. 2020. Genomic instability and cancer: lessons from Drosophila. *Open Biol* **10**: 200060.
17. Gervais L, Bardin AJ. 2017. Tissue homeostasis and aging: new insight from the fly intestine. *Curr Opin Cell Biol*.
18. Griffin R, Binari R, Perrimon N. 2014. Genetic odyssey to generate marked clones in Drosophila mosaics. *Proc Natl Acad Sci U S A* **111**: 4756–4763.
19. Haber JE. 2018. DNA Repair: The Search for Homology. *BioEssays* **40**: 1–12.
20. Haigis KM, Caya JG, Reichelderfer M, Dove WF. 2002. Intestinal adenomas can develop with a stable karyotype and stable microsatellites. *Proc Natl Acad Sci U S A* **99**: 8927–8931.
21. Howarth K, Ranta S, Winter E, Teixeira A, Schaschl H, Harvey JJ, Rowan A, Jones A, Spain S, Clark S, et al. 2009. A mitotic recombination map proximal to the APC locus on chromosome 5q and assessment of influences on colorectal cancer risk. *BMC Med Genet* **10**: 1–11.
22. Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman P V, Mar BG, Coleman Lindsley R, Mermel CH, Burtt N, Chavez A, et al. 2014. Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes NIH Public Access. *N Engl J Med* **371**: 2488–2498. http://cancer.sanger.ac.uk/cancergenome/projects/cosmic.
23. Jiang H, Patel PH, Kohlmaier A, Grenley MO, McEwen DG, Edgar BA. 2009. Cytokine/Jak/Stat Signaling Mediates Regeneration and Homeostasis in the Drosophila Midgut. *Cell* **137**: 1343–1355. http://dx.doi.org/10.1016/j.cell.2009.05.014.
24. Jin Y, Patel PH, Kohlmaier A, Pavlovic B, Zhang C, Edgar BA. 2017. Intestinal Stem Cell Pool Regulation in Drosophila. *Stem Cell Reports* **8**: 1479–1487. http://dx.doi.org/10.1016/j.stemcr.2017.04.002.
25. Knudson AG. 1971. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* **68**: 820–823.
26. Kohlmaier A, Fassnacht C, Jin Y, Reuter H, Begum J, Dutta D, Edgar B. 2015. Src kinase function controls progenitor cell pools during regeneration and tumor onset in the Drosophila intestine. *Oncogene* **34**: 2371–2384. www.nature.com/onc (Accessed June 24, 2020).
27. Lane ME, Sauer K, Wallace K, Jan YN, Lehner CF, Vaessin H. 1996. Dacapo, a cyclin-dependent kinase inhibitor, stops cell proliferation during Drosophila development. *Cell* **87**: 1225–1235. http://dx.doi.org/10.1016/S0092-8674(00)81818-8.
28. Langley CH, Stevens K, Cardeno C, Lee YCG, Schrider DR, Pool JE, Langley SA, Suarez C, Corbett-Detig RB, Kolaczkowski B, et al. 2012. Genomic variation in natural populations of Drosophila melanogaster.

Genetics **192**: 533–598.
29. Lapunzina P, Monk D. 2011. The consequences of uniparental disomy and copy number neutral loss-of-heterozygosity during human development and cancer. *Biol Cell* **103**: 303–317.
30. Lee PS, Petes TD. 2010. Mitotic gene conversion events induced in G1-synchronized yeast cells by gamma rays are similar to spontaneous conversion events. *Proc Natl Acad Sci U S A* **107**: 7383–7388.
31. Lucchetta EM, Ohlstein B. 2018. Amitosis of polyploid cells regenerates functional stem cells in the Drosophila intestine. *Cell Stem Cell* **20**: 609–620.
32. Mable BK, Otto SP. 2001. Masking and purging mutations following EMS treatment in haploid, diploid and tetraploid yeast (Saccharomyces cerevisiae). *Genet Res* **77**: 9–26.
33. Martincorena I, Fowler JC, Wabik A, Lawson ARJ, Abascal F, Hall MWJ, Cagan A, Murai K, Mahbubani K, Stratton MR, et al. 2018. Somatic mutant clones colonize the human esophagus with age. *Science*.
34. Martincorena I, McLaren S, Tubio JM, Campbell PJ, Van Loo P, Alexandrov LB, Jones PH, Fullam A, Roshan A, Widaa S, et al. 2015. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science (80-)* **348**: 880–886.
35. McKerrell T, Park N, Moreno T, Grove CS, Ponstingl H, Stephens J, Crawley C, Craig J, Scott MA, Hodkinson C, et al. 2015. Leukemia-Associated Somatic Mutations Drive Distinct Patterns of Age-Related Clonal Hemopoiesis. *Cell Rep* **10**: 1239–1245.
36. Micchelli CA, Perrimon N. 2006. Evidence that stem cells reside in the adult Drosophila midgut epithelium. *Nature* **439**: 475–479.
37. Ohlstein B, Spradling A. 2006. The adult Drosophila posterior midgut is maintained by pluripotent stem cells. *Nature* **439**: 470–474.
38. Otto SP, Goldstein DB. 1992. Recombination and the evolution of diploidy. *Genetics* **131**: 745–751.
39. Patel PH, Dutta D, Edgar BA. 2015. Niche appropriation by Drosophila intestinal stem cell tumours. *Nat Cell Biol*.
40. Perrot V, Richard S, Valero M. 1991. Transition from haploidy to diploidy. *Nature* **351**.
41. Riddiford N. 2020. Evolution and genomic signatures of spontaneous somatic mutation in Drosophila intestinal stem cells.
42. Ryland GL, Doyle MA, Goode D, Boyle SE, Choong DYH, Rowley SM, Li J, Ovarian A, Study C, Bowtell DDL, et al. 2015. Loss of heterozygosity : what is it good for ? *BMC Med Genomics* 1–12. http://dx.doi.org/10.1186/s12920-015-0123-z.
43. Saavedra P, Perrimon N. 2019. Drosophila as a model for tumor-induced organ wasting. *Adv Exp Med Biol* **1167**: 191–205.
44. Saeki H, Kitao H, Yoshinaga K, Nakanoko T, Kubo N, Kakeji Y, Morita M, Maehara Y. 2011. Copy-neutral loss of heterozygosity at the p53 locus in carcinogenesis of esophageal squamous cell carcinomas associated with p53 mutations. *Clin Cancer Res* **17**: 1731–1740.
45. Sakofsky CJ, Roberts SA, Malc E, Mieczkowski PA, Resnick MA, Gordenin DA, Malkova A. 2014. Break-induced replication is a source of mutation clusters underlying kataegis. *Cell Rep* **7**: 1640–1648.
46. Serra E, Rosenbaum T, Nadal M, Winner U, Ars E, Estivill X, Lázaro C.

2001. Erratum: Mitotic recombination effects homozygosity for NF1 germline mutations in neurofibromas (Nature Genetics (2001) 28 (294-296)). *Nat Genet* **29**: 100.
47. Shetzer Y, Kagan S, Koifman G, Sarig R, Kogan-Sakin I, Charni M, Kaufman T, Zapatka M, Molchadsky A, Rivlin N, et al. 2014. The onset of p53 loss of heterozygosity is differentially induced in various stem cell types and may involve the loss of either allele. *Cell Death Differ* **21**: 1419–1431.
48. Siudeja K, Bardin AJ. 2017. Somatic recombination in adult tissues: What is there to learn? *Fly (Austin)* **11**: 121–128. http://dx.doi.org/10.1080/19336934.2016.1249073.
49. Siudeja K, Nassari S, Gervais L, Skorski P, Lameiras S, Stolfa D, Zande M, Bernard V, Frio TR, Bardin AJ. 2015. Frequent Somatic Mutation in Adult Intestinal Stem Cells Drives Neoplasia and Genetic Mosaicism during Aging. *Cell Stem Cell* **17**: 663–674. http://dx.doi.org/10.1016/j.stem.2015.09.016.
50. Siudeja K, van den Beek M, Riddiford N, Boumard B, Wurmser A, Stefanutti M, Lameiras S, Bardin AJ. 2020. Somatic transposition in the Drosophila intestine. *bioRxiv* 1–38.
51. Spradling AC. 2017. Polytene Chromosome Structure and Somatic Genome Instability. *Cold Spring Harb Symp Quant Biol* **82**: 293–304.
52. St. Charles J, Petes TD. 2013. High-Resolution Mapping of Spontaneous Mitotic Recombination Hotspots on the 1.1 Mb Arm of Yeast Chromosome IV. *PLoS Genet* **9**.
53. Stern C. 1936. *SOMATIC CROSSING OVER AND SEGREGATION I N DROSOPHILA MELANOGASTER* THE ACTION OF hflNUTE FACTORS Blond-Minute Autosomal Minutes and sex-linked spots The specificity of the effects of sex-linked and autosomal Minutes*.
54. Thiagalingam S, Laken S, Willson JKV, Markowitz SD, Kinzler KW, Vogelstein B, Lengauer C. 2001. Mechanisms underlying losses of heterozygosity in human colorectal cancers. *Proc Natl Acad Sci U S A* **98**: 2698–2702.
55. Tomasetti C, Vogelstein B. 2015. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science (80-)*.
56. Tuna M, Knuutila S, Mills GB. 2009. Uniparental disomy in cancer. *Trends Mol Med* **15**: 120–128.
57. Villegas SN. 2019. One hundred years of Drosophila cancer research: No longer in solitude. *DMM Dis Model Mech* **12**.
58. Wang F, Jiang L, Chen Y, Haelterman NA, Bellen HJ, Chen R. 2015. FlyVar: A database for genetic variation in Drosophila melanogaster. *Database* **2015**: 1–7.
59. Wang L. 2018. Loss of Tumor Suppressor Gene Function in Human Cancer : An Overview. 2647–2693.
60. Xu W, Mulligan L, Ponder M. 1992. Loss of NF1 alleles in phaeochromocytomas from patients with type 1 neurofibromatosis. *Genes Chromosom Cancer* **4**: 337–342.
61. Yarosh W, Spradling AC. 2014. Incomplete replication generates somatic DNA alterations within Drosophila polytene salivary gland cells. 1840–1855.
62. Yin Y, Dominska M, Yim E, Petes TD. 2017. High-resolution mapping of heteroduplex DNA formed during UV-induced and spontaneous mitotic recombination events in yeast. *Elife* **6**: 1–34.

63. Yokoyama A, Kakiuchi N, Yoshizato tetsuichi, Nannya Y, Suzuki H, takeuchi Y, Shiozawa Y, Sato Y, Aoki K, Ki Kim S, et al. 2019. Age-related remodelling of oesophageal epithelia by mutated cancer drivers. *Nature*. https://doi.org/10.1038/s41586-018-0811-x.
64. Zhou J, Boutros M. 2020. JNK-dependent intestinal barrier failure disrupts host-microbe homeostasis during tumorigenesis. *Proc Natl Acad Sci U S A* **117**: 9401–9412.
65. Zink F, Stacey SN, Norddahl GL, Frigge ML, Magnusson OT, Jonsdottir I, Thorgeirsson TE, Sigurdsson A, Gudjonsson SA, Gudmundsson J, et al. 2017. Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. *Blood*.

Chapter 3 references

1. Al-Hadid Q, Yang Y. 2016. R-loop: An emerging regulator of chromatin dynamics. *Acta Biochim Biophys Sin (Shanghai)* **48**: 623–631.
2. Alecki C, Chiwara V, Sanz LA, Grau D, Pérez OA, Armache K, Chédin F, Francis NJ. 2020. RNA-DNA strand exchange by the Drosophila Polycomb complex PRC2. *Nat Commun* 1–14. http://dx.doi.org/10.1038/s41467-020-15609-x.
3. Amon JD, Koshland D, States U. 2016. RNase H enables efficient repair of R-loop induced DNA damage. *Elife* 1–20.
4. Bayona-Feliu A, Casas-Lamesa A, Reina O, Bernués J, Azorín F. 2017. Linker histone H1 prevents R-loop accumulation and genome instability in heterochromatin. *Nat Commun* **8**. http://dx.doi.org/10.1038/s41467-017-00338-5.
5. Boulianne B, Robinson ME, May PC, Castellano L, Blighe K, Thomas J, Reid A, Müschen M, Apperley JF, Stebbing J, et al. 2017. Lineage-Specific Genes Are Prominent DNA Damage Hotspots during Leukemic Transformation of B Cell Precursors. *Cell Rep* **18**: 1687–1698.
6. Boyle M, Wong C, Rocha M, Jones DL. 2007. Decline in Self-Renewal Factors Contributes to Aging of the Stem Cell Niche in the Drosophila Testis. *Cell Stem Cell* **1**: 470–478.
7. Brand AH, Perrimon N. 1993. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**: 401–415.
8. Cerritelli SM, Crouch RJ. 2009. Ribonuclease H: The enzymes in eukaryotes. *FEBS J* **276**: 1494–1505.
9. Chappidi N, Nascakova Z, Boleslavska B, Zellweger R, Isik E, Andrs M, Menon S, Dobrovolna J, Balbo Pogliano C, Matos J, et al. 2020. Fork Cleavage-Religation Cycle and Active Transcription Mediate Replication Restart after Fork Stalling at Co-transcriptional R-Loops. *Mol Cell* **77**: 528-541.e8.
10. Chedin F, Benham CJ, Musier-Forsyth K. 2020. Emerging roles for R-loop structures in the management of topological stress. *J Biol Chem* **295**: 4684–4695.
11. Chernikova SB, Game JC, Brown JM. 2012. Inhibiting homologous recombination for cancer therapy. *Cancer Biol Ther* **13**: 61–68.
12. Choate KA, Lu Y, Zhou J, Choi M, Elias PM, Farhi A, Nelson-Williams C, Crumrine D, Williams ML, Nopper AJ, et al. 2010. Mitotic Recombination in Patients with Ichthyosis Causes Reversion of Dominant Mutations in KRT10. *Science (80-)* **60**: 94–97.
13. Depinho RA. 2000. The age of cancer. **408**.
14. Erceg J, Abed JA, Goloborodko A, Lajoie BR, Fudenberg G, Abdennur N, Imakaev M, Mccole RB, Nguyen SC, Saylor W, et al. 2019. Chromosome pairing in early Drosophila embryos. *Nat Commun*. http://dx.doi.org/10.1038/s41467-019-12211-8.
15. Gavande NS, Vandervere-carozza PS, Hinshaw HD, Jalal SI, Sears CR, Pawelczak KS, Turchi JJ. 2016. DNA repair targeted therapy : The past or

future of cancer treatment ? *Pharmacol Ther* **160**: 65–83.
16. Hamperl S, Bocek MJ, Saldivar JC, Swigut T, Cimprich KA. 2017. Transcription-Replication Conflict Orientation Modulates R-Loop Levels and Activates Distinct DNA Damage Responses. *Cell* **170**: 774-786.e19. http://dx.doi.org/10.1016/j.cell.2017.07.043.
17. Hartono SR, Malapert A, Pénélope L, Bernard P, Chédin F, Vanoosthuyse V. 2018. The Affinity of the S9.6 Antibody for Double-Stranded RNAs Impacts the Accurate Mapping of R-Loops in Fission Yeast. *J Mol Biol* **3**.
18. Howarth K, Ranta S, Winter E, Teixeira A, Schaschl H, Harvey JJ, Rowan A, Jones A, Spain S, Clark S, et al. 2009. A mitotic recombination map proximal to the APC locus on chromosome 5q and assessment of influences on colorectal cancer risk. *BMC Med Genet* **10**: 1–11.
19. Hudry B, Khadayate S, Miguel-Aliaga I. 2016. The sexual identity of adult intestinal stem cells controls organ size and plasticity. *Nature* **530**.
20. Jongmans MCJ, Diets I, Quarello P, Garelli E, Kuiper R, Pfundt R. 2018. Somatic reversion events point towards RPL4 as a novel disease gene in a condition resembling Diamond-Blackfan anemia. *Haematologica* **103**: 607–609.
21. Kelley M, Logsdon D, Fishel M. 2014. Targeting DNA repair pathways for cancer treatment: what's new? *Futur Oncol* **10**: 1215–1237.
22. Kuzminov A. 2018. When DNA topology turns deadly: *Trends Genet* **34**: 139–148.
23. LaFave MC, Andersen SL, Stoffregen EP, Holsclaw JK, Kohl KP, Overton LJ, Sekelsky J. 2014. Sources and structures of mitotic crossovers that arise when BLM helicase is absent in Drosophila. *Genetics* **196**: 107–118.
24. Lange SS, Takata K, Wood RD. 2011. DNA polymerases and cancer. *Nat Rev Cancer* **11**: 96–110.
25. Lee PS, Petes TD. 2010. Mitotic gene conversion events induced in G1-synchronized yeast cells by gamma rays are similar to spontaneous conversion events. *Proc Natl Acad Sci U S A* **107**: 7383–7388.
26. Lee T, Luo L. 2001. Mosaic analysis with a repressible cell marker (MARCM) for Drosophila neural development (Trend in Neurosciences (May 2001)). *Trends Neurosci* **24**: 385.
27. Lucchetta EM, Ohlstein B. 2018. Amitosis of polyploid cells regenerates functional stem cells in the Drosophila intestine. *Cell Stem Cell* **20**: 609–620.
28. Malig M, Hartono SR, Giafaglione JM, Sanz LA, Chedin F. 2020. Ultra-Deep Coverage Single-Molecule R-loop Footprinting Reveals Principles of R-loop Formation. *J Mol Biol*. http://search.ebscohost.com/login.aspx?direct=true&db=edselp&AN=S0022283620301649&site=eds-live.
29. Mateos-Gomez PA, Gong F, Nair N, Miller KM, Lazzerini-Denchi E, Sfeir A. 2015. Mammalian polymerase θ promotes alternative NHEJ and suppresses recombination. *Nature* **518**: 254–257. http://dx.doi.org/10.1038/nature14157.
30. Matunis EL, Stine RR, de Cuevas M. 2012. Recent advances in Drosophila male germline stem cell biology. *Spermatogenesis* **2**: 137–144.
31. O'connell K, Jinks-Robertson S, Petes TD. 2015. Elevated Genome-Wide Instability in Yeast Mutants Lacking RNase H Activity. *Genetics* **201**: 963–975.

32. Pan Y, Feng C, Wang H, Lee M, Tang Z, Lin Z. 2020. Ichthyosis with confetti caused by new and recurrent mutations in KRT10 associated with varying degrees of keratin 10 mis-localization. *J Dermatol Sci* **98**: 35–40. https://doi.org/10.1016/j.jdermsci.2020.02.005.
33. Patel PH, Dutta D, Edgar BA. 2015. Niche appropriation by Drosophila intestinal stem cell tumours. *Nat Cell Biol*.
34. Polleys EJ, House NCM, Freudenreich CH. 2017. Role of recombination and replication fork restart in repeat instability. **56**: 156–165.
35. Pommier Y. 2006. Topoisomerase I inhibitors : camptothecins and beyond. *Nat Rev* **6**: 789–802.
36. Preston CR, Flores C, Engels WR. 2006. Report Age-Dependent Usage of Double-Strand-Break Repair Pathways. *Curr Biol* 2009–2015.
37. Reeg S, Grune T. 2015. Protein Oxidation in Aging : Does It Play a Role in Aging Progression ? *Antioxidants Redox Signal* **23**: 239–255.
38. Regan JC, Khericha M, Dobson AJ, Bolukbasi E, Rattanavirotkul N, Partridge L. 2016. Sex difference in pathology of the ageing gut mediates the greater response of female lifespan to dietary restriction. *Elife*.
39. Reiff T, Jacobson J, Cognigni P, Antonello Z, Ballesta E, Tan KJ, Yew JY, Dominguez M, Miguel-aliaga I, Miguel U. 2015. Endocrine remodelling of the adult intestine sustains reproduction in Drosophila. 1–19.
40. Rera M, Clark RI, Walker DW. 2012. Intestinal barrier dysfunction links metabolic and inflammatory markers of aging to death in Drosophila. *Proc Natl Acad Sci U S A* **109**: 21528–21533.
41. Rondón AG, Aguilera A. 2019. What causes an RNA-DNA hybrid to compromise genome integrity? *DNA Repair (Amst)* **81**: 102660. https://doi.org/10.1016/j.dnarep.2019.102660.
42. Roy D, Lieber MR. 2009. G Clustering Is Important for the Initiation of Transcription-Induced R-Loops In Vitro, whereas High G Density without Clustering Is Sufficient Thereafter. *Mol Cell Biol* **29**: 3124–3133.
43. Roy D, Zhang Z, Lu Z, Hsieh C-L, Lieber MR. 2010. Competition between the RNA Transcript and the Nontemplate DNA Strand during R-Loop Formation In Vitro: a Nick Can Serve as a Strong R-Loop Initiation Site. *Mol Cell Biol* **30**: 146–159.
44. Shrivastav M, Haro LP De, Nickoloff JA. 2008. Regulation of DNA double-strand break repair pathway choice. **18**: 134–147.
45. Smolka JA, Sanz LA, Hartono SR, Chédin F. 2020. Recognition of cellular RNAs by the S9.6 antibody creates pervasive artefacts when imaging RNA:DNA hybrids. *bioRxiv*.
46. St. Charles J, Petes TD. 2013. High-Resolution Mapping of Spontaneous Mitotic Recombination Hotspots on the 1.1 Mb Arm of Yeast Chromosome IV. *PLoS Genet* **9**.
47. Stack S, Brown W. 1969. Somatic pairing, reduction and recombination: An evolutionary hypothesis of meiosis. *Nature* **224**: 177–178. http://adsabs.harvard.edu/abs/1969Natur.224..177K.
48. Stadtman ER. 1992. Protein Oxidation and Aging. *Science (80-)* **257**.
49. Statistics O for N. 2019. Cancer Registration Statistics, England.
50. Stein H, Hausen P. 1969. Enzyme from Calf Thymus Degrading the RNA Moiety of DNA-RNA Hybrids: Effect on DNA-Dependent RNA Polymerase. *Science (80-)* 393–395.
51. Stern C. 1936. *SOMATIC CROSSING OVER AND SEGREGATION I N*

DROSOPHILA MELANOGASTER THE ACTION OF hflNUTE FACTORS Blond-Minute Autosomal Minutes and sex-linked spots The specificity of the effects of sex-linked and autosomal Minutes*.

52. Suijkerbuijk SJE, Kolahgar G, Kucinski I, Piddini E. 2016. Cell competition drives the growth of intestinal adenomas in Drosophila. *Curr Biol*.
53. Thomas M, White RL, Davis RW. 1976. Hybridization of RNA to double stranded DNA: Formation of R loops. *Proc Natl Acad Sci U S A* **73**: 2294–2298.
54. Toriumi K, Tsukahara T, Hanai R. 2013. R-loop formation in trans at an AGGAG repeat. *J Nucleic Acids* **2013**.
55. Vanoosthuyse V. 2018. Strengths and weaknesses of the current strategies to map and characterize R-loops. *Non-coding RNA* **4**.
56. Venugopal P, et al. 2017. Self-reverting mutations partially correct the blood phenotype in a Diamonf Blackfan anemia patient. *Haematologica* **102**: 117–120.
57. Wahba L, Costantino L, Tan FJ, Zimmer A, Koshland D. 2016. S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. *Genes Dev* 1327–1338.
58. White RR, Vijg J. 2016. Do DNA double-strand breaks drive aging? *Mol Cell* **63**: 729–738.
59. Wong AC. 2019. New Approaches to Microbiome-Based Therapies. *Host-microbe Biol* 1–5.
60. Yan Q, Sarma K. 2020. MapR : A Method for Identifying Native R-Loops Genome Wide. *Curr Protoc Mol Biol* **130**.
61. Zielke N, Korzelius J, vanStraaten M, Bender K, Schuhknecht GFP, Dutta D, Xiang J, Edgar BA. 2014. Fly-FUCCI: A Versatile Tool for Studying Cell Proliferation in Complex Tissues. *Cell Rep* **7**: 588–598. http://dx.doi.org/10.1016/j.celrep.2014.03.020.
62. Zink F, Stacey SN, Norddahl GL, Frigge ML, Magnusson OT, Jonsdottir I, Thorgeirsson TE, Sigurdsson A, Gudjonsson SA, Gudmundsson J, et al. 2017. Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. *Blood* **130**: 742–752.

Résumé substantiel (en français) des chapitres du mémoire de thèse

La thèse est divisée en 3 chapitres, comportant également une section « matériels et méthodes » ainsi qu'une partie « bibliographie ».

_Le **chapitre 1** est une introduction à la thèse divisée en 3 sections couvrant trois thèmes majeurs :

-1.1. L'importance de mieux comprendre les génomes altérés des cellules souches.

-1.2. Les lacunes de nos connaissances vis-à-vis des mécanismes conduisant aux altérations du génome.

-1.3. L'intestin de la *Drosophile* utilisé comme modèle pour aborder les questions concernant les mécanismes d'altération du génome dans les cellules souches.

Le chapitre 1.1 commence par un aperçu général des lésions de l'ADN dans les cellules souches et de la façon dont elles peuvent provoquer une altération du génome perturbant donc leur auto-renouvellement et leur capacité à se différencier. Cela entraîne des changements dans le statu quo des différents tissus, avec pour conséquence un impact sur le vieillissement et l'initiation de cancer. Il existe des mécanismes spécifiques aux cellules et tissus par lesquels les cellules souches se protègent des dommages et des mutations. Malgré ces mécanismes de protection, les dommages et l'acquisition de mutations se produisent toujours. Des questions fondamentales demeurent quant à la façon dont les dommages à l'ADN et les mutations somatiques des cellules souches peuvent être manipulés pour ralentir le vieillissement et retarder ou même éviter l'apparition de cancer. Comment empêcher l'accumulation dans les cellules souches de dommages génomiques ? Peut-on exploiter les mécanismes de compétition cellulaire pour remplacer des cellules mutantes potentiellement dangereuses par des cellules thérapeutiques ? Les facteurs extrinsèques aux tissus, tels que les changements environnementaux, peuvent-ils être manipulés pour contrôler les expansions clonales ? Ces questions, soulevées dans ce chapitre, restent ouvertes aujourd'hui et seront des domaines de recherche manifestes bénéficiant d'études sur divers systèmes de modèles génétiques.

Dans le chapitre 1.2, il est plutôt question d'une cause courante d'altération du génome : la perte d'hétérozygotie (LOH), sa responsabilité face à la maladie et notamment au déclenchement de cancer. En général, au niveau de l'organisme, un certain degré de

protection est assuré par la "diploïdie", qui permet de disposer de deux copies de chaque gène pour se protéger des effets d'une mutation somatique. Si une mutation se produit dans une des deux copie, la deuxième copie " Wild-Type " fournit une sauvegarde et maintient la fonction. Ainsi, l'état hétérozygote masque les effets des mutations délétères récessives. J'explique ici la perte de l'état hétérozygote protecteur, pouvant conduire à des cancers ou troubles pathologiques et qui se produit également dans les tissus humains normaux. En particulier, je détaille ce que l'on sait des mécanismes pouvant conduire à la perte d'hétérozygotie (LOH) en me concentrant principalement sur la "recombinaison mitotique"/ "mitotic recombination" (MR en anglais) comme mécanisme et en mentionnant également l'aneuploïdie comme autre mécanisme. La Recombinaison Mitotique (MR) est un mécanisme de réparation de l'ADN, également connu sous le nom de recombinaison homologue, il permet l'utilisation d'un chromosome homologue comme matrice dans le but de réparer une cassure double brin sur un chromosome. Bien que l'étude de la levure ait mis en lumière les caractéristiques des séquences et les différents mécanismes qui régissent la recombinaison mitotique, la question demeure de savoir s'ils sont conservés et fonctionnent dans les cellules souches adultes qui entretiennent l'homéostasie des tissus chez les eucaryotes supérieurs. En outre, bien que des efforts aient été déployés pour comprendre comment la MR apparaît dans les tissus somatiques, elle n'a été observé, pour la plupart, qu'à l'aide de techniques à faible résolution chez les eucaryotes supérieurs. De plus, les questions liées à la façon dont les facteurs environnementaux pourraient contribuer à la LOH induite par la MR serait une perspective précieuse dans le domaine. Il est donc nécessaire de combler les lacunes de ces connaissances en conjuguant les résultats obtenus chez la levure unicellulaire et ceux obtenus chez les eucaryotes supérieurs afin de comprendre les voies qui favorisent et empêchent la MR dans un tissu complexe in vivo. L'utilisation d'un système in vivo dans lequel nous pouvons tester des hypothèses et voir l'impact des variables changeantes sera important pour élucider la MR depuis l'initiation génomique jusqu'aux moteurs tissulaires.

En outre, j'ai présenté comment la LOH peut être induite par un autre mécanisme appelé "aneuploïdie" étant la perte d'un chromosome entier. J'ai souligné que si l'aneuploïdie devrait, en théorie, entraver la prolifération cellulaire, elle favorise paradoxalement la progression des tumeurs. Cela soulève des questions concernant les mécanismes compensatoires qui peuvent se mettre en place pour modifier le déséquilibre du matériel génétique et rétablir l'homéostasie protéomique. Ainsi, un système est nécessaire pour tester les aneuploïdies spontanées et les potentiels mécanismes tampons qui pourraient doter la cellule d'une tolérance face à un complément cellulaire anormal.

Enfin, au chapitre 1.3, je présente le système modèle de la *Drosophile* que j'emploie pour répondre à d'importantes questions portant sur la perte d'hétérozygotie dans les cellules souches. J'aborde brièvement la structure de l'intestin de la *Drosophile* adulte, ses types de cellules, la communication entre celles-ci qui spécifie son lignage cellulaire. Je décris ensuite certains changements comme le vieillissement ainsi que facteurs extrinsèques qui pouvant l'influencer, comme l'environnement extérieur. Enfin, je termine cette section en soulignant les avantages de l'utilisation de ce système modèle en décrivant brièvement les outils génétiques disponibles qui font de ce dernier un modèle génétique puissant pour répondre aux questions des points 1.1 et 1.2. De façon succincte, concernant ce modèle nous pourrions résumer ceci : l'intestin de la *Drosophile* est un tissu régénératif maintenu par environ 1000 cellules souches intestinales multipotentes étant sujettes à de fréquentes mutations spontanées. Les mutations spontanées se traduisent par des néoplasies phénotypiquement visibles qui sont détectées chez les mouches âgées. Le laboratoire a montré qu'il s'agit de clones mutants apparaissant spontanément à partir d'une cellule souche mutante et que ce système récapitule les expansions clonales des tissus humains (Siudeja et al. 2015). Globalement, l'intestin de la *Drosophile* peut être utilisé pour déterminer comment les mutations somatiques surviennent dans les cellules souches adultes. La petite taille du génome de la *Drosophile* se trouve être un autre avantage, permettant un séquençage plus rentable, une durée de vie significativement courte de 6 semaines, rendant les expériences de vieillissement chez les mouches plus rapides que les expériences de vieillissement chez la souris. Plus important encore, ce qui fait de la *Drosophile* un modèle puissant en général, est son adaptabilité génétique à des outils disponibles pour la manipulation in vivo et la possibilité de modifier les conditions environnementales. Malgré la divergence physiologique entre la *Drosophile* et les vertébrés, la modélisation des maladies intestinales humaines est possible en raison du haut degré de conservation des voies de signalisation entre ces deux modèles.

_Le **chapitre 2** est le chapitre couvrant mes résultats en deux sections. La première section (chapitre 2.1) est mon article en préparation, dont je suis la première autrice, traitant de la recombinaison mitotique en tant que mécanisme à même de conduire la LOH spontanée dans les cellules souches intestinales de la *Drosophile*. La deuxième section (chapitre 2.2) présente le deuxième mécanisme, l'aneuploïdie, par lequel la LOH peut se produire.

Résumé de l'article **:**

Les cellules somatiques peuvent subir une altération du génome entraînant une perte d'hétérozygotie (LOH). Ce phénomène se produit dans les tissus humains normaux,

les troubles pathologiques ainsi que les cancers. Bien que de précédentes études sur la levure aient permis de mieux comprendre les différents mécanismes de la perte d'hétérozygotie, les détails mécanistiques eux font défauts concernant les organismes multicellulaires aux tissus complexes. A l'aide du modèle vivant de la *Drosophile* et plus précisément des cellules souches in vivo, nous étudions ici les mécanismes qui initient la perte d'hétérozygotie, comblant par la même occasion le fossé existant entre la levure unicellulaire et les eucaryotes supérieurs. Grâce à des données de séquençage génomique des événements somatiques de la LOH, du profilage des modifications du nombre de copies et des modifications de l'hétérozygotie des polymorphismes d'un seul nucléotide, nous avons démontré que la LOH se produit principalement par recombinaison mitotique et bien plus rarement par aneuploïdie.

En conséquence, nous avons trouvé une implication de la protéine de réparation de l'ADN, Rad51, dans la recombinaison mitotique de la LOH. La cartographie fine des sites de recombinaison n'a pas révélé les accumulations de mutations qui surviennent couramment avec un mécanisme de réplication induit par une rupture, mais a plutôt montré des exemples clairs d'enjambements de chromosomes issus d'événements croisés générés par une réparation basée sur une double jonction de Holliday. Les régions de recombinaison cartographiées ont également permis de mieux comprendre les caractéristiques potentielles des séquences génomiques susceptibles de favoriser la recombinaison mitotique, notamment une association avec la région répétée Histone Locus Cluster et les régions précédemment cartographiées pour former des « R-loops ». Nous avons également étudié comment les facteurs environnementaux peuvent influencer ce processus et démontré que l'infection par la bactérie entéropathogène Ecc15 augmentait la fréquence de perte d'hétérozygotie. Cette étude permet de mieux comprendre mécaniquement comment la recombinaison mitotique se produit dans les cellules souches in vivo, elle identifie des facteurs intrinsèques et extrinsèques qui peuvent entraîner la LOH, tout ceci fournissant un support important d'informations quant à l'initiation du cancer et des potentielles stratégies préventives et thérapeutiques.

Le chapitre 2.2 souligne qu'en plus de la recombinaison mitotique, nous avons trouvé des preuves d'un autre mécanisme de LOH, l'aneuploïdie. Nous avons établi un système pour mieux comprendre les aneuploïdies spontanées donnant lieu à une LOH dans les cellules souches intestinales. Cela nous permettra d'évaluer les mécanismes tampons potentiels qui pourraient doter la cellule d'une tolérance à un complément cellulaire anormal.

_Le **chapitre 3** est un chapitre de discussion qui commence par une critique de mes expériences, soulignant certaines des réserves techniques et expérimentales tout en proposant des suggestions d'amélioration. Le chapitre se poursuit par une discussion de mes résultats à la lumière des résultats précédents afin de les placer dans le contexte de ce qui est connu. Enfin, je termine en donnant quelques perspectives sur l'orientation que pourraient prendre les recherches futures sur la LOH dans les cellules souches et sur les implications qu'elles pourraient avoir en clinique. Les résultats scientifiquement exploitables issus de mon projet de doctorat se résument en la mise au point du tout premier système permettant d'étudier systématiquement la LOH à l'aide d'une technique à haute résolution chez un animal vivant entraînant une croissance néoplasique. Notre système modèle de cellules souches in vivo de l'intestin de la *Drosophile* fourni un lien important dans la conservation évolutive des mécanismes de MR entre la levure unicellulaire et les tissus complexes des eucaryotes supérieurs (dans lesquels l'étude des mécanismes moléculaires de la LOH s'est avérée techniquement difficile). Nous avons également identifié des facteurs intrinsèques et extrinsèques à la cellule susceptibles d'entraîner la LOH et apporter ainsi d'importantes informations quant à l'initiation du cancer, informations qui pourraient avoir une implication d'un point de vue thérapeutique dans des études encore plus approfondies.

Les personnes particulièrement vulnérables aux cancers induits par la LOH, comme celles qui sont nées avec une mutation germinale dans un gène suppresseur de tumeur, peuvent hypothétiquement bénéficier d'un traitement favorisant une voie alternative de réparation des cassures double brin. Nos recherches apportent également une plus grande compréhension de la relation entre les bactéries pathogènes et le LOH. Les bactéries qui composent le microbiote ont la capacité d'être des initiateurs importants à de nombreux cancers et sont mal comprises. Grâce aux progrès des thérapies du microbiome actuellement en cours de développement, les patients à risque peuvent bénéficier de la neutralisation et de la correction des effets négatifs de la dysbiose.

La thèse se clôture sur une section traitant l'ensemble des matériels et méthodes utilisés dans l'étude, mettant en évidence les stocks de *Drosophiles* utilisés, les techniques de vieillissement des expériences, les explications des expériences génétiques, les traitements bactériens et autres, l'immunofluorescence, la quantification des expériences, la méthode de séquençage et l'analyse bio-informatique. Enfin, la bibliographie figure à la fin de la thèse et est divisée en références pour les chapitres 1, 2 et 3 séparément.

RÉSUMÉ

Les cellules somatiques peuvent subir une altération du génome entraînant une perte d'hétérozygotie (LOH). Ce phénomène se produit dans les tissus humains normaux, les troubles pathologiques et les cancers. Bien que de précédentes études sur la levure aient permis de mieux comprendre les différents mécanismes de la perte d'hétérozygotie, les détails mécanistiques eux font défaut concernant les organismes multicellulaires aux tissus complexes. A l'aide du modèle vivant de la *Drosophile* et plus précisément des cellules souches *in vivo*, nous étudions ici les mécanismes qui initient la perte d'hétérozygotie, comblant par la même occasion le fossé existant entre la levure unicellulaire et les eucaryotes supérieurs. Grâce à des données de séquençage génomique des événements somatiques de la LOH, du profilage des modifications du nombre de copies et des modifications de l'hétérozygotie des polymorphismes d'un seul nucléotide, nous avons démontré que la LOH se produit principalement par recombinaison mitotique et bien plus rarement par aneuploïdie.

En conséquence, nous avons trouvé une implication de l'enzyme de réparation de l'ADN, Rad51, dans la recombinaison mitotique de la LOH. La cartographie fine des sites de recombinaison n'a pas révélé les accumulations de mutations qui surviennent couramment avec un mécanisme de réplication induit par une rupture, mais a plutôt montré des exemples clairs d'enjambements de chromosomes issus d'événements croisés générés par une réparation basée sur une double jonction de Holliday. Les régions de recombinaison cartographiées ont également permis de mieux comprendre les caractéristiques potentielles des séquences génomiques susceptibles de favoriser la recombinaison mitotique, notamment une association avec la région répétée *Histone Locus Cluster* et les régions précédemment cartographiées pour former des « R-loops ». Nous avons également étudié comment les facteurs environnementaux peuvent influencer ce processus et démontré que l'infection par la bactérie entéropathogène Ecc15 augmentait la fréquence de perte d'hétérozygotie. Cette étude permet de mieux comprendre mécaniquement comment la recombinaison mitotique se produit dans les cellules souches *in vivo*, elle identifie des facteurs intrinsèques et extrinsèques qui peuvent entraîner la LOH, tout ceci fournissant un support important d'informations quant à l'initiation du cancer et des stratégies préventives et thérapeutiques potentielles.

MOTS CLÉS

Perte d'hétérozygotie, recombinaison mitotique, aneuploïdie, cellules souches intestinales, homéostasie tissulaire, vieillissement, cancer, séquençage, *Drosophile*, *in vivo,* réparation de l'ADN.

ABSTRACT

Somatic cells can undergo a genome alteration leading to loss of heterozygosity (LOH). This phenomenon occurs in normal human tissues, pathological disorders, and cancers. Although previous studies in yeast have provided a substantial insight into different mechanisms of LOH, mechanistic details are lacking in multicellular organisms with complex tissues. Here we investigate the mechanisms giving rise to LOH, bridging the gap between unicellular yeast and higher eukaryotes using an *in vivo* stem cell model system in *Drosophila*. Through whole-genome sequencing of somatic LOH events, profiling copy number changes and changes in heterozygosity of single-nucleotide polymorphisms, we demonstrated that LOH arises primarily via mitotic recombination and more rarely through aneuploidy.

Consistent with this, we found involvement of the DNA repair enzyme Rad51 in mitotic recombination-driven LOH. Fine mapping of recombination sites did not reveal mutational pile-ups that commonly arise with a break-induced replication mechanism and instead showed clear examples of chromosomes arising from cross-over events generated by double-Holliday junction-based repair. The mapped recombination regions also provided insight into potential genomic sequence features that may promote mitotic recombination, including an association with the repeated region of the *Histone Locus Cluster* and regions previously mapped to form R loops. We further explored how environmental factors can influence this process and demonstrate that infection with the enteric pathogenic bacteria, *Ecc15*, increased LOH frequency. This study provides a better mechanistic understanding of how mitotic recombination arises in stem cells *in vivo,* and identifies intrinsic and extrinsic factors that can drive LOH, thus providing important insight into cancer initiation and potential preventative and therapeutic strategies.

KEYWORDS

Loss of heterozygosity, mitotic recombination, aneuploidy, intestinal stem cells, tissue homeostasis, aging, cancer, whole-genome sequencing, *Drosophila,* in vivo, DNA repair.