Addendum

Cover Photo: Iron G4 ▶

THESIS SUMMARY

Microsatellites and G-quadruplex motifs are DNA sequences that are prevalent throughout the genome and are linked to diseases such as cancer and neurodegenerative disorders. Microsatellites can be defined as short tandem repeats with units of 1-8 base pairs (bp) long. G-quadruplex motifs (also known as G4 DNA) can be defined as sequences that contain four tracts of three or more guanines interspaced by at least one random nucleotide. Although DNA is organized as a double helical structure, G-quadruplex motifs have the unique capacity to form a four-stranded DNA structure (named a G-quadruplex structure) in which the guanines interact with each other through Hoogsteen base pairing. Microsatellites and G-quadruplexes motifs are sequences that are intrinsically difficult to replicate because of their repetitive nature and capacity to form alternative DNA structures. As a result, they can trigger mutations upon replication. Microsatellites and G-quadruplexes play a significant role in the initiation of the aforementioned devastating diseases. However, many aspects about G-quadruplex and microsatellite instability (MSI) are incompletely understood. For example, what determines the degree of their instability, since some microsatellites and G-quadruplexes are more mutagenic than others? Which genes and pathways prevent microsatellite and G-quadruplex instability? What are the direct genetic consequences, and which molecular mechanisms act to produce genomic changes at these sequences? Answers to these questions will be of great importance in the development of new and better treatments of microsatellite- and G-quadruplex-related diseases. In this thesis I provide new insights into the biology behind microsatellite and G-quadruplex instability, by making use of various model organisms and newly developed molecular tools.

Using an MSI-reporter system based on fluorescence and optimized for mammalian cells, I demonstrate in **Chapter 2** that MSI is greatly influenced by the tract length and the nucleotide composition of the microsatellite. Other factors, such as its strand orientation or its transcriptional status, appeared not to affect the stability of a microsatellite. Furthermore, I show that the MSI-reporter system can be a useful tool to screen for MSI-inducing compounds, as well as to screen for genes that protect the genome against MSI. By testing a library of ±450 different miRNAs (small RNA-molecules that influence gene expression), I show that overexpression of one of these miRNAs, named miR-21, results in increased MSI. By additional experiments I show that miR-21 targets and thereby reduces the expression of the Lynch syndrome- and MSI-associated gene MSH2, explaining the increased levels of MSI observed in miR-21 overexpressing cells.

In **Chapter 3** I present a new genetic tool that enables us to trace single cells and also to study tumor development in living zebrafish (). I show that genes can be stochastically activated by placing their coding sequence downstream of a microsatellite. The gene of interest is placed such that only after a mutation in the microsatellite, a so-called frameshift, the gene becomes expressed. Low frequency frameshifting stochastically,

but occasionally, activates the gene of interest in cells. I show that when the activated gene of interest in these cells is tagged with a fluorescent protein, these cells and their progeny can be followed over time in a living animal. Using the same principle I also describe that microsatellite-dependent stochastic activation of an oncogene, in this case oncogenic H-RAS, results in the formation of tumors within 5 days. Since zebrafish are transparent during embryonic development, this technique provides the opportunity to study the early stages of tumor development in a living animal.

In **Chapter 4**, I focus on G-quadruplex instability in the nematode. Previous studies have shown that G-quadruplexes can induce 50-300 bp deletions in the genome of mutant worms that lack the helicase DOG-1 (the worm counterpart of the human Fanconi anemia-associated gene FANCJ), however, the underlying mechanism of this process was unknown. In this study, I reveal the molecular mechanism that explains the formation of these G-quadruplex-induced deletions. I show that a polymerase, named polymerase theta, plays an essential role in the formation of these G-quadruplex-induced deletions. Polymerase Theta-Mediated End-Joining (TMEJ) of G-quadruplex-induced DNA damage results in deletions that are characterized by a) an extremely narrow size distribution between 50-300 bp, b) one nucleotide homology at the junctions and c) occasional insertions of up to 20 base pairs templated from the flanking sequences. In the absence of polymerase theta, G-quadruplexes can lead to large deletions spanning several kilobases. By comparing the genomes of worms that lived geographically separated for millions of years, I provide evidence that G-quadruplex-induced genomic deletions occurred also during the normal evolution of wild type worms.

In **Chapter 5**, I investigate the stability of G-quadruplexes in human cells. I provide preliminary data that G-quadruplex motifs can be highly polymorphic between humans. Using fluorescence-based G-quadruplex-instability reporters, I further show that G-quadruplex-dependent genomic instability is increased upon transcription and upon treatment with G-quadruplex-binding ligands. Finally, I provide data that argues that G-quadruplexes can lead to small genomic deletions in human cells. Although it is unclear at the moment whether this genomic instability phenotype is FANCJ and polymerase theta dependent, the newly developed G-quadruplex-instability reporters presented in this chapter will, in combination with functional genetic screens, facilitate an answer to this question in the future.

My thesis ends with a summarizing discussion and future perspectives on microsatellite and G-quadruplex instability and their link to disease. For example, I speculate how polymerase theta-activity can lead to the expansion of a specific G-quadruplex motif that is associated with Amyotrophic Lateral Sclerosis (ALS) and how small molecules that bind G-quadruplexes or inhibit polymerase theta activity can aid in the treatment of cancer and other microsatellite- and G-quadruplex-related diseases.