THE PENNSYLVANIA STATE UNIVERSITY

THE SCHREYER HONORS COLLEGE

DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOLOGY

POSSIBLE SPECIFIC ROLE OF TRIPLEX AND HAIRPIN DNA STRUCTURES IN CAUSING MICROSATELLITE INSTABILITY THROUGH DIFFERENT MODES OF REPLICATION CYCLE

QIAO KAI LAW
SPRING 2012

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Biotechnology
with honors in Biotechnology

Reviewed and approved* by the following:

Maria Krasilnikova
Research Assistant Professor of Biochemistry and Molecular Biology
Thesis Supervisor

David Gilmour
Professor of Biochemistry & Molecular Biology
Honors Advisor

Wendy Hanna-Rose
Associate Department Head for Undergraduate studies
Department of Biochemistry and Molecular Biology

* Signatures are on file in the Schreyer Honors College

Abstract

Microsatellite is a tandem array of repeated nucleotides that consists of six or more repeated motif units. They are present all over the genome and are relatively unstable compared to normal DNA sequences. When microsatellite exceeds the threshold length, it can undergo large-scale expansions/contractions, which is often referred to as microsatellite instability. Many microsatellite repeats, especially trinucleotides, have been linked to over 20 types of neurological disorders and cancers. However, the mechanism of microsatellites instability in the genome that sometimes leads to the diseases is still enigmatic. Here we studied the effect of di- and tetra-nucleotides repeats in the first replication cycle as well as in the subsequent replication cycles. Replication of SV40-origin based plasmid transiently transfected in 293A or Cos-1 cells, served as model system for the first and subsequent replication cycles respectively. Plasmid replication was analyzed using two-dimensional gel electrophoresis to visualize the replication effects of different di- and tetranucleotide repeats in two different replication modes. We showed that the first replication cycle was more prone to stalling by triplex-forming repeats, while the subsequent replication cycle was mostly stalled by the repeats with hairpin-forming capacity. The replication stalling capacity of different repeats that we analyzed was consistent with the microsatellite mutability rates observed in the human-chimpanzee genomic evolution, indicating that replication stalling may be an important factor driving the instability of those repeats. It is interesting that triplex-forming, and not hairpin-forming potential of repeats, had a better correlation with the previously observed mutability rate in tetranucleotides. However, for dinucleotides, it seemed that hairpin formation correlated better with instability while triplexes had only a

limited role. We suggest that triplex structure may have an impact on genomic instability when chromatin structure is loose, which occurs in the first replication cycle of transfected DNA. Comparison between the replication stalling effects of microsatellite in different replication mode with the mutability in chimpanzee-human evolution may indicate the specific role of different DNA secondary structures in causing microsatellite instability.

Acknowledgements

Here I would like to thank Dr. Maria Krasilnikova for guiding me throughout the years in completing this project and thesis as both my thesis advisor and project monitor. She has been very generous and kind in giving her guidance in leading us to carry out and understand the experiments. She has also taught me to be critical and analytical in the experiment, skills which could prove useful in my future career. Lastly, she has also been a very caring supervisor to us in our lab that not only acts as a mentor, but a friend.

In addition, I would like to thank my colleagues - Kian Hui Yeoh, Su Jen Khoo, Abhinaya, Nari, Stephen for their help at difficult times and being there for me whenever I needed help. I would like to especially thank Su Jen Khoo, Abhinaya and Nari for helping to contribute the data for dinucleotides.

Finally, I would like to express my gratuity to my honors advisor, Dr. David Gilmour, for being so helpful and patient in guiding me towards completing the thesis and as my academic advisor. Without his support, I definitely would not be able to complete my thesis that well.

Table of Content

Abstract……………………………………………………………………………………………….………i-ii

Acknowledgements……………………………………………………………………………..iii

Introduction……………………………………………………………………………………......1-11

Materials and Methods……………………………………………………………………………12-17

Results………………………………………………………………………………………………18-28

Discussion and Future Directions………………………………………………………..29-38

References…………………………………………………………………………………………39-44

Introduction

Microsatellites, also known as short sequence or short tandem repeats (STR), are nucleotide repeats of 1-6 bases in DNA, i.e. $(A)_n$ where A represents a repeated nucleotide sequence motif, and n the repeat number. STRs are found prevalently in both eukaryotes and prokaryotes, where they are randomly distributed throughout the genome [27]. STRs are often used as genetic markers due to their polymorphism [3, 12, 40]. There are different types of STRs depending on the sequence of the repeated motif. Microsatellites are categorized according to the number of nucleotides in a minimum unit of nucleotide motif, such as mono-, di-, tri- and tetra- nucleotides (Table 1). In addition, some STRs have interruption in them and can be further divided into different categories. Interrupted microsatellite are STRs with a mutation occurring in the microsatellite stretches. Imperfect microsatellites are STR repeats that are disrupted by an insertion and compound microsatellites are STRs that consist of more than one repeating motif. Each of these microsatellites has its own characteristics, and in majority of the cases, trinucleotide repeats are the most studied due to their clinical relevance.

Categories	Annotation	Example
Mononucleotide:	$(A)_{12}$	AAAAAAAAAAAA
Dinucleotide:	$(AT)_8$	AT AT AT AT AT AT AT AT
Trinucleotide	$(AAT)_5$	AAT AAT AAT AAT AAT
Tetranucleotide	$(ATGC)_4$	ATGC ATGC ATGC ATGC

Imperfect	$(AG)_2(CA)_2(AG)_4$	AGAGCACAAGAGAGAG
Interrupted	$(AG)_2TG(AG)_3$	AGAGTGAGAGAG
Compound	$(AG)_4(CT)_4$	AGAGAGAGCTCTCTCT

Table 1: Examples of microsatellites categories. There are different types of microsatellites according to their motifs and composition. Different categories of microsatellites might have different characteristics and thus are often studied separately in the researches.

The reason for the increased interest in microsatellite repeats is due to their role in numerous neurological disorders (Table 2). The neurological diseases that are linked to microsatellites include human neurodegenerative disorders (Fragile X syndrome, Huntington's disease, myotonic dystrophy, Friedrich's Ataxia, etc.) and hereditary nonpolyposis colorectal carcinoma [35, 45]. It is worth noting that most of these diseases are caused by a dynamic expansion of a trinucleotide (triplet) microsatellite, with the exception of colorectal carcinoma, which was linked to mono/dinucleotides [27, 35].

Disease	Chromosome	Repeats	Normal copy number	Disease copy Numbers
Fragile X(A) (Verkerk et al. 1991; Kremer et al.1991)	Xq27.3	CGG	6-54	>200
Myotonic Dystrophy (Brook et all. 1992)	19q13.3	CTG	5-27	>50
Huntington's (The Huntington's disease Collaborative Research Group	4p16.3	CAG	11-39	36-121

1993)				
Spinal Bulbar muscular atrophy	Xq	CAG	13-30	39-60
Dentalorubral-Palidoluysian atrophy	12	CAG	8-25	54-68

Table 2: Neurological disorders that are associated by microsatellites and its threshold range. Noticeably, most of the neurological diseases related to microsatellites are mostly caused by an expansion in tri-nucleotides. This is also why they are the repeats that are of most interests and widely studied by the scientists.

Despite the fact that these diseases manifest themselves with different unrelated symptoms, and the mechanism of how a microsatellite affects the gene expression is different in each specific case, they all have a similar underlying non-mendelian pattern of inheritance, which is often called dynamic instability [32, 39]. The number of repeated units in a microsatellite is transmitted from one generation to the next without any length changes and does not pose any threat, up until the repeat's length exceeds a threshold number of units [18, 35, 39]. Over the threshold length, the repeat length often changes in the intergenerational transmissions: it can start undergoing large expansions or contractions [35]. An expansion or contraction refers to an increase or decrease in the number of repeated unit in a microsatellite, typically between 10 to 3000 bases [31]. After a few rounds of replication, these repeats may accumulate more than a thousand of repeated units and subsequently affect some important gene functions that can lead to diseases [35]. With each large increase in repeats length, the probability of dynamic expansion rises significantly along with disease severity [31, 35, 25, 18]. It is interesting that in each of the disorders just one microsatellite was found to be expanded: the repeats

in other locations in the genome were not affected by an expansion [31]. Another common feature of the trinucleotide repeat diseases is the delayed onset: the age of onset inversely correlates with the number of units in the expanded pathogenic repeat [7, 36]. There is still no cure for those diseases, and the mechanism of expansion is yet to be uncovered. So it is no wonder why microsatellites, especially those disease-causing repeats, are the focus of many research groups.

In Friedreich's ataxia disease, the triplet repeats of $(GAA)_n$ within the first intron of the gene frataxin were identified as a driving force behind the development of the disease [21, 23, 25]. These STRs within normal length limits (3-30 unts) [25] are always present inside the intron of frataxin and are passed on to the next generation with a relatively high fidelity. When an expansion event occurs and STRs length exceeds the threshold length of $(GAA)_{39}$, these repeats acquire a significant probability of undergoing dynamic expansion/contraction that can result in Friedreich's ataxia [25].

Microsatellite are inherently unstable [45]. Studies showed that the stability of a microsatellite is determined, in part, by the repeated minimum unit, motif and length [3, 21, 23, 45]. The observed patterns showed that these factors are not mutually exclusive and very often, they work together in determining the mutability of the STR [12, 31]. There is one common theme to these observed characteristic: the motifs that are prone to form stable secondary structures such as hairpins and triple helices (Figure 1) have a higher mutability rate as compared to other motifs, given the same total repeated unit and length [1, 23]. In the paper by Kelkar et al 2008 [23], a genome-wide study on evolution of microsatellites in human and chimpanzee, showed a strong correlation between

repeat's length and motif and its respective stability. As mentioned also, those motifs with a strong capacity of forming secondary structure such as $(GGAA)_n$ and $(GAAA)_n$ with triplexes [11] outside the conventional duplex B-DNA structure, are more likely to have a steeper slope of mutability.

All these observations naturally lead to a logical hypothesis of the DNA secondary structure's role in the instability nature of microsatellites. To form a stable secondary structure with a double-stranded DNA, there is a minimum free energy level requirement that needs to be fulfilled [31]. This nucleation energy is associated with a necessity to unwind the double helix and disassemble a part of it into two separate strands before a structure such as a triplex DNA or a hairpin can form [18, 40]. If the final structure is more energetically favorable, it will be stable, but there are always some unpaired nucleotides at the borders of a non-B and Watson-Crick structures [18] that have a destabilizing effect. With an increase in the total length of a non-B structure, the relative input of these unpaired bases decreases, which leads to the higher overall stability. This may explain a critical length threshold before dynamic expansion/contraction of microsatellites occur since a certain copy number length is needed to overcome the energy barrier so that the structure is energetically favorable [31]. Some STR stretches such as $(CTG)_n$ and $(GGC)_n$ in the well studied neurological diseases are shown to be more prone to form hairpins structures *in vitro* [16]. These non-canonical DNA structures could stall the DNA replication process. They could readily form during the lagging strand synthesis within a single-stranded part that temporarily exist in Okazaki fragment synthesis [18, 25, 47]. They could also potentially form in a superhelical double-stranded

DNA before a helicase associated with the replication fork even arrived at this position [37]. It is also possible that flanking sequences may influence the stability of STRs by affecting their nucleation energy: stabilizing or destabilizing the double-stranded helix in the vicinity of a repeat [16, 40].

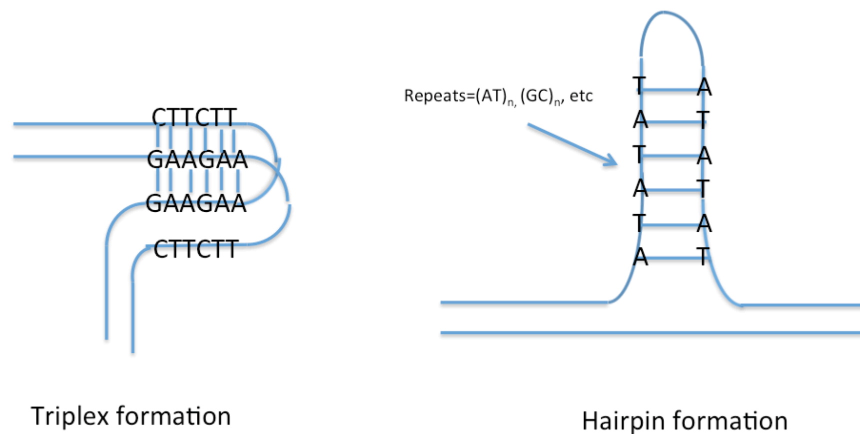

Figure 1: Unusual DNA secondary structure formation. Some microsatellites were shown to be able to facilitate formation of unusual DNA structures such as hairpin and triplex structures *in vitro*. This is proposed to be due to the repetitive sequences that could form a self-complementary pairing under certain circumstances.

As mentioned, stable formation of secondary structures was thought to be a crucial factor of consideration in proposing the mechanism model for microsatellites instability. Currently, there are several models proposed for the dynamic expansion of microsatellites. Among them, DNA slippage is the most prevalent and widely accepted mechanism that could account for the microsatellite instability [2]. In this model, the repeated nature of microsatellite sequences can provide various sites for self-

complementation [4, 18, 31] when the DNA duplex dissociates into single strands, which does take place during the DNA replication process. This phenomenon happens more noticeably in the nascent (newly synthesized) strand of DNA, where the STRs can self-anneal and form a hairpin-like structure. Upon the hairpin structure formation, it is possible that DNA polymerase skips the part that was involved in a hairpin resulting in a deletion. If the hairpin forms in a nascent strand, this can result in the expansion of a repeated stretch upon DNA repair. On top of this, the DNA polymerase tends to slip, especially on the long and repetitive STR tracts [38]. The dissociation and misaligned re-association of DNA polymerase is not a novel event and has often been observed in the prokaryotes [45]. These two combined, the DNA polymerase may 'miss' the self-annealed STRs and continue to incorporate the redundant nucleotides. When a second round of replication occurs, the previously self-annealed structure unwinds to be copied into the daughter strands. Thus, expansion occurs as a result. Slippage of DNA usually involves the breakage of many bonds, which tends to be very energetically unfavorable [23]. However, when a repeat is long enough and a secondary structure formed is stable, the energy difference in between the duplex state and slipped state is minimized, allowing a large-scale slippage (and expansion) to occur [35]. So, this model is consistent with the observation that the higher propensity of microsatellite to form secondary structure, the more likely it is to be unstable and to disrupt the otherwise faithful replication process.

One of the mechanisms proposed for the instability of microsatellites is the recombination model. This proposed mechanism also involves the repair system after the replication stalling was encountered. According to this model, the replication fork is

stalled by the formation of an unusual secondary DNA structure. To bypass this replication blockage, the DNA has to be repaired. One of the systems to repair DNA blockage includes recombination and fork reversal [8]. Recombination repair sometimes results in an unequal crossing-over that confers more repeat to one strand than the other [27, 45]; as a result, an event of contraction/expansion is observed. An observation that favors this model is that microsatellites become increasingly unstable when they are located within Alu repeats that mediate recombination at high frequency [21]. However, the caveat is that most triplet diseases are caused by expansion. If the recombination repair model were in play, we would expect to observe equal events of expansion and contraction. However, that is not the case [3, 12, 31].

Another proposed mechanism is the template-switching model proposed by Goldfless et al., 2006 and Shishkin et al., 2009 [41]. In this model, it was proposed that the newly synthesized DNA in the leading strand could continue its elongation by switching its template from the template strand to the nascent lagging DNA strand by chance ($\sim 10^{-3}$ per replication) [41]. Due to the repetitive nature of the microsatellite on both strands, the nucleotides are complementary to each other thus making this switch more feasible, as both the strand could make pairing at multiple points along the repeats tract [41]. This 'switched-template' replication could go on until it reaches the end of the Okazaki fragment. At that point, the polymerase could switch back to its original leading strand template in order to continue replication. Repeats previously incorporated based on the nascent lagging strand template would have to self-anneal and adopt a folded form in order for elongation to continue. In addition, it was also believed that template switching

could occur more than one time in a single replication cycle [41], which makes it a more likely explanation for the catastrophic expansions (large-scale expansions) observed in neurological diseases including Friedreich's ataxia.

Another potential determinant in the stability of microsatellite is the mismatch repair system. Cell's repair system is thought to have a significant, but contradicting, role in the stability of a microsatellite. There are some studies that showed the presence of mismatch repair system (MMR) stabilizes the microsatellite repeat [26, 42] but there are also some experiments that showed that MMR actually promotes the instability of microsatellites. With simple repeats in yeast, mammalian and bacteria cells, the rate of mutability of microsatellites greatly increases when the mismatch repairs system is compromised, indicating the role of MMR in stabilizing the low copy microsatellites [31, 27, 32, 42]. This phenomenon explains the MMR's role in the length dependence of instability. At lower lengths, DNA polymerase exonuclease proofreading activity and MMR are more effective in reducing errors in DNA synthesis [26]. In addition, the looped-out DNA hairpin intermediate is short and less stable at lower length and thus less favorable to be formed [26]. However, as the repeat length goes up, the repair system cannot catch up and thus cannot repair as effectively: an unusual DNA secondary structure inhibits the proper function of MMR proteins and also lowers the efficiency of MMR proteins binding [42]. However, there are some studies that showed that in the presence of MMR system, the $(CAG)_n$ was observed to be more expanded, not stabilized [33]. This was proposed to be caused by the preferential binding of MMR proteins to the hairpin intermediate loop formed by the repeats, thereby stabilizing the structure [33].

Despite many proposed mechanisms, none has been confirmed as the definitive mechanism for microsatellites instability. And most probably, there is no single explanation that could account for the instability of microsatellites because different microsatellites are unique in their abilities to form secondary structures and in their interaction with the replication machinery [13].

Recent discovery of an alternative replication mode that takes place in the newly transfected DNA has pointed out that an expanded GAA repeat can have different effects on replication depending on specific conditions inside the cell [5]. This alternative replication mode was observed to happen in the first replication cycle of transfected DNA, while chromatin was still at its formation stage, and may also play a role in the first replication of the zygote that happens right after the sperm fertilizes the egg [5]. In this first cycle of zygotic division, the DNA also does not have a mature chromatin structure and it was shown that the chromatin assembly was not completed even after its first replication [22]. Contrary to the first replication cycle, $(GAA)_n$ repeat had little or no effect on the subsequent replication cycles of the transfected DNA [23]. In fact, most of the instability of GAA repeats in Friedreich's ataxia was shown to occur during the first embryonic divisions [9]. Also, in this first replication cycle of transfected DNA, random initiation of replication was observed to occur throughout the plasmid [5]. Preliminary studies showed that this unique replication mode presumably initiates at the AT rich region and requires the presence of single-stranded DNA [5]. Interestingly, these single-stranded DNA regions could also be provided by the formation of secondary structures

such as hairpin or triplex. The formation of these structures is more likely in DNA with a loose chromatin structure, that may occur immediately after fertilization, and possibly in cancer cells since they were also shown to have an abnormal chromatin structure [5].

Here we studied how a structure-forming potential of different STR motifs (dinucleotides and tetranucleotides) correlated with their ability to affect replication in the first and subsequent replication cycles of transfected DNA. We also examined the correlation between the relative strength of replication stalling at different microsatellites with their mutability previously estimated by comparing human and chimpanzee genomes.

Materials and Methods

Strains and cell lines:

A bacterial strain that was used in cloning and purification of repeats-containing plasmids was *E. coli* XL1-blue (Stratagene).

Replication analysis of plasmids was carried out in two mammalian cell lines: 293A human embryonic kidney cells (Invitrogen) and COS-1 African Green monkey kidney cells (Sigma). COS-1 is a T antigen-expressing cell line that provides for replication initiation at the SV40 origin.

Plasmid design

Two plasmid vectors were used in the experiment: pYES-Bsg [25] and pUCneoH. The pYES-Bsg plasmid was used for generating plasmids containing desired copy number of repeats because it was shown previously to be able to maintain the inserted repeats stably in the cells [25]. The generated multiple copy number repeats were then transferred into a second plasmid vector, pUCneoH, to carry out the replication analysis. PUCneoH is a derivative of pUCneo [6] that has a SV40 origin for replication in the cells expressing T antigen and was obtained in a two-step process. First, a unique HindIII site of pUCneo was inactivated by religation of a blunt-ended HindIII-digested pUCneo. Second, a fragment containing a new HindIII site was inserted at the blunt-ended AatII. The obtained pUCneoH plasmid allowed a more efficient repeats cloning into the sticky HindIII site. To this end, fragments containing different repeats sequences were obtained by HindIII digestion of corresponding repeats-containing pYES plasmids described below. These fragments were cloned into the HindIII site of the pUCneoH vector.

Generating plasmids with long microsatellite repeats

Various plasmids with expanded repeats were obtained in two steps. In the first step, an oligonucleotide with a short 10 units repeat (Figure 2) was inserted into the Eco81I site of pYES-Bsg (Figure 3).

```
5′- TCAGAATTCGTTGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAAGCT -3′
   3′- CTTAAGCAACTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTTCGAAGT -5′
```

Figure 2: Example of oligonucleotides used in repeat-containing plasmids cloning that contained $[AATG/TTAC]_{10}$ repeat(underlined). These complementary oligonucleotides form an overhanging sticky end of Eco81I restriction site upon annealing, which ease the insertion of these oligonucleotides into the plasmid pYes-Bsg containing single Eco81I cut site.

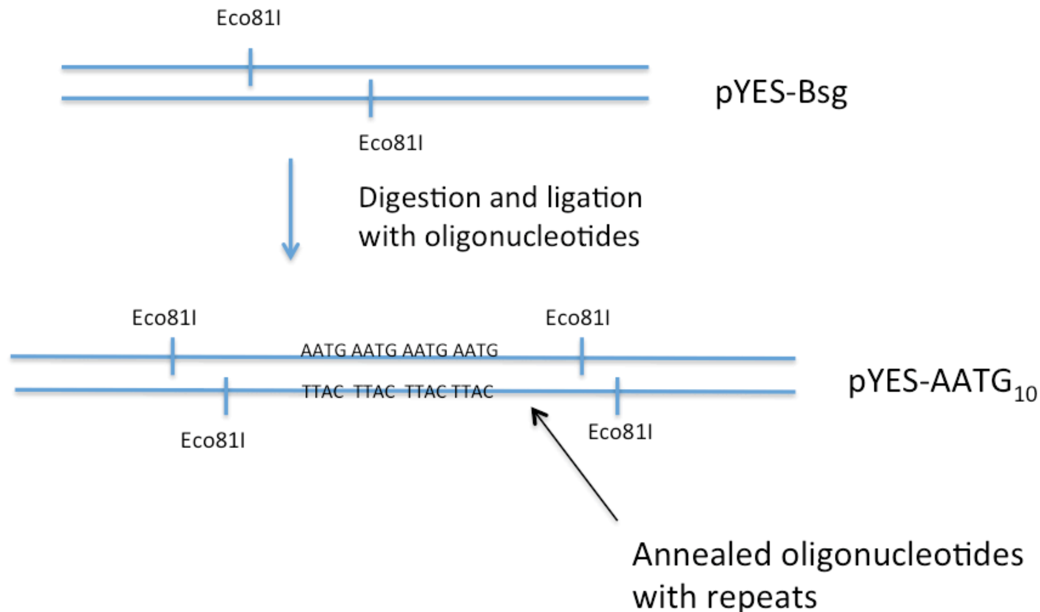

Figure 3: Ligation of designed oligonucleotide repeat into pYES-Bsg plasmid vector. The pYes-Bsg plasmid contains a single cut site for Eco81I restriction enzyme. Thus, after digestion with Eco81I into linear DNA, pYes-Bsg will be a vector plasmid with perfect sticky end for ligating our repeat of interest that also contains an overhanging Eco81I restriction site.

In the second step, to generate expanded repeats, we took advantage of the BsgI, a restriction endonuclease that recognizes a certain sequence, and then cuts in approximately 16 nucleotides away from it (Figure 4). In pYes-Bsg plasmid, two BsgI recognition sites were located in the flanking sequences from both sides of the repeat so that digestion with BsgI would result in a fragment that consists only of a microsatellite repeat without any flanking sequences. For generating elongated microsatellites, pYES-Bsg derivative that contained 10 repeated units was partially digested with BsgI to obtain a linear plasmid with a single cut at the end of a repeat, and then dephosphorylated (Figure 5). Ligation of this vector with the BsgI-BsgI fragment of the same plasmid resulted in different constructs with various numbers of tandemly inserted fragments. The exact numbers of repeats were confirmed by sequencing.

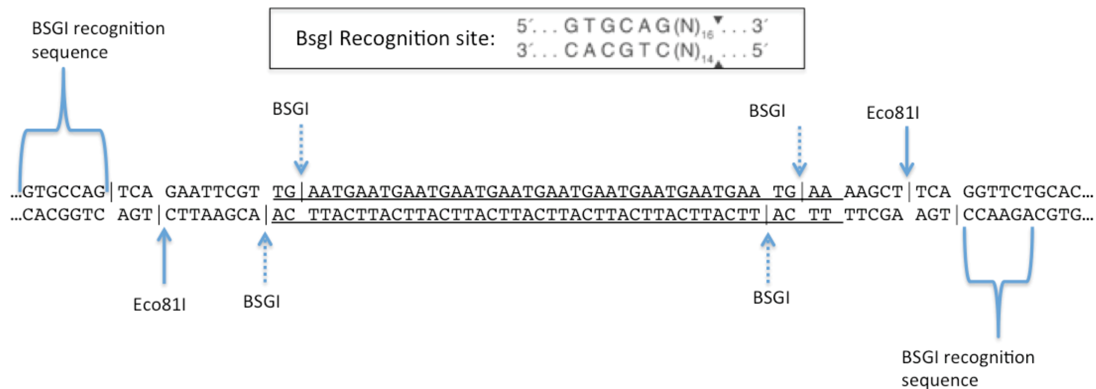

Figure 4: Digestion scheme that cuts the pure repeats stretch from the plasmid. There are two BsgI recognition site on the pYes-Bsg plasmid. By digesting with BsgI, the plasmid was cut at 16 nucleotides downstream after the recognition sequence (GTGCAG). This will generate a pure stretch of DNA repeat (underlined and separated by vertical line) as shown in the figure. The Eco81I sites were shown for reference as to where the oligonucleotide was inserted into the vector.

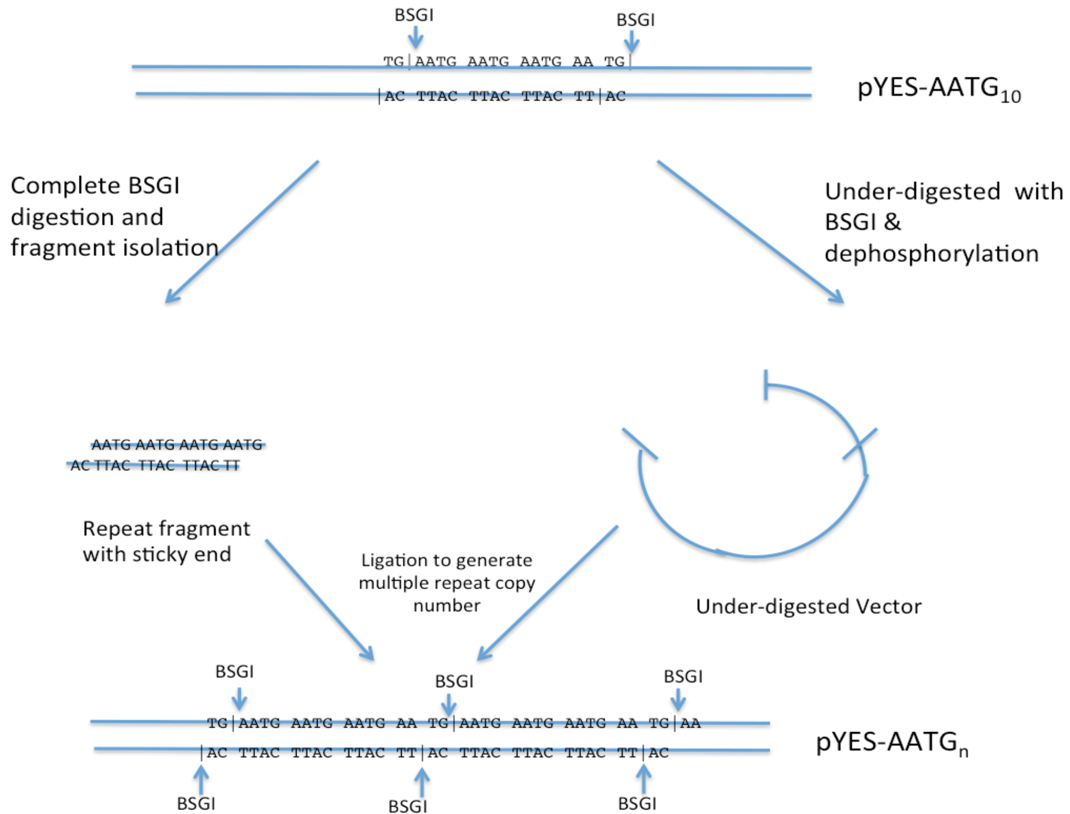

Figure 5: A cloning scheme for creating multiple copy repeats. The digestion scheme utilizes the underdigestion of the previous plasmid with inserted DNA repeat in order to generate a linear vector plasmid with repeat of interest at one end. Then, the plasmid vector is ligated with the pure repeat fragments obtained as described in figure 4. Because BsgI generated pure fragment repeats with similar sticky ends, there is a probability that the vector plasmid would be ligated with more than one pure repeat fragment.

Mammalian cell culture and transfection

COS-1 cells and 293A cells were grown in Dulbecco's modified Eagle medium DMEM supplemented with 10% newborn calf serum (COS-1) or fetal Bovine serum (293A).

COS-1 cells were transfected using Turbofect (Fermentas) at the recommended 50% confluence by manufacturer's instruction for observation of subsequent replication cycles (SV40 origin based replication). The cells were grown for 30 hours before

harvesting.

For the analysis of first replication cycle, 293A cells were transfected at 60% density using Lipofectin (Invitrogen) according to the manufacturer's protocol. The cells were grown for 6 hours before harvesting.

Isolation of replication intermediates

Replication intermediates were isolated using a modified Hirt extraction protocol [21]. Transfected cells were washed with TBS (50 mM Tris–HCl pH 7.0, 150 mM NaCl) once and lysed by adding 750 μl of lysis solution (50 mM Tris– HCl pH 7.0, 20mM EDTA, 10mM NaCl, 10% SDS, 0.2mg/ml proteinase K) per one 100mm culture dish for 20 minutes at room temperature. Then, 0.1875 ml of 5M NaCl was added to the lysates, and chromosomal DNA was precipitated overnight at 4 ^{o}C. Next day, the chromosomal DNA was pelleted at 4 ^{o}C at 15,000 rpm for 50 min, the supernatant was transferred in a 2 ml Eppendorf tube and incubated with 5 μl of proteinase K for 2 hours at 55°C. Plasmid DNA was then purified by phenol/chloroform extraction twice, followed by a chloroform extraction. The purified plasmid DNA was then ethanol-precipitated, and dissolved in 30 μl of TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA).

Neutral-neutral two-dimensional (2D) gel electrophoresis of Replication Intermediates and Southern transfer

10μl of replication intermediate samples isolated as described above were digested with AflIII and were separated in the first dimension gel (0.4% gel agarose, Sigma-Aldrich) in 1xTBE at 1 V/cm for 15h. The gel lanes fragments that contained replication intermediates spanning from one to two AflIII fragment sizes were cut out from the gel, and embedded in a second dimension 1% agarose gel. The second dimension separation was performed at 4° C, 2.5V/cm for 15h. The gel was transferred to the Zeta-probe nylon membrane (Biorad) according to the manufacturer's protocol. The membrane was cross-linked and then hybridized with P^{32} labeled radioactive fragments corresponding to the fragment of interest (Figure 6). The labeling was performed with DecaLabel™ DNA Labeling Kit (Fermentas). The radiolabeled membrane was then exposed to an X-ray film at -80°C overnight to 3 days.

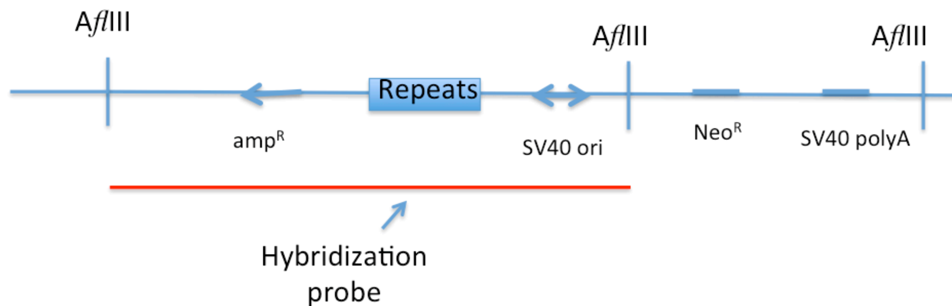

Figure 6: Hybridization probe (red line) corresponding to fragment of interest. The probe was prepared by digesting the plasmid with AflIII restriction enzyme. Then the digested fragments were separated on gel and extracted from the gel based on the length of fragment of interest. The fragment was then radiolabeled using P^{32}.

Results

In mammalian genome, microsatellites of different motifs and lengths are characterized by different mutability levels [23]. Also, for some microsatellites, relatively long repeats are abundant in the genome, while others are present only as short versions with low numbers of repeated units [23]. To analyze the correlation between the mutability of microsatellites and replication stalling, we used a model system where microsatellite repeats were located within the SV40-based plasmid pUCneoH. To study the effect of microsatellite repeats on replication, replication intermediates were isolated from the growing cultures, digested with AflII and analyzed using 2D gel electrophoresis. Repeats were positioned 1/3 of the way within the AflIII fragment. This would mostly generate a Y arc since SV40 origin was located at the very end of the AflIII fragment (Figure 7). While in COS-1 cells replication was initiated by T antigen at the SV40 origin, in 293A cells that lacked T antigen, the plasmid was replicated transiently upon an alternative replication mode that initiated randomly throughout the plasmid [5]. Replication in 293A cells was used as a model system for the situation when chromatin was not assembled in a regular structure. It was previously shown that alternative replication, which occurs at a loose chromatin structure with only a few nucleosomes present at scattered positions, was affected by $(GAA)_n$ [6] and $(Fra16B)_n$ repeats [49] much more than the SV40-based replication [5]. To model this situation as well, we analyzed replication progression through the repeats in 293A cells where we expected only the first replication cycle driven by alternative replication mode to occur. The alternative replication mode would also generate mostly a Y arc upon the AflIII digest, since replication forks mostly entered

from the side of the fragment, and only initiation within the middle third of it would result in the bubble arc (figure 8). For both alternative and SV40 driven replication, replication stalling at a repeat would result in a bulge along the replication arc that was otherwise smooth. The intensity of the bulge relative to the intensity of the arc itself was interpreted as the strength of the stalling [15]. The presence of a bulge at the stalling position was due to the fact that replication fork's progression was impeded at that point, leading to accumulation of a replication intermediate of a particular form and thus resulting in accumulation of the radioactive signal.

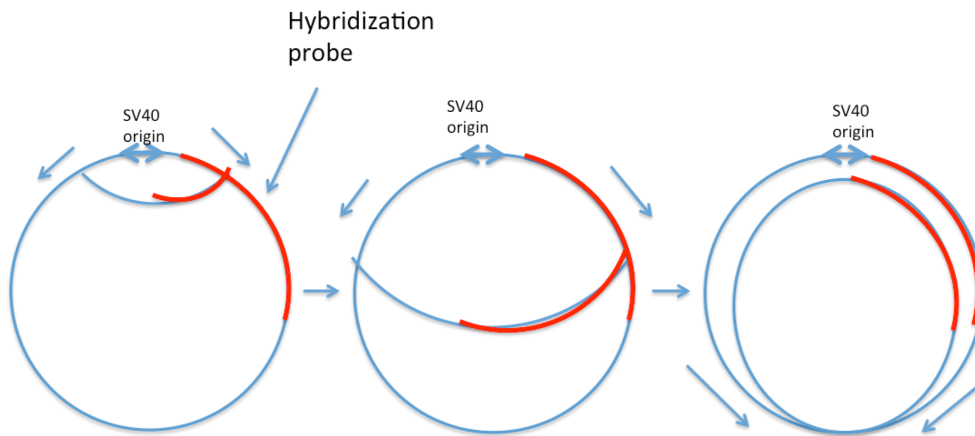

Figure 7: Replication fork progression. Since the replication forks start outside the probed region, the replication intermediates will be detected as a Y arc as indicated in the figure by red y-shaped line. As the replication goes to completion, we will detect linear plasmid as revealed by 2D gel electrophoresis (figure 8). The bubble arc would only be observed if and only if the replication starts from the inside of the probed region.

Combined 2D gel pattern

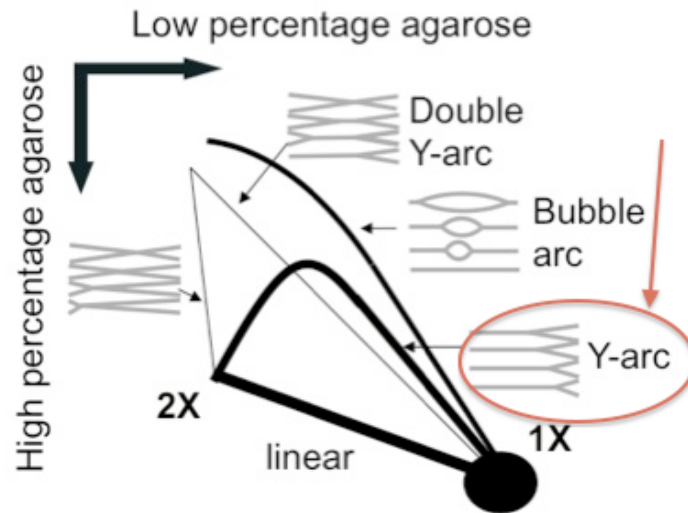

Figure 8: Interpretation of a 2D gel pattern with its associated replication intermediates shapes. The 2D gel electrophoresis separates the replication intermediates by weight in the 1st-dimensional low percentage agarose gel and then followed by a shape separation in 2nd-dimensional high percentage agarose. This will result in a 2D-gel replication intermediate shape as shown in the figure. The intensity of each arc on the gel is the representation of the total accumulated replication intermediate population at the point of lysis. Thus, a replication stalling at a certain point in the plasmid will give a more intense signal at a specific spot that would be interpreted as a bulge along the line. The Y arc pointed by the red arrow will be the focus of our study, as any replication stalling in the probe area would be revealed on the Y arc as a bulge or intensified area.

Replication stalling observed in first replication cycle for microsatellites with high tendency of forming triplex DNA structure

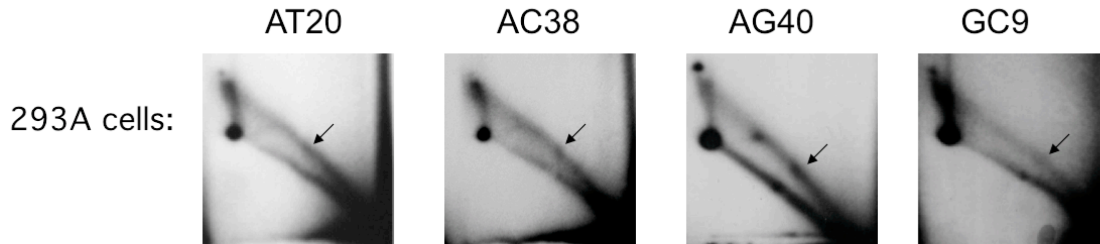

Figure 9: 2D gel analysis of di-nucleotide microsatellites in the first replication cycle. The gel result showed that only AG_{40},that have the capability of forming triplex structure, have the most prominent replication stalling in the first replication cycle. AT_{20}, AG_{40} and AC_{38} that could only form hairpin had very little effect of replication stalling as pointed by the arrows.

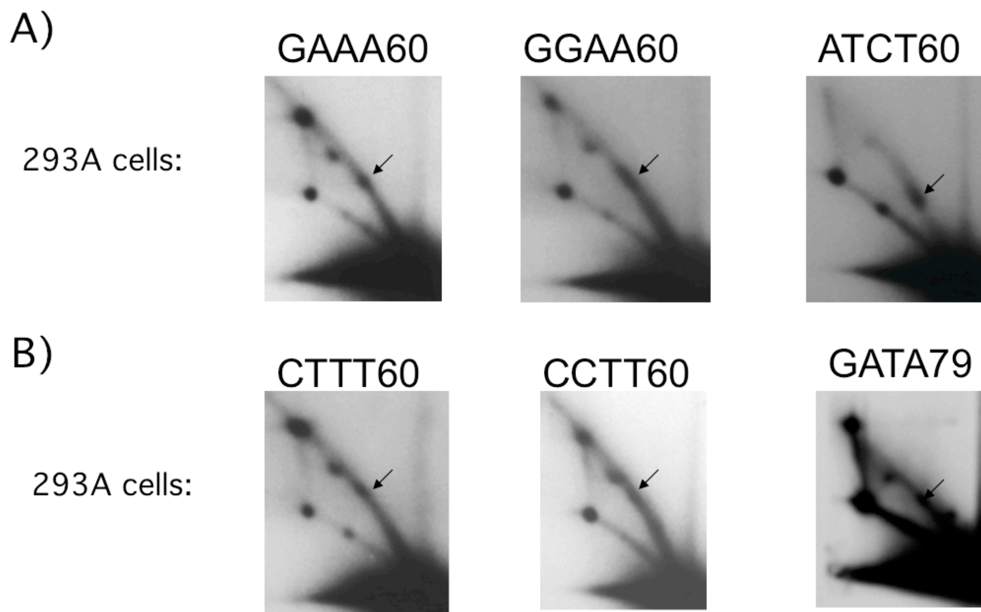

Figure 10A & 10B: 2D gel analysis of tetra-nucleotide microsatellites in first replication cycle (The high mutability repeat group in human-chimpanzee genome comparative analysis). **10A:** The results showed that $GAAA_{60}$, $GGAA_{60}$ and $ATCT_{60}$ all had very strong replication stalling effect on the first replication, which

was indicated by the intense bulge (arrows) along the Y arc. **10B:** The repeats on panel B are complementary to those repeats in panel A, but inserted into the plasmid in the opposite orientation. Despite the different repeats orientation, the results still showed evident stalling effects in the first replication, indicating that the replication stalling from the repeats were independent of the repeats orientation.

Note: On some of the Y-arcs, there was a second bulge observed. This was suspected to be caused by stalling of inverted repeats from the other direction. The repeats of interest were inserted at about 1/3 of the probe and thus any replication stalling caused by the inserted repeats would be expected to form at the 1/3 position on the Y arcs.

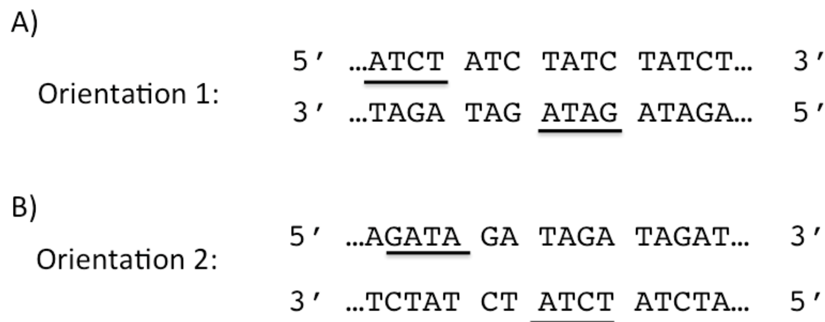

Figure 11: Due to the nature of the microsatellite repeats, $(ATCT)_n$ could be also read as $(TCTA)_n$ or $(CTAT)_n$, depending on the reading frame. This also holds true for other repeats. For example, $(GC)_n$ could also be read as $(CG)_n$ and $(ACAT)_n$ could also be read as $(ATAC)_n$.Figure 11A and 11B showed that repeats inserted in both orientation, where $(GATA)_n$ is $(ATCT)_n$ inserted in an opposite orientation.

Our results showed that AG_{40}, $GAAA_{60}$, $GGAA_{60}$, and $ATCT_{79}$ caused the most prominent stalling effect when they undergo replication through alternative replication mode in the 293A cells (Figure 9 & 10). The 293A cells had no T antigen expression,

thus the subsequent replication cycles based on SV40 origins did not take place in this system [5]. Coincidentally, most of the microsatellite repeats that exhibited a strong stalling effect in the first replication cycle in 293A cells, with the exception of $ATCT_{79}$ and $GATA_{79}$(GATA is the opposite orientation of ATCT) (Figure 11), were also the repeats that were shown to form triplex DNA structure *in vitro* [12]. Among those were all the polypurine-polypirimidine sequences such as $(GA)_n$ [12], $(GAA)_n$ [25], $(GAAA)_n$ [12] and $(GGAA)_n$ [12] repeats. The repeated nature of the sequence could also allow for the mirror symmetry of the structure. For $(GATA)_n$,it was not shown to form any DNA secondary structures *in-vitro*. However, there exists a transcription binding factor that binds specifically to the sequence $(GATA)_n$, namely GATA transcriptional factor[50], which we suspect had a role in the replication stalling in the first replication cycle. Surprisingly, the repeats that have high tendency to form hairpin structures such as AT_{20} and GC_9 [11] had shown very little replication stalling effect in the first replication model. This strongly suggested that the specific formation of a triplex structure could affect the progression of replication fork in first replication cycle.

Replication stalling in subsequent replication cycle is most probably caused by hairpin structure formation

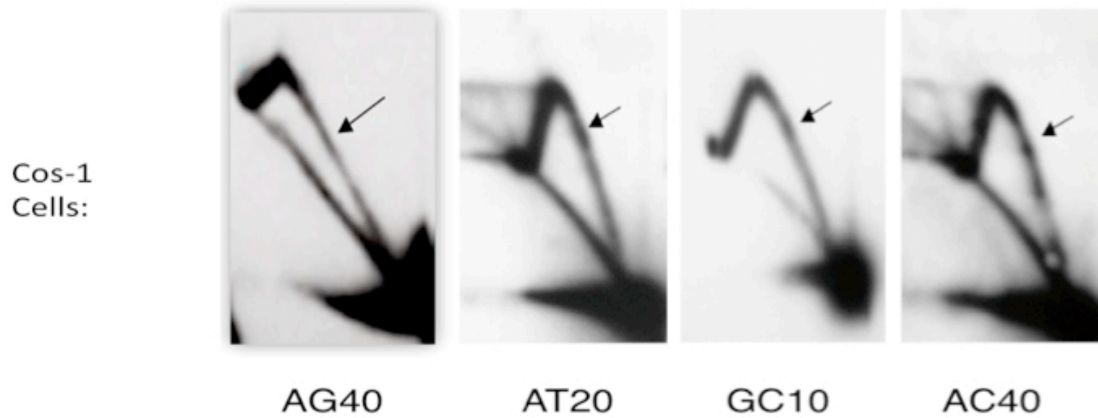

Figure 12: 2D gel analysis of di-nucleotide microsatellites in subsequent replication cycle. The di-nucleotide repeats (AT_{20},GC_{10},AC_{40}) had shown the most evident replication stalling effect in the subsequent replication cycle. The arrows pointed out the spots where the replication forks were stalled and thus accumulating the signals of replication intermediates of the same shape, revealed by 2D gel separation. **Note:** There was also another second bulge on the Y arc, which was suspected to be caused by the inverted repeats that stall the replication from another direction. We focused on the bulge on 1/3 position on the Y arc because that is where the repeats were located.

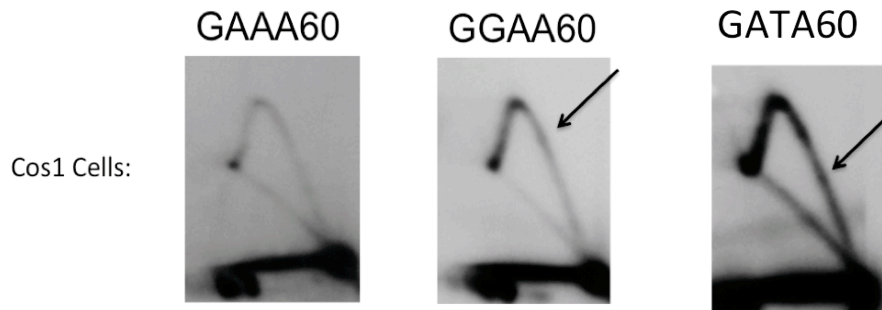

Figure 13: 2D gel analysis of tetra-nucleotide microsatellites in subsequent replication cycle (High mutability group in human-chimpanzee genome comparative analysis). In the subsequent replication cycle, the repeats ($GAAA_{60}$, $GGAA_{60}$, $GATA_{60}$) that previously gave significant stalling effect in our first replication model only had a minor effect in subsequent replication.

In COS-1 cells, microsatellite repeats of AT_{20}, GC_{10}, AC_{40}, and $GATA_{60}$ were observed to cause replication stalling in SV40-origin-based replication, during the subsequent replication cycle (Figure 12 & 13). These microsatellite repeats had a propensity to form hairpins by self-complementarity within the sequence [11]. For the dinucleotide microsatellites, the strength of the stalling correlated with the ability of the sequence to fold into the hairpin. This result supported the hypothesis that hairpin structure may be responsible for replication stalling and thus indirectly promoting microsatellite instability. Surprisingly, $GAAA_{60}$ repeat that had given significant replication stalling in the first replication cycle did not have an effect on the replication in the subsequent cycles, and both $GGAA_{60}$ and $GATA_{60}$ repeats only had a mild replication stalling effect in the subsequent replication cycle (Figure 14).

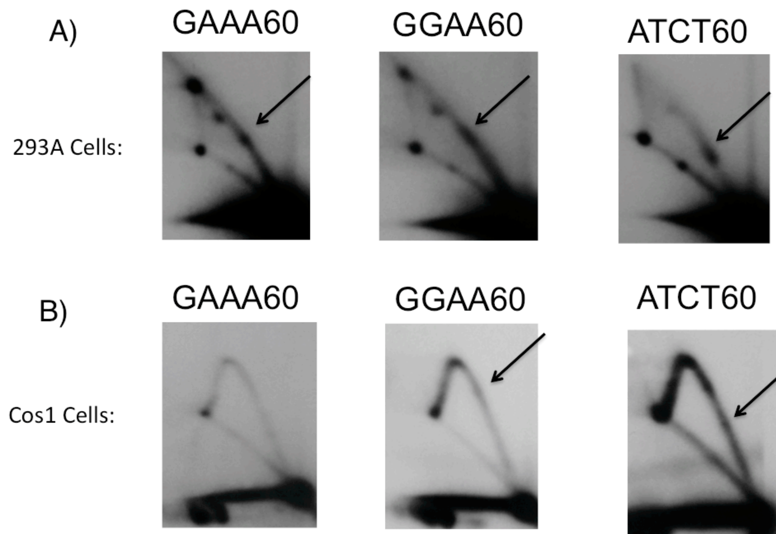

Figure 14: The above figure compares the replications stalling effect of high mutability tetra-nucleotide repeats in both the first replication cycle (panel A) and the subsequent replication cycle (panel B). It is interesting that these repeats had different extent of replication stalling effect, depending on the different replication modes. These tetra-nucleotides were the repeats that experienced high mutability in the human-chimpanzee population genomic evolution. The prominent stalling effect in the first replication and the lack of replication stalling effect in the subsequent replication cycle may indicate that the more significant role of first replication, rather than subsequent replication, in accounting for genomic microsatellite instability.

Repeats that have low mutability and low tendency to form secondary structures did not have significant replication stalling effect

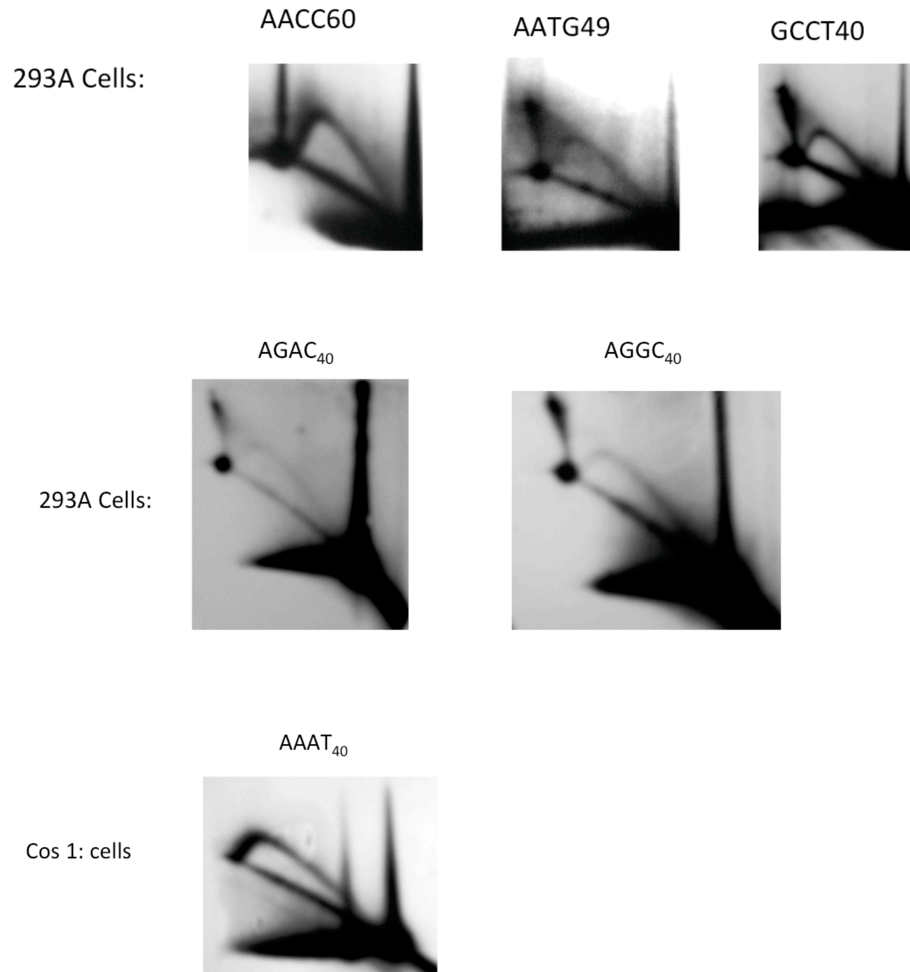

Figure 15: 2D gel analysis of tetra-nucleotide microsatellites in first replication cycle and subsequent replication cycle (low mutability group in human-chimpanzee genome comparative analysis). As predicted from the mutability of these tetra-nucleotides, there was no significant replication stalling effect, whether in the first replication cycle or in the subsequent replication cycle, as shown in the result. Any replication stalling would be interpreted as a bulge or intensified spot along the Y arc (Figure 8).

Consistent with our expectation, repeats that were not observed to form stable non-canonical structures, such as hairpin and triplex ($AACC_{60}$, $AATG_{49}$, $GCCT_{40}$, $AGAC_{40}$,

$AGGC_{40}$, $AAAT_{40}$), did not show much effect on the replication progression, at least up to the length analyzed (Figure 15). There was no significant bulge observed along the Y-arc (refer to Figure 8 and Figure 15). The replication effect of these repeats was compared to their mutability calculated based on the human-chimpanzee comparative analysis [23]. Indeed, the comparison showed that these microsatellites had relatively low mutability in human-chimpanzee evolution, indicating that they were transmitted stably in the population with no dramatic changes. Moreover, they are only detected at short lengths in the genome (Figure 18), which was due to their low mutability as there is no dynamic expansions or contractions that contribute to length variation [1, 23].

Discussion and Future Direction

Most studies on microsatellite instability were performed on model systems that focused on replication that took place for more than one round [2, 47]. However, the fact that replication may proceed differently before the chromatin assembly was completed on a plasmid or on a zygote DNA cannot be ignored [5]. Apparently, the first replication cycle of the zygote is hard to study since it occurs only once after the egg was fertilized. In the first replication cycle right after fertilization, the sperm's chromatin undergoes remodeling through replacing the protamine of paternal chromatin with histone [17, 43]. Maternal chromatin is being remodeled as well [17, 43]. During this chromatin remodeling stage, the chromatin remains loosely packed, in order to allow access for DNA remodeling enzymes [43]. Coincidentally, at this point most of the genomic instability was observed as well, including the instability of the GAA repeat involved in Friedreich Ataxia [9, 24].

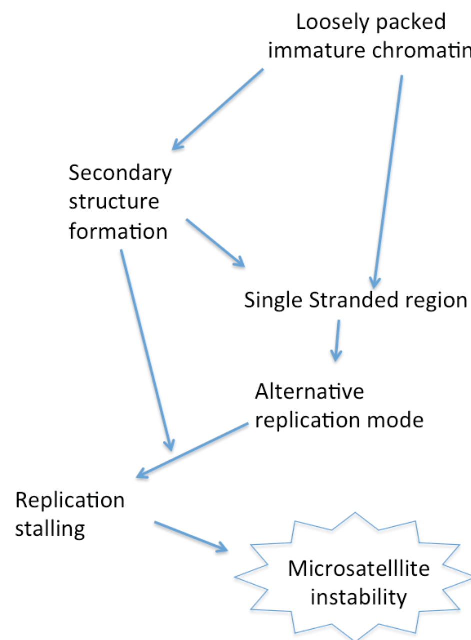

Figure 17: Schematic presentation of the hypothesis on the increased instability of microsatellites in early zygotic divisions. We are proposing that the loosely packed immature chromatin structure is one of the main driving force contributing to the higher tendency of microsatellites in forming secondary structure, as well as to enhance the alternative replication mode which is more prone to replication stalling. The replication stalling then eventually leads to the microsatellite instability that includes dynamic expansions and contractions of microsatellite repeats.

This chromatin formation stage also takes place in newly transfected plasmids, during the first 24 hours post transfection. In this plasmid model system, a novel mode of replication initiation, which occurred simultaneously with chromatin formation, have been recently reported. This replication mode was referred to as alternative replication mode [5] since it was shown to initiate randomly throughout the plasmid until the chromatin assembly is complete, and the presence of single stranded DNA was found to increase its efficiency [5]. Interestingly, it was found to occur even in the G0/G1 phase prior to S phase [5]. Alternative replication mode has been observed in many different

cell lines as well as primary cells so it is very likely to be relevant *in vivo* [5]. Early observation suggested that this replication mode could also start from a single-stranded region on the plasmid associated with increased superhelicity or provided by a non-B DNA conformation, which would also more likely form in condition of immature chromatin structure, although this fact has never been directly demonstrated [5]. We suggest that the tight chromatin structure that can exist on the plasmid during its subsequent cycles can increase the nucleation energy barrier for forming stable secondary structures like triplexes. When replication is initiated by the alternative replication mode, the preformed triplex structure may inhibit the replication fork progression thus causing replication stalling that may eventually lead to microsatellites instability (Figure 17). Interestingly, formation of secondary structures like hairpin and triplex structures can also provide some single-stranded DNA that could potentially increase the initiation efficiency of the alternative replication mode. So, it is possible that the secondary structures formation at the repeats and the alternative replication mode could work hand-in-hand to increase the instability of microsatellites.

Prior studies have shown that the formation of secondary structures were tightly linked to the instability of microsatellites [1, 19, 29, 37]. In our study, we went one step further and showed that the potential of a microsatellite to fold into a particular non-B structure such as a triplex or a hairpin, correlates with replication stalling in the SV40 origin-based or alternative replication mode. We have shown that it is important to distinguish between these two distinct replication modes because microsatellites ability to form non-B structures may depend on the plasmid chromatin state. For instance, for the tetra-

nucleotides we observed that the same repeats that had caused significant replication stalling in the first replication cycle, surprisingly, did not have similar effect in the subsequent replication cycle. In particular, $(GAAA)_{60}$ strongly impeded the replication fork progression in 293A cells (first replication model) but had almost no effect in the Cos-1 cells (subsequent replication model). A similar observation was also made for $(GGAA)_{60}$ and $(ATCT)_{60}$ (Figure 14).

We observed that the triplex-structure-forming microsatellites were the ones that had most prominent replication stalling effect in the first replication cycle, while hairpin structure formation propensity was tightly linked to strong replication stalling in the subsequent replication cycles. For instance, $(GAAA)_{60}$ and $(GGAA)_{60}$, a polypurine mirror repeat sequences that were capable of forming stable triplex structure *in vitro* [14], give prominent replication stalling in the first replication cycle but had almost no effect in the subsequent replication (Figure 14). On top of that, the subsequent replication cycles were only impeded mostly by the hairpin-forming repeats, such as $(AT)_{20}$, AC and $(GC)_{9}$. This is most likely due to the differences in chromatin states in the first and subsequent replication cycles.

Under the condition of a tightly packed chromatin, the formation of triplex structures may be less favorable because local unwinding of a significant stretch of DNA double helix within the microsatellite is required for a stable pyrimidine/purine/ purine or pyrimidine/purine/pyrimidine to form a triplex structure [18]. Formation of a stable triplex structure is a two-stage process: first, a nucleation energy is required to break

hydrogen bonds within a certain area of a microsatellite in order to use a single-stranded DNA as a third DNA strand in the triplex formation [18, 37]. This would be energetically unfavorable in tightly packed chromatin structure. Second, the free energy of the triplex structure formation must be close or lower than the free energy of forming the conventional B-DNA in order to push the equilibrium towards the stable triplex formation [30]. Otherwise, the DNA would remain in a duplex rather than a triplex due to the free energy level difference. Thus, a condition of a highly negative supercoiling is usually needed in order to offset the high nucleation energy by the facilitating the unwinding of the double helix [18]. In fact, *in vitro*, an increase in the negative supercoiling by -0.025±10% was shown to increase the formation probability of a triplex by 30% [46]. Different isoforms of the triplex structure were shown to be stabilized by magnesium ions or slightly acid pH [18]. We hypothesize that the loosely packed immature chromatin structure allowed the local unwinding and participation of third DNA strand in the triplex that would not be possible in subsequent replication cycles. The chromatin state of a transfected plasmid may be reminiscent of that of a newly formed zygote when it is going through the chromatin remodeling stage, although the chromatin of a human zygote has never been studied in details to make a conclusive statement.

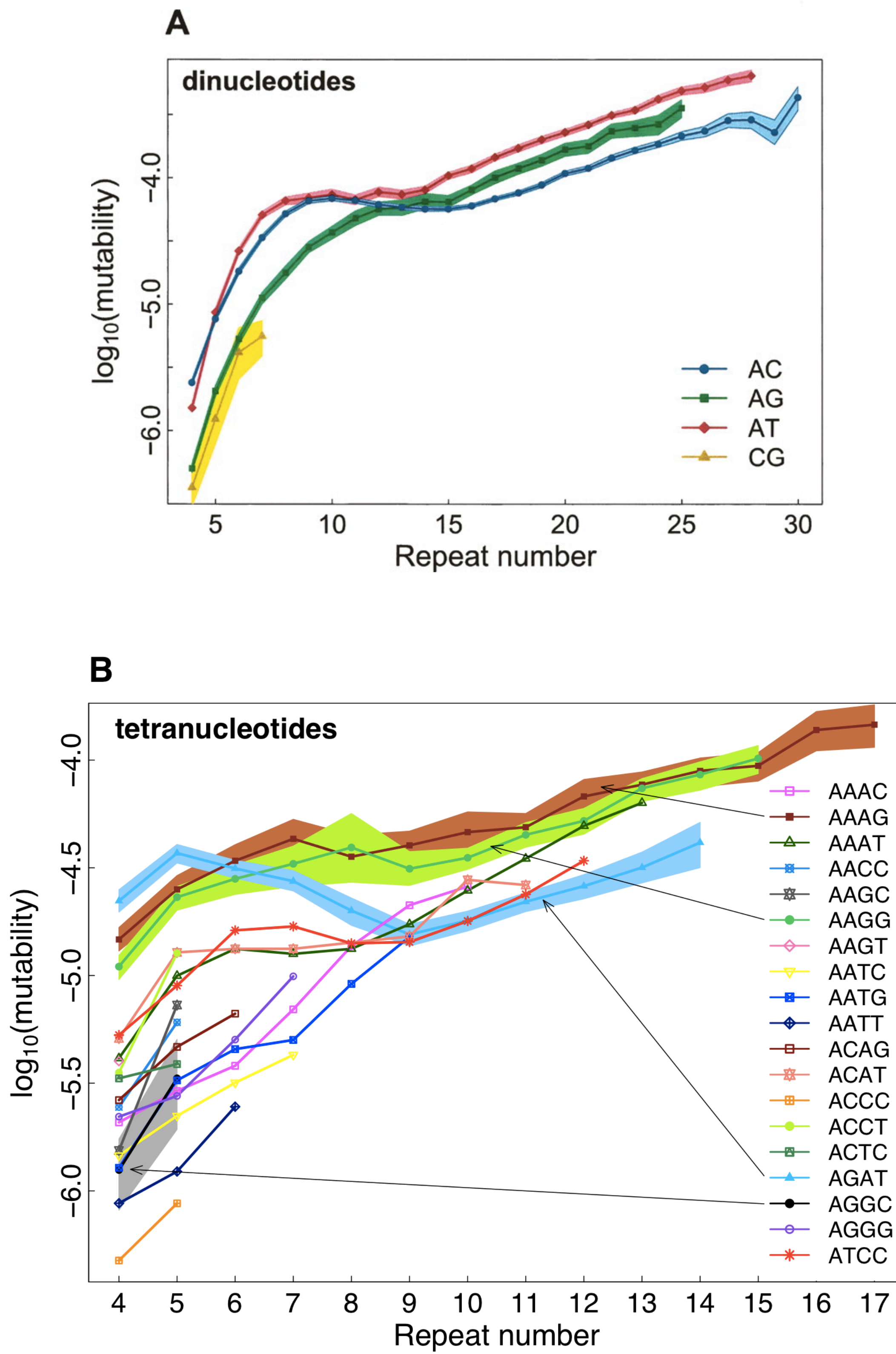

Figure 18: Study of microsatellite instability in chimpanzee-human genomic evolution comparison for di-

and tetra-nucleotide repeats in Kelkar et al 2008 [23]. Data available on Genome Research (www.genome.org). **Figure 18A:** $(AT)_n$ had the highest mutability in the evolution, followed closely by $(AG)_n$ and $(AC)_n$. $(GC)_n$ however only occured at low length (shorter than 10 repeats) but had the widest spread in the genome (indicated by the standard deviation area). **Figure 18B:** On the graph, $(AAAG)_n$ and $(AAGG)_n$ were the repeats with high mutability and had most spreaded lengths distribution; while $(AACC)_n$, $(AATG)_n$, $(AGGC)_n$ are on the lower portion of the graph with very little spread.

Our data for the first time shows the correlation of replication stalling and mutability for a wide range of microsatellite repeats. Our results of replication stalling at microsatellites were compared to the data previously published by Kelkar et al on the microsatellites changes in human-chimpanzee evolution (Figure 18) [23]. In the paper, a graph of microsatellites mutability depending on their lengths was generated for each sequence motif by collecting and comparing the genomic data of both human and chimpanzee. We have analyzed all the high mutability repeats (AAGG, AAAG, AGAT) and some representative repeats from the low mutability repeats group (AGGC, AACC, AATG) of the tetranucleotide microsatellites (Figure 18B). From this comparison, we could clearly see the positive correlation between replication stalling in the first replication cycle and instability of microsatellite transmission. The repeats that have strong stalling effect in replication especially in the first replication cycle, such as $GAAA_{60}$, $GGAA_{60}$ and $ATCT_{60}$ (complementary to AGAT), also had the highest mutability. Noticeably, all three tetranucleotide repeats, with the exception of AGAT, are the microsatellites that had been shown capable of forming stable DNA triplex structure *in vitro* [14]. We suspect that AGAT may have a capacity of forming a secondary structure that is similar in nature to triplex, possibly a AT and GA duplex that was suggested by Vorlickova et al [44, 48].

The other possibility is the GATA transcriptional factor that binds to the GATA sequence that eventually leads to replication stalling [50]. The result that the most mutable repeats are also causing pronounced stalling in the first replication cycle suggests a role of replication in the condition of a loose chromatin structure in causing instability of microsatellites. Interestingly, for tetranucleotides, the hairpin formation capacity did not seem to have an effect on mutability. Hairpin-forming tetranucleotide motifs such as AATT had low mutability and did not spread widely in the population (Figure 18B). So, for tetra-nucleotides, we are proposing that the instability of microsatellite transmission is most probably through the first replication cycle and is mostly caused by triplex DNA structure formation.

For the different motifs of dinucleotide repeats, the calculated mutability rates comparison was AT>AG>AC. CG repeat (also can be read as GC) did not spread widely in longer length in the genome so the data collected was not sufficient to analyze mutability of CG for the repeat numbers that were used in our replication analysis (Figure 18A). These mutability data correlate with the replication stalling that we detected for the SV40-based subsequent replication cycles. For the replication stalling in this system, we obtained the following: CG>AT>GA=AC. However, in the first replication cycle only the GA repeat showed a considerable effect as expected based on its triplex-forming potential [11]. A limited input from the triplex structure formation in the first replication cycle can explain the higher mutability level of GA comparing to AC, but mostly the stalling in the subsequent replication cycles could account for the mutability differences.

Even though there seems to be a lot of data linking replication stalling to microsatellite instability [25, 28, 35, 47], the exact mechanism of how instability occurs is still not known. One of the most plausible explanations is that when the replication fork is stalled, it has to undergo repair in order complete the replication [8], which may involve fork reversal or restarting the replication fork [35]. The fork regression can introduce genomic instability [35]. In addition, the repair system that involves disassembling-restarting the fork is a step-wise process [8]. During DNA repair, the complementary strand has to be separated by an unknown mechanism for fork reversal thus increasing the chance that a region of DNA is single-stranded [34]. This increases the probability of part of a microsatellite looping out and forming some misaligned pairs [20]. The general consensus is that the longer the pausing, the higher the chance for misalignment and mutation to occur [10, 20, 28]. Even when the replication blockage is repaired, the previously formed looped out repeats may become expanded in the newly synthesized DNA strands.

The disparity in the microsatellite instability during the first and subsequent replication cycles are important to study because in some diseases including Friedreich's ataxia, the dynamic expansion/deletion were traced to the first embryonic cells division, when replication and chromatin formation occur at the same time. [9]. The study on these differences may reveal the underlying mechanism of disease progression.

To summarize, all our results are consistent with the important role of triplex and hairpin DNA structures in genomic instability. It seems likely that there is a specific role of DNA

triplex structure in causing microsatellite instability for tetranucleotides and DNA hairpin formation in causing microsatellite instability for dinucleotides in the genomic transmission. In the future studies, it would be interesting to directly confirm the formation of those secondary structures inside the mammalian cell. It would also be important to demonstrate that the first replication cycle of a zygote, similar to the first cycle of the transfected DNA, is indeed a source of increased mutability in microsatellites sequences.

References:

1. Bacolla, A. (2008-10). Abundance and length of simple repeats in vertebrate genomes are determined by their structural properties. Genome research, 18(10), 1545-53.doi:10.1101/gr.078303.108
2. Bacolla, A. (2009-04). Non-B DNA conformations as determinants of mutagenesis and human disease. Molecular carcinogenesis, 48(4), 273-85.doi:10.1002/mc.20507
3. Bhargava, A. (2010-03). Mutational Dynamics of Microsatellites. Molecular biotechnology, 44(3), 250-266.doi:10.1007/s12033-009-9230-4
4. Brewer, B. J. (2011-03). Origin-dependent inverted-repeat amplification: a replication-based model for generating palindromic amplicons. PLoS genetics, 7(3), e1002016.doi:10.1371/journal.pgen.1002016
5. Chandok, G. S. (2011-03). A distinct first replication cycle of DNA introduced in mammalian cells. Nucleic acids research, 39(6), 2103-15.doi:10.1093/nar/gkq903
6. Chandok, G. S. (2012-01). Effects of Friedreich's ataxia GAA repeats on DNA replication in mammalian cells. Nucleic acids researchdoi:10.1093/nar/gks021
7. Chang, D. J. (2009-02). DNA damage tolerance: when it's OK to make mistakes. Nature chemical biology, 5(2), 82-90.doi:10.1038/nchembio.139
8. De Michele, G. (1994-08). Late onset Friedreich's disease: clinical features and mapping of mutation to the FRDA locus.. Journal of neurology, neurosurgery and psychiatry, 57(8), 977-979.doi:10.1136/jnnp.57.8.977
9. De Michele, G. (1998-11). Parental gender, age at birth and expansion length influence GAA repeat intergenerational instability in the X25 gene: pedigree

studies and analysis of sperm from patients with Friedreich's ataxia. Human molecular genetics, 7(12), 1901-1906.doi:10.1093/hmg/7.12.1901

10. Dere, R. (2004-06). Hairpin Structure-forming Propensity of the (CCTG{middle dot}CAGG) Tetranucleotide Repeats Contributes to the Genetic Instability Associated with Myotonic Dystrophy Type 2. The Journal of biological chemistry, 279(40), 41715-41726.doi:10.1074/jbc.M406415200
11. Eckert, K. A. (2000-07-15). Mutational analyses of dinucleotide and tetranucleotide microsatellites in Escherichia coli: influence of sequence on expansion mutagenesis. Nucleic acids research, 28(14), 2831-8.
12. Eckert, K. A. (2002-07). Mutation rate and specificity analysis of tetranucleotide microsatellite DNA alleles in somatic human cells. Molecular carcinogenesis, 34(3), 140-50.doi:10.1002/mc.10058
13. Eckert, K. A. (2009-04). Every microsatellite is different: Intrinsic DNA features dictate mutagenesis of common microsatellites present in the human genome. Molecular carcinogenesis, 48(4), 379-88.doi:10.1002/mc.20499
14. Frank-Kamenetskii, M. D. (1995). Triplex DNA structures. Annual review of biochemistry, 64(1), 65-95.doi:10.1146/annurev.bi.64.070195.000433
15. Friedman, K. L. (1995). DNA Replication [46] Analysis of replication intermediates by two-dimensional agarose gel electrophoresis. Methods in enzymology, 262, 613-627.doi:10.1016/0076-6879(95)62048-6
16. Gacy, A. M. (1995-05-19). Trinucleotide repeats that expand in human disease form hairpin structures in vitro. Cell (Cambridge), 81(4), 533-40.

17. Grenier, L. (2010-08-01). Paternal exposure to cyclophosphamide affects the progression of sperm chromatin decondensation and activates a DNA damage response in the prepronuclear rat zygote. Biology of reproduction, 83(2), 195-204.doi:10.1095/biolreprod.109.083345
18. Hanvey, J. C. (1988-09). Intramolecular DNA triplexes in supercoiled plasmids. PNAS : Proceedings of the National Academy of Sciences, 85(17), 6292-6.
19. Heidenfelder, B. L. (2003-01-24). Hairpin formation in Friedreich's ataxia triplet repeat expansion. The Journal of biological chemistry, 278(4), 2425-31.doi:10.1074/jbc.M210643200
20. Hile, S. E. (2004-01-16). Positive correlation between DNA polymerase alpha-primase pausing and mutagenesis within polypyrimidine/polypurine microsatellite sequences. Journal of molecular biology, 335(3), 745-59.
21. Hirt,B. (1967) Selective extraction of polyoma DNA from infected mouse cell cultures. J. Mol. Biol., 26, 365–369.
22. Imschenetzky, M. (1995-10). Sea urchin zygote chromatin exhibit an unfolded nucleosomal array during the first S phase. Journal of cellular biochemistry, 59(2), 161-7.doi:10.1002/jcb.240590205
23. Kelkar, Y. D. (2008-01). The genome-wide determinants of human and chimpanzee microsatellite evolution. Genome research, 18(1), 30-8.doi:10.1101/gr.7113408
24. Kovtun, I. (2004-10). Somatic deletion events occur during early embryonic development and modify the extent of CAG expansion in subsequent generations. Human molecular genetics, 13(24), 3057-3068.doi:10.1093/hmg/ddh325

25. Krasilnikova, M. M. (2004-03). Replication stalling at Friedreich's ataxia (GAA)n repeats in vivo. Molecular and cellular biology, 24(6), 2286-95.
26. Kroutil, L. C. (1996-01-23). Exonucleolytic proofreading during replication of repetitive DNA. Biochemistry (Easton), 35(3), 1046-53.doi:10.1021/bi952178h
27. Li, Y. (2002-12). Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Molecular ecology, 11(12), 2453-65.
28. Lomonosov, M. (2003-12). Stabilization of stalled DNA replication forks by the BRCA2 breast cancer susceptibility protein. Genes & development, 17(24), 3017-3022.doi:10.1101/gad.279003
29. Lopes, M. (2006-01-06). Multiple mechanisms control chromosome integrity after replication fork uncoupling and restart at irreparable UV lesions. Molecular cell, 21(1), 15-27.doi:10.1016/j.molcel.2005.11.015
30. Matsumoto, N. (1995). Dynamics in isomerization of intramolecular DNA triplexes in supercoiled plasmids. Nucleic acids symposium series (1979), 34, 209-10.
31. McMurray, C. T. (1995-10). Mechanisms of DNA expansion. Chromosoma, 104(1), 2-13.doi:10.1007/BF00352220
32. McMurray, C. T. (1999-03). DNA secondary structure: A common and causative factor for expansion in human disease. PNAS : Proceedings of the National Academy of Sciences, 96(5), 1823-1825.doi:10.1073/pnas.96.5.1823
33. McMurray, C. T. (2008-07-01). Hijacking of the mismatch repair system to cause CAG expansion and cell death in neurodegenerative disease. DNA repair, 7(7), 1121-34.doi:10.1016/j.dnarep.2008.03.013

34. McMurray, C. T. (2010-11). Mechanisms of trinucleotide repeat instability during human development. Nature reviews. Genetics, 11(11), 786-799.doi:10.1038/nrg2828

35. Mirkin, E. V. (2007-03). Replication Fork Stalling at Natural Impediments. Microbiology and molecular biology reviews, 71(1), 13-35.doi:10.1128/MMBR.00030-06

36. Myers, R. (1983-01). Maternal Transmission in Huntington's Disease. The Lancet (British edition), 321(8318), 208-210.doi:10.1016/S0140-6736(83)92587-4

37. Pearson, C. E. (2005-10). Repeat instability: mechanisms of dynamic mutations. Nature reviews. Genetics, 6(10), 729-742.doi:10.1038/nrg1689

38. Reuschenbach, M. (2012-02). Absence of Mismatch Repair Deficiency–Related Microsatellite Instability in Non-Melanoma Skin Cancer. Journal of investigative dermatology, 132(2), 491-493.doi:10.1038/jid.2011.326

39. Rose, O. (1998-05). A threshold size for microsatellite expansion. Molecular biology and evolution, 15(5), 613-5.

40. Schlotterer, C. (2001-05). Microsatellite Instability. Encyclopedia of Life Sciencesdoi:10.1038/npg.els.0000840

41. Shishkin, A. A. (2009-07). Large-Scale Expansions of Friedreich's Ataxia GAA Repeats in Yeast. Molecular cell, 35(1), 82-92.doi:10.1016/j.molcel.2009.06.017

42. Sia, E. A. (1997-01-31). Genetic control of microsatellite stability. Mutation research, 383(1), 61-70.

43. Spinaci, M. (2004-10). Maternal chromatin remodeling during maturation and after fertilization in mouse oocytes. Molecular reproduction and development, 69(2), 215-21.doi:10.1002/mrd.20117

44. Suen, I. S. (1999-01-18). Structural properties of Friedreich's ataxia d(GAA) repeats. Biochimica et biophysica acta, 1444(1), 14-24.

45. Toth, G. (2000-07). Microsatellites in Different Eukaryotic Genomes: Survey and Analysis. Genome research, 10(7), 967-981.doi:10.1101/gr.10.7.967

46. Vetcher, A. A. (2002-10-18). Sticky DNA, a long GAA.GAA.TTC triplex that is formed intramolecularly, in the sequence of intron 1 of the frataxin gene. The Journal of biological chemistry, 277(42), 39217-27.doi:10.1074/jbc.M205209200

47. Voineagu, I. (2008-07). Replication stalling at unstable inverted repeats: Interplay between DNA hairpins and fork stabilizing proteins. PNAS : Proceedings of the National Academy of Sciences, 105(29), 9936-9941.doi:10.1073/pnas.0804510105

48. Vorlickova, M. (1998-03). Conformational properties of DNA strands containing guanine-adenine and thymine-adenine repeats. Nucleic acids research, 26(6), 1509-1514.doi:10.1093/nar/26.6.1509

49. Yu, S. (1997-02-07). Human chromosomal fragile site FRA16B is an amplified AT-rich minisatellite repeat. Cell (Cambridge), 88(3), 367-74.

50. Zarin, A. A. (2012-03-29). A GATA/homeodomain transcriptional code regulates axon guidance through the Unc-5 receptor. Development (Cambridge)doi:10.1242/dev.070656 (Cambridge)doi:10.1242/dev.070656

Academic Vita

Qiao Kai Law
518 University Drive, Apt 111,
State College, PA 16801.
lawqk@hotmail.com
814-753 0310

Education

- BS in Biotechnology, Penn State University (2009-2012)
  - Minor
    - Economics
    - Microbiology
    - Biochemistry and Molecular Biology
  - Dean's list (Fall 2010-Fall 2011)

Achievements

- The Evan Pugh Scholar Award (2011)
  - Top 0.5% of the class for outstanding academic performance
- Scholar of Public Service Department scholarship (2009-2012)
  - Full-scholarship for abroad study in United States by Malaysia Government
- Maintained consecutive Dean's list throughout undergraduate tenure
- Scholar of Schreyer's Honors College (2010-2012)

Experiences

- Involved actively in independent research in Krasilnikova's Lab (Spring 2010-Present)
  - Microsatellite project
  - Utilized microbial techniques, tissue culture, 2D gel electrophoresis etc.
- Part-time East Hall Dining Commons (Feb 2011-Present)
  - 12-16 hours commitment every week
- Teaching Assistant for BISC 001 (Fall 2011)
  - Managed some class duties and simple tutoring with the students
- Volunteering as Orientation Assistant (Spring 2012)
  - Helped with assisting new students and document check-ins

Skills

- Computer skills
  - Excellent competency in using Microsoft Offices
  - Adequate computer literacy to troubleshoot some minor computer issues
- Spoken and written languages
  - English
  - Mandarin
  - Malay