THE PENNSYLVANIA STATE UNIVERSITY

THE SCHREYER HONORS COLLEGE

DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOLOGY

UNSCHEDULED REPLICATION AT MICROSATELLITE REPEATS

YEE VOAN TEO
SPRING 2014

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

Reviewed and approved* by the following:

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

Sarah Ades
Associate Professor of Biochemistry & Molecular Biology
Honors Advisor

Richard Frisque
Interim Department Head for Biochemistry and Molecular Biology

* Signatures are on file in the Schreyer Honors College.

Abstract

Friedreich's ataxia is an autosomal recessive neurodegenerative disorder that is primarily caused by the expansion of GAA trinucleotide repeats. This expansion was found to occur during the first pre-embryonic cell division, when the chromatin is not fully mature. The presence of GAA repeats was found to increase the efficiency of unscheduled replication in mammalian cells. Due to its repetitive nature, the GAA repeat can potentially form triplexes that are accompanied by single-stranded DNA regions, which can then serve as loading docks for replication initiation protein. In order to determine the molecular mechanism of the unscheduled replication initiation at the GAA repeats, we used a plasmid with GAA_{57} repeats as a model system to determine which replication initiation protein would bind to the GAA repeats when the chromatin is not fully assembled. We observed the recruitment of three replication initiation proteins—DNA polymerase α, origin recognition complex (ORC) and mini chromosome maintenance (MCM) to the DNA in the proximity of the GAA repeat during early embryogenesis, which supports the hypothesis that unscheduled replication occurs at the GAA repeat in the absence of a complete chromatin structure. Short regions of DNA that are synthesized from this unscheduled replication are expelled from the genome in the form of linear DNA fragments and induce recombination with other genomic regions, which could then lead to genomic instability. In addition, we also showed that tetranucleotide repeats that form triplex structures stall the unscheduled replication in length dependent manner, which is indicative of the same stalling mechanism. We also studied the progression of mammalian replication fork through several tetranucleotides in the first and subsequent replication cycles. We observed that tetranucleotides that could potentially form triplex structures significantly stalled the first replication cycle whereas tetranucleotides

that might form hairpin structures stalled subsequent replication cycles. Lastly, we also compared the strength of replication stalling with the relative mutability of different tetranucleotides at orthologous locations in human and chimpanzee, which showed that stalling intensity is motif and length dependent.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my thesis advisor, Dr. Maria Krasilnikova, for guiding and supporting me in completing this undergraduate research project and thesis. Her enthusiasm and motivation has helped me throughout the time of research and the writing of this thesis. She has been very patient in providing practical guidance and knowledge in leading me to conduct the experiments. It would not have been possible for me to complete this thesis without the guidance and encouragement from her.

I would like to thank my previous and current lab mates in the Krasilnikova lab for their endless help and encouragement throughout. In addition, I would also like to acknowledge the financial support from the College of Science and the Office of Undergraduate Education Undergraduate that has provided recognition for me to conduct high-quality work.

I would also like to thank my honors advisor, Dr. Sarah Ades, for her guidance in completing the thesis. Without her support, this thesis would not have been possible. Last but not least, I would like to thank my family—my parents and sisters—for their endless encouragement and support.

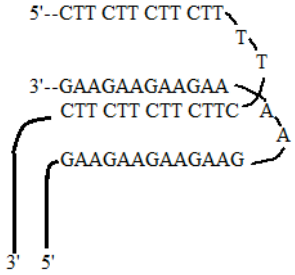
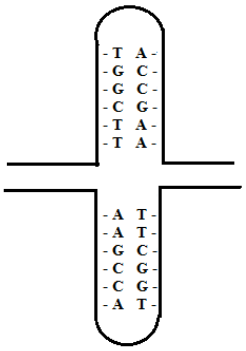
Table of Content

Introduction

Repetitive DNA elements such as microsatellites, minisatellites and inverted repeats are found abundantly in the human genome. These repeated and mobile DNA elements made up approximately 43% of the euchromatic portion and 18% of the heterochromatic portion of the human genome [41]. Some of the repetitive DNA elements can undergo homologous recombination that will result in the duplication or deletion of the sequences that lie in between the repeats [20]. Among all the repetitive DNA elements, microsatellites are the most prevalent repetitive DNA motifs, where they made up approximately three percent of the human genome. Microsatellites are short DNA sequences of 1-8 bases that are repeated in tandem. They are distributed non-randomly across exons, introns, 5'-untranslated regions (UTR) and 3'UTR. Microsatellites have high mutation rates due to their repetitive nature that is prone to DNA strand slippage. This slippage will in turn lead to the expansion of the sequence and replication stalling. This explains why microsatellites could lead to genomic instability in mammalian cells [26].

Unlike most of the canonical right-handed B conformation of DNA, some of these repetitive DNA elements can adopt non-B DNA structures such as cruciforms, triplexes and G-quadruplexes due to their repetitive nature (Table 1) [48]. These structures have been known to contribute to genomic instability by the disruption of DNA replication and transcription [4]. For instance, it has been reported that the cruciform structures aligning at the end of *lacI* were found to result in the deletion of this gene in *Escherichia coli* [15, 49]. Palindromic AT-rich repeats were found to be susceptible to double-strand breaks that could eventually lead to chromosomal translocation. Over half of the breakpoints that are involved in cancer-specific chromosomal translocation occur in fragile sites that are associated with AT-rich repeats. [11, 21, 49].

Mapping and sequencing of DNA sequences with deletion breakpoints have indicated the involvement of repetitive DNA motifs [20]. Thus, this indicates the possible roles of alternative DNA structures in generating those deletion breakpoints. It has been proposed that the genomic instability caused by these secondary structures is due to the stalling at the secondary structure that in turn blocks the DNA replication process [5].

Name	**Conformation**	**General Sequence requirement**	**Example of sequence**
Triplex		Homopurine-homopyrimidine repeats	CTTCTT TTCTTC GAAGAA AAGAAG
Cruciform		Inverted repeats	

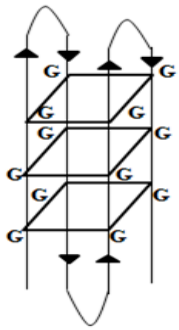
G-quadruplex		$(G_{3+}N_{1-7})_3$ tracts	$(CTAGGG)_n$ single strand

Table 1: Examples of non-canonical DNA structures formed by various repetitive DNA motifs

One of the proposed replication models for repeat expansions is the 5' flap model. During the Okazaki fragment processing, repeats found in the 5' flap can fold back onto itself to form a hairpin structure that impairs replication. The hairpin would then ligate to the 5' Okazaki fragment and cause repeat expansion after the next replication round. It has been previously shown that the deletion of C-terminal flap endonuclease *RAD27* resulted in the expansions of CAG/CTG repeats and CGG/CCG repeats that could form hairpin structures. The 5' flap model is independent of the starting tract length [23].

Another mechanism for the repeat expansion is the 3' replication slippage. The slippage of polymerase at the 3' end of the nascent DNA strand could result in hairpin formation and eventually replication stalling. Upon the next round of replication, the repeated DNA sequences would then contribute to an expansion. An analysis of this replication slippage model in yeast showed that the formation of a hairpin would signal the recruitment of the helicase Srs2 to unwind the secondary structure and post-replication repair (PRR) to prevent this repeat expansion. Mutations of components involved in PRR have shown to increase the efficiency of CAG/CTG repeat expansion [13, 23].

Some other proposed models are the fork reversal and template switching models. Replication stalls at the DNA repeats due to the formation of secondary structures in the lagging strand. This would result in the fork reversal that generates a four-way junction. The 3' repeated end then folds back on itself to form a hairpin structure and continues replication. The repeats would be expanded after the next round of replication. In the template switching model, DNA polymerase in the leading strand can switch to the nascent lagging strand and use it as a template. At the end of the Okazaki fragment, DNA polymerase would switch back to the leading strand and continue the replication. The use of the nascent Okazaki fragment as a template during the switch would result in the expansion of repeat approximately the size of an Okazaki fragment [23].

Although variations in microsatellites can be detrimental and affect normal cell function and phenotype (Fig. 1), several normal and non-harming functions of microsatellites have been proposed [32]. First of all, microsatellites are enriched at certain positions in chromosome and have similar distribution patterns along the chromosome irrespective of their motif. They make up a large portion of centromere regions in many species. These suggested that microsatellites might play a role in chromatin organization [31]. Another role of microsatellites is that they are most likely to be involved in recombination due to their high affinity to recombination enzymes. Some of the dinucleotide repeats such as GT, CA, CT, GA, GC and AT were found to be involved in recombination through the formation of secondary DNA structures. Microsatellites were also found to be involved in regulation of gene activity. For instance, $(TC)_n$ dinucleotide repeat in the promoter region of the *Drosophila* heat-shock protein gene *hsp26* is a transcriptional element. Deletion of the repeat decreases the transcriptional activity of *hsp26*

gene [31]. Another role of microsatellites is that they serve as regulatory protein binding sites. In 1993, Aharoni *et al.* observed that GT and GA-repeated sequences could act as binding sites for proteins in human fibroblasts [31].

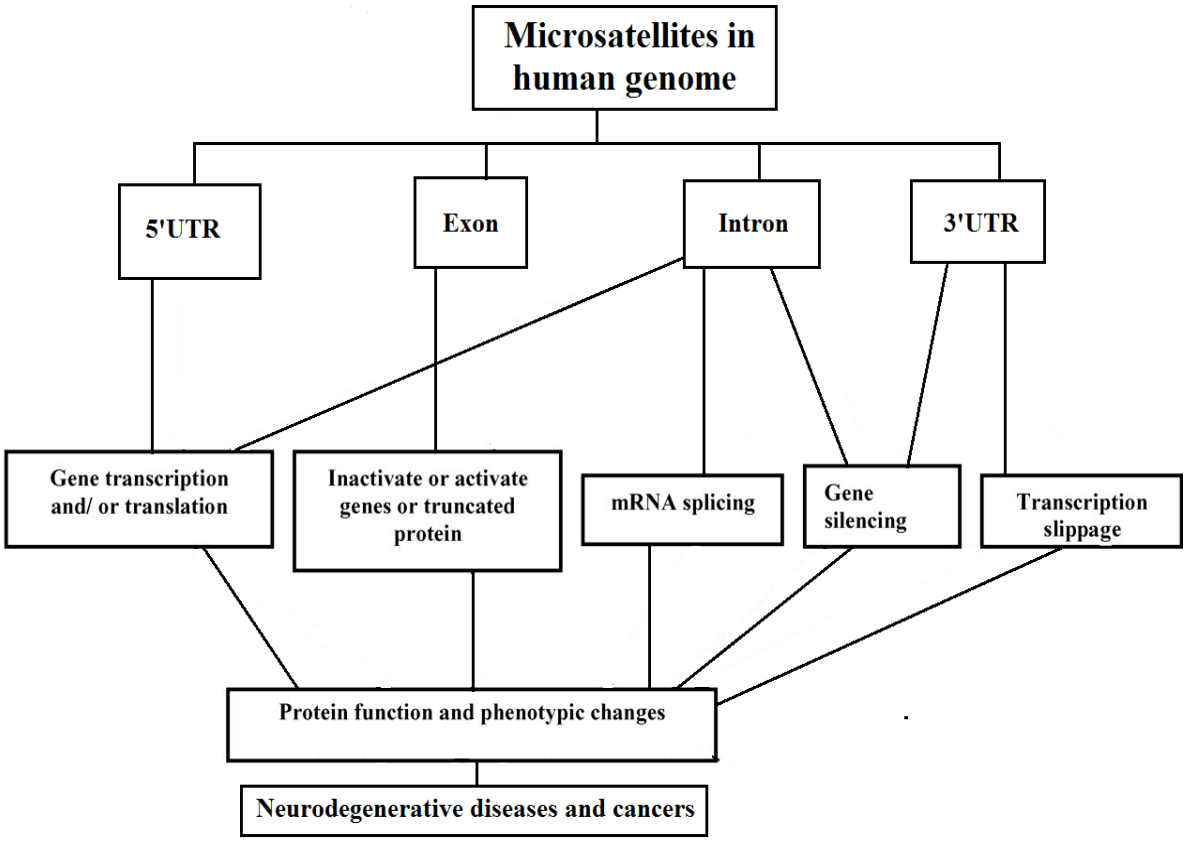


Figure 1: Microsatellites regulatory functions in 5'UTR, 3'UTR, exon and intron

Trinucleotide repeats constitute an important class of miscrosatellites [26]. They are the most widely studied microsatellites because expanded trinucleotide repeats are often associated with neurodegenerative disorders (Table 2) [33]. Short three nucleotide tandem repeats are

commonly found in healthy individuals. They are annotated as $(A)_n$, in which A symbolizes a motif of three nucleotides repeat and n symbolizes the number of repeats. Due to trinucleotide repeats' high mutability rate, they are prone to undergo changes in intergenerational transmissions, which can be either expansions, or contractions. One of the important characteristics of trinucleotide repeats is that their expansions can cause neurological diseases. The expansion, which is also known as dynamic mutation, usually occurs in a non-Mendelian inheritance pattern. When trinucleotide repeats exceed a crucial threshold length, they can expand or contract substantially during parent-child transmission as well as in early embryogenesis. These changes were suggested to take place due to the formation of non-B DNA secondary structures that were shown to promote genomic instability. These expansions in turn cause severe neurological diseases such as Huntington's disease, fragile X syndrome, and Friedrich's ataxia (Table 2) [33]. The normal function of trinucleotide repeats in human genome is not known yet [28].

Trinucleotide repeat diseases in coding sequences

Disease	**Trinucleotide repeat**	**Sequence location**	**Normal repeat number**	**Pre-mutation repeat number**	**Disease repeat number**
Huntington's disease	CAG	*HTT* (exon1)	6-29	29-37	38-180
Dentatorubral-pallidoluysian atrophy	CAG	*ATN1* (exon 5)	6-35	35-48	49-88
Spinocerebellar ataxia type 1	CAG	*ATXN1* (exon 8)	6-39	40	41-83

Trinucleotide repeat diseases in non-coding sequence					
Disease	**Trinucleotide repeat**	**Sequence location**	**Normal repeat number**	**Pre-mutation repeat number**	**Disease repeat number**
Friedreich's ataxia	GAA	*FXN* (intron 1)	5-30	31-100	70-1000
Fragile X syndrome	CGG	*FMR1* (5'UTR)	6-50	55-200	200-4000
Myotonic Dystrophy type 1	CTG	*DMPK* (3'UTR)	5-37	37-50	>50

Table 2: Features of trinucleotide repeat diseases

It has been shown that the length of an expanded trinucleotide repeat was strongly associated with the severity of the respective disease; the greater the repeat expansion, the more severe the disorder [34]. In each successive generation carrying an expanded trinucleotide, there is an increase in severity and onset of the respective disease. This phenomenon is known as anticipation [33]. The mechanism that leads to anticipation is yet to be known.

Herein, we focused on one of the trinucleotide repeat diseases, Friedreich's ataxia (FRDA). FRDA occurs in about 1 in 50,000 individuals, making it the most common inherited ataxia [34]. FRDA is an autosomal recessive degenerative neurological disorder that is primarily caused by homozygous expansions of the GAA trinucleotide repeats in the first intron of the frataxin (*FXN*) X25 gene on chromosome 9q13 [2]. *FXN* gene is expressed in all cells, but mostly in cells that are rich in mitochondria, such as neuron cells. This gene encodes for frataxin,

a protein that is involved in the iron homeostasis. A large expansion of the GAA repeat causes suppression in transcription of the *FXN* gene and thus, leads to a decrease in the levels of frataxin [8, 11, 27]. It has been postulated that non-B DNA structures play a role in the transcriptional decrease [36, 39]. The significantly reduced frataxin mRNA and protein levels result in mitochondrial iron accumulation, abnormal iron homeostasis and increased sensitivity towards oxidative stress [35]. This eventually leads to the death of sensory neurons in dorsal root ganglia, pinocerebellar and corticospinal tract [19]. FRDA is characterized by neuropathological disabilities including muscle weakness, sensory loss and dysarthria [12]. Patients show symptoms as early as two years old, but these symptoms typically become apparent within 10 years [34].

The length of the GAA repeat in FRDA is proportional to the reduction in frataxin and the severity of FRDA symptoms. Healthy persons carry approximately 5-30 GAA repeats whereas FRDA-associated patients with severe multi-system abnormalities carry 70-1000 GAA repeats [33]. The exact mechanism of the GAA repeat expansion is still not known. However, the expansion of the FRDA GAA repeat in mammalian cells was found to occur during the first pre-embryonic cell division, when the chromatin is not fully mature. The GAA repeat that has exceeded the threshold length can cause stalling of the first replication cycle in mammalian cells but not the subsequent cycles [11]. Previous studies performed in our lab has shown that the first replication cycle in mammalian cells goes differently from the subsequent replication cycles in that the first replication cycle initiates randomly across the DNA without a specific initiation site and in the absence of a mature chromatin structure. This unscheduled replication mode has shown to be blocked by the GAA repeats [11, 44].

GAA repeats are able to form an alternative DNA structure, known as the DNA triple helix or triplex. It has been proposed that the replication stalling at GAA repeats was caused by this unusual DNA structure [11]. The DNA triplex consists of three strands: two of the DNA double helix and a third strand, which winds around the major groove of the double helix. Triplexes can either be intermolecular, where the third strand comes from a single-stranded oligonucleotide or from a different DNA molecule, or intramolecular, when the DNA partially unwinds and folds back, making a complex with the remaining part of the double helix [38]. For a triplex to form, DNA is required to contain a sequence of mirror symmetry that consists of only purine or only pyrimidine nucleotides. The purine strand of the double helix always runs as the central strand that provides binding sites for the third strand through Hoogsteen or reverse-Hoogsteen hydrogen bonding. The triplex is of 'pyrimidine motif' (Y·R·Y) if the third strand consists of a stretch of pyrimidines that runs parallel to the central purine strand. In contrast, if the third strand is of a stretch of purines that runs antiparallel to the central purine strand, the arrangement is known as the 'purine motif' (R·R·Y). Triplex of purine motif forms readily under physiological pH condition in the presence of bivalent metal ions, while triplex of pyrimidine motif forms readily under acidic pH condition [49].

The DNA triplex formed by an expanded GAA repeat is stabilized by negative supercoiling and can be of two types: (i) purine strand with the GAA repeats that folds back to the central purine strand (purine motif), or (ii) pyrimidine strand with the CTT repeats that runs parallel to the central purine strand (pyrimidine motif) [9]. These purine and pyrimidine motifs are stabilized by divalent cations such as Mg^{2+}, and polyamines such as spermine and

spermidine, which serve to reduce the unfavorable electrostatic forces between the negative phosphate charges of the three DNA strands [9].

Apart from inhibiting the elongation stage of the unscheduled replication, GAA repeats also increase the efficiency of the unscheduled replication initiation [11]. This is most likely due to the formation of an intramolecular triplex that provides single stranded regions, which serve as a loading lock for the replication initiation proteins to initiate the alternative replication elsewhere. Another possibility is that the formation of a DNA-RNA complex during transcription provides single-stranded DNA regions that could induce unscheduled replication initiation. It is also possible that the unscheduled replication mode is associated with DNA repair as the replication mode was detected at the G1/S border besides the S-phase [11].

Herein, we studied molecular mechanisms of unscheduled replication initiation at the GAA repeat. We proposed the mechanism for the GAA repeats instability in mammalian cells as such: unscheduled replication occurs at the GAA repeat in the absence of a complete and mature chromatin structure through the recruitment of three different replication initiation proteins—DNA polymerase α, origin recognition complex (ORC) and mini chromosome maintenance (MCM) to the DNA in proximity of GAA repeat. Unscheduled replication initiation at the *FXN* gene GAA repeat can promote instability in Friedreich's ataxia. This study is important because through the understanding of the mechanism of unscheduled replication at the GAA repeats, we can have a better insight in the mechanism of the genomic instability at the FRDA GAA repeats.

While most of the research is focused on the instability of the trinucleotides and other microsatellites that undergo expansion in inherited disorders, however, there are also many other repeats in the genome that are unstable as well. It is still unknown whether the mechanisms of

the instability of all the repeats are the same. The instability of repeats of different motifs was evaluated by comparing microsatellites at orthologous locations in human and chimpanzee [22]. Their instability is correlated with the strength of secondary structures that they can adopt. Our lab's recent data showed that the instability of dinucleotide repeats is also correlated with the strength of replication stalling at the repeats positioned in the plasmid in mammalian cells [1]. Dinucleotide repeats instability has been previously explored in our lab, and we concluded that the strength of replication stalling is motif sequence, length and replication mode dependent. For a given length of repeats, the strength of replication stalling in subsequent replication cycle is in the following order: GC/CG > AT/TA > GT/CA > TC/AG. However, in plasmids' first replication cycle upon transfection, TC/AG is the only dinucleotide repeat that causes significant replication stalling [1].

To expand the previous study on dinucleotide repeats, herein, we studied the progression of mammalian replication fork through several tetranucleotides in the first and subsequent replication cycles of transfected plasmids. These tetranucleotides might potentially form secondary structures such as triplex and hairpin structure, which could lead to replication stalling and genomic instability. The results confirmed the previous conclusions from the analysis of replication through dinucleotide repeats: repeats that can adopt triplex structure stalled replication in the absence of a mature chromatin, while repeats that can form hairpins affected replication when chromatin assembly was already established.

Materials and Methods

Bacterial Strain

Cloning of repeat-containing plasmids was carried out in *E.coli* XL1 Blue strain (Stratagene).

Plasmids

A 188 bp blunt-ended EcoRI-BamHI fragment of Yep24 was inserted into the blunt-ended XhoI site of pYES+ to obtain a pYES-control plasmid that does not contain any apparent repeats.

The blunt-ended EcoRI-HindIII fragment from pBluescript-GAA_{57} was inserted into the XhoI site of pYES+ to obtain a pYES-TTC_{57} plasmid.

The blunt-ended EcoRI-HindIII fragment from pBluescript-GAA_{57} was inserted into the blunt-ended Eco81I-site of pYES-Bsg plasmid. The repeat (underlined) was flanked by two BsgI sites in an inverted orientation (shaded) and two EcoRI sites (italicized).

*GAATTC*TGCAGATATCCATCACACTGGCGGCCGCTCGAGTGCAGACCTCAAATTCGA
T(GAA)$_{57}$GATCAAGCTTCAGGTTCTGCACATCGAGCATGCATCTA*GAATTC*.

The pUCneoGAA$_{57}$ and pUCneoCTT$_{57}$ plasmids were generated by the insertion of repeat-containing EcoRI-EcoRI fragments of pYES-GAA$_{57}$ and PYES-CTT$_{57}$ into the blunt-ended AatII site of pUCneo.

pRepControl plasmid was generated by the insertion of HindIII-Eco91I fragment of pRep4 that contains OriP and EBNA-1 gene into the ApaI-NdeI site of pUCneo.

Cloning of tetranucleotides

For tetranucleotide replication analysis, oligonucleotides containing $(AGCT)_{10,}$ $(ACGT)_{10}$, $(AATT)_{10}$, $_{(}GGCC)_{10}$, (CTTT)10, and (CCTT)10 were first cloned and elongated in pYES plasmid as described for GAA 57 repeat above. Plasmid pUCneoH was obtained by inactivating the existing HindIII site of pUCneo by HindIII digestion and religation of the blunt-ended vector, and inserting oligonucleotides containing HindIII at the blunt-ended AatII site. Then, the plasmids pUCneoH$(AGCT)_{10}$, pUCneoH$(ACGT)_{10}$, pUCneoH$(AATT)_{10}$, pUCneoH$_{(}GGCC)_{10,}$ pUCneoH(CTTT)10, and pUCneoH (CCTT)10 were obtained by inserting the corresponding repeat-containing HindIII-HindIII fragments of pYES into the HindIII site of pUCneoH.

Cell cultures and transfections

COS-1 cells, 293-EBNA cells and 293A embryonic kidney fibroblasts cells were grown in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% of fetal bovine serum (FBS) to 50% confluence.

For the analysis of chromatin immunoprecipitation, 293A cells were transfected using Lipofectin (Invitrogen) for 5 hours while 293-EBNA cells were transfected using Turbofect (Fermentas) for 46 hours based on the manufacturer's instructions.

For the analysis of the replication progression in third and subsequent replication cycles, COS-1 cells were transfected using Turbofect (Fermentas) for 30 hours based on the manufacturer's instructions. For the analysis of the first replication cycle, 293A cells were transfected using Lipofectin (Invitrogen) for 5 hours. Following 5 hours of transfection for 293A cells, 10mL of

DMEM supplemented with 10% FBS was added to the transfection mix on the plate and incubated for one hour prior to lysis.

Isolation of replication intermediates

The transfected cells were washed with TBS (50mM Tris-HCl pH 7.0, 150mM NaCl) and lysed in 750 μL of lysis solution (50mM Tris-HCl pH 7.0, 20mM EDTAA, 10mM NaCl, 10% SDS, 0.2mg/mL proteinase K) for 20 min at room temperature. 187 μL of 5M NaCl was added to precipitate the chromosomal DNA overnight at 4 ℃. Next, the chromosomal DNA was centrifuged for 50 min at 15000 rpm at 4 ℃. The supernatant was collected and 5 μL of proteinase K (20mg/mL) was added and incubated at 55 ℃ for 2 hours. Plasmid DNA was purified by phenol/chloroform extraction twice and chloroform extraction once and precipitated with isopropanol and ethanol. The pellets were vacuum-dried and then dissolved in 30 μL of TE buffer.

2-D gel electrophoresis and Southern blotting

10 μL of samples prepared as above were digested with AflIII. The samples were separated in the first dimension gel electrophoresis (0.4% agarose gel, Sigma-Aldrich) in 1x TBE at 1 V/cm for 15 hours, followed by a second dimension gel (1.0% agarose gel) at 4 ℃, 2.5 V/cm for 15 hours.

The gel was transferred to a Zeta-probe nylon membrane (Biorad) based on the manufacturer's instructions. The membrane was cross-linked with the UV light and P^{32}-labeled radioactive probes were obtained by using DecaLabel^{TM} DNA labeling kit (Fermentas), then fragments of interest were hybridized with the radioactive probes. The radioactive signals were visualized on

x-ray film by exposing the labeled membrane to the film at -80 °C for 3 days. Quantitative analysis was made using Storm 860 PhosphorImager with the ImageQuant software (Molecular Dynamics).

Cross-linking and Chromatin Preparation

Proteins were cross-linked to DNA by incubating cells with 353.86 µL of 37% formaldehyde at room temperature for 10 min. 884.8 µL of 2M glycine was added to the media at room temperature for 5 min. After that, the medium was removed and the cells were washed three times with cold 1x phosphate-buffered saline (PBS) supplemented with 1mM PMSF.

The cells were harvested in 1mL of Nuclei Washing Buffer (NWB) (5 mM $MgCl_2$, 10 mM Tris Cl pH 8.0, 0.32 M sucrose, 1mM PMSF) supplemented with 0.5% Triton X-100 and pelleted at 4^{o}C at 3000 rpm for 5 min. The pellet was resuspended in 1 mL of NWB and was again pelleted at 4^{o}C at 3000 rpm for 5 min. The pellet was resuspended in 4mL of IP buffer (0.1 M NaCl, 66.7 mM Tris-Cl pH 8.0, 5 mM EDTA, 0.33% SDS, 1.67% Triton X-100, 1mM PMSF) .

The sample was fragmented at ten 15-sec pulses of 50% amplitude with 1 min rest interval (Fisher Scientific Sonic Dismembrator Model 500).

Chromatin Pre-clearing and Immunoprecipitation

The sample was centrifuged at 4^{o}C at 14000 rpm for 15 min. 2 mL of the supernatant were pre-cleared with 120 µL of 50% slurry Protein A agarose/ Salmon Sperm DNA (Millipore) by rotating for 1-2h at 4°C. The beads were pelleted at 4°C x 4000 rpm x 1min and the supernatants were centrifuged at 4°C x 14000 rpm x 15min to discard aggregates.

1 μg of each antibody (DNA pol α rabbit polyclonal IgG (Santa Cruz Biotechnology), MCM 4 rabbit polyclonal IgG (Santa Cruz Biotechnology) and ORC 2 rabbit polyclonal IgG (Santa Cruz Biotechnology)) was added to each 500 μL of supernatant, respectively. Another 500 μL of supernatant was used as the control without antibodies. All the tubes were incubated overnight with rotation at 4 °C.

The samples were centrifuged at 4°C at 14000 rpm for 15 min. 30 μL of the control sample was collected as "Total input". 15 μL of 50% slurry Protein A agarose/ salmon sperm DNA (Millipore) were added to each of the tube except for "Total input", and rotated at 4°C for 1-2 hours. The beads were pelleted at 4000 rpm at 4°C for 1 min. Supernatants were discarded.

Washes

Collected beads were washed with 1mL of the solution listed in order and rotated gently at 4°C. Three 5 min washes in mixed micelle buffer (0.15 M NaCl, 20 mM Tris-Cl pH 8.0, 5 mM EDTA, 5.2% sucrose, 0.2% SDS, 1% Triton X-100, 1 mM PMSF) were followed by two 5 min washes in buffer 500 (10 mM Tris-Cl pH 8.0, 50 mM HEPES, 0.5 M NaCl, 1% Triton X-100, 0.1% sodium deoxycholate, 1 mM EDTA, 1mM PMSF), two 5 min washes in LiCl/Detergent buffer (10 mM Tris-Cl pH 8.0, 250 mM LiCl, 0.5% NP40 Nonidet, 0.5% sodium deoxycholate, 10 mM EDTA, 1mM PMSF,) and two 5 min washes in TE (10 mM Tris-Cl, pH 8.0, 1 mM EDTA, 1 mM PMSF). The beads were pelleted at 4000 rpm at 4°C for 1 min in between washes.

DNA elution and purification

The pelleted beads were resuspended in 150 μL of SDS/bicarbonate buffer (0.1 M Na HCO_3, 1% SDS) and shaken at setting three of the Vortex-Genie 2 for 15 min. The beads were pelleted at

4000 rpm for 1 min and the supernatants were collected. The previous steps were repeated to elute DNA in another 150 µL of SDS/bicarbonate buffer. 270 µL of SDS/bicarbonate buffer was added to the "Total input" to bring the final volume up to 300 µL. All of the samples were reverse cross-linked at 65℃ overnight.

On the next day, each sample was incubated with 1 µL RNAse at 37℃ for 30 min, followed by the incubation with 1/100 volume of Proteinase K for 1 h at 55℃. DNA was then purified by phenol/chloroform extraction twice and chloroform extraction once.

DNA was precipitated with 1/10 volume of 3M NaOAC, 2 µg of glycogen and 2.5 volumes of 100% ethanol, pelleted at 14000 rpm for 30min at 4℃, and washed twice with 70% ethanol. Then, the samples were dried for 10 min and dissolved in 30 µL of TE; the "Total input" sample was dissolved in 120 µL of TE.

Polymerase Chain Reaction (PCR)

Immunoprecipitated DNA was amplified in a total volume of 25 µL with Maxima Hot Start *Taq*DNA Polymerase (Fermentas) based on the manufacturer's instructions. PCR was performed in a DeltaCycler II system (Ericomp). The following PCR cycle was used: denaturation/ enzyme activation at 95℃ for 2 min, followed by 31 cycles of denaturation at 94℃ for 1 min, annealing at 50℃ for 1 min, and extension at 72℃ for 1 min. The final extension was performed at 72℃ for 10min. The resulting PCR products were resolved by gel electrophoresis using 1.5% agarose gel (Sigma-Aldrich).

Results

Analysis of the recruitment of replication initiation protein by GAA repeats

We used chromatin immunoprecipitation and PCR to analyze the recruitment of replication initiation proteins to different regions of pUCneoGAA57 and control pUCneo plasmids that are transiently transfected in human embryonic kidney cell line 293A. High concentration of DNA fragment observed after the PCR amplification reflects a high concentration of antibody-protein-DNA complex, which would then imply that the respective region recruits the analyzed replication proteins to the plasmid. To confirm the recruitment of replication initiation proteins to the GAA repeat and surrounding regions on the plasmid, PCR amplification results for the region surrounding the GAA repeat were compared to a control plasmid that does not contain GAA or CTT repeats.

The enrichment of replication initiation proteins in each tested region was compared to a negative control (-) (with no antibody added) and the "Total Input" sample (T) that contained the total DNA isolated from transfected 293A cells. This allowed us to analyze whether replication initiation proteins were recruited to other regions of the plasmids, or just to polypurine/polypyrimidine sequences. We analyzed the recruitment of replication initiation proteins to regions surrounding the repeat's position and some other regions of the pUCneo plasmid.

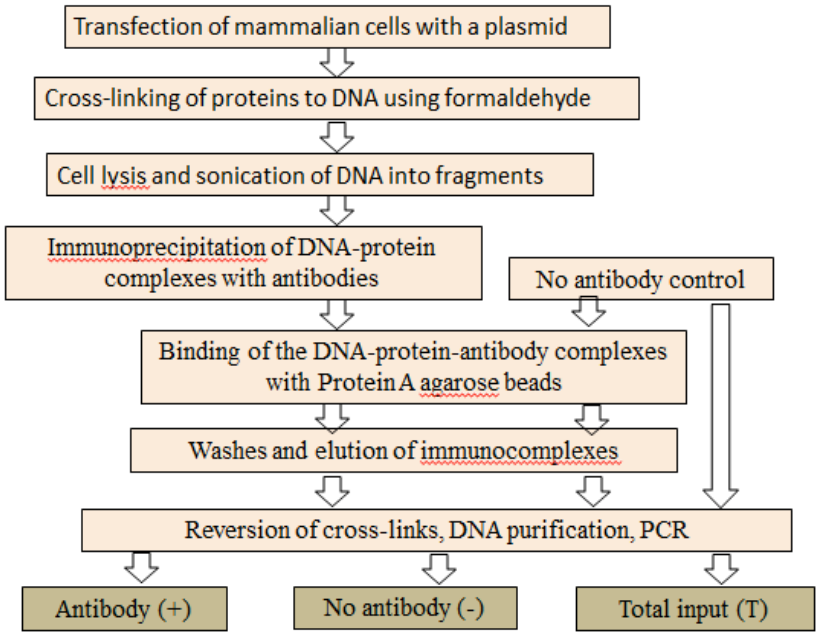


Figure 2: Schematic representation of the use of chromatin immunoprecipitation (ChIP) method to analyze the enrichment of replication initiation proteins.

We first verified that the chromatin immunoprecipitation method is instrumental in analyzing the recruitment of replication initiation proteins by immunoprecipitating the Epstein-Barr virus (EBV)-based plasmid, pREP (isolated from 293-EBNA), with two antibodies: MCM 4 rabbit polyclonal IgG and ORC 2 rabbit polyclonal IgG. We were able to detect MCM4 and ORC2 at the OriP replication origin of pREP (Fig 3). This confirmed that our method worked in analyzing the recruitment of specific proteins to certain DNA regions and that our antibodies were functional.

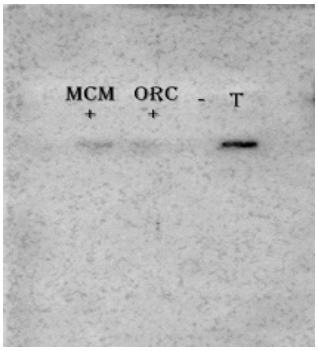


Figure 3: PCR products of the OriP origin region of the pREP plasmid following ChIP with antibodies against MCM 4 and ORC 2. ChIP assay indicated the enrichment of MCM and ORC in the OriP origin region. (+): with antibody, (-): without antibody, T: Total input

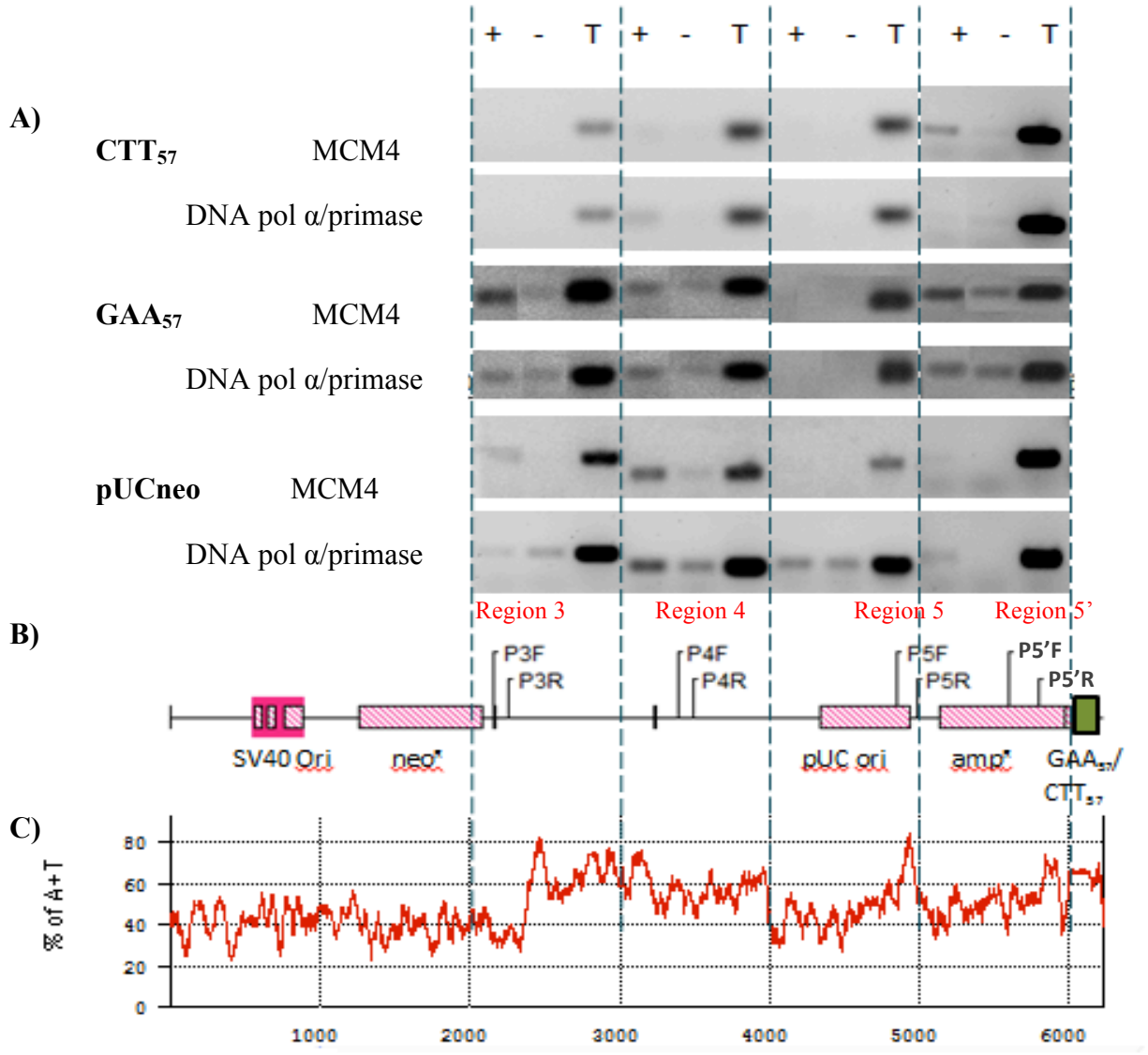


Figure 4: (a) ChIP with antibodies against MCM4 and DNA polymerase α/primase. ChIP assay of four regions (region 3, 4, 5 and 5') in pUCneo CTT_{57}, pUCneo GAA_{57} and pUCneo control. (+): with antibody, (-): without antibody, T: Total input

(b) The map of the pUCneo plasmid with the positions of PCR forward and reverse primers indicated by P*x*F as forward primer and P*x*R as reverse primer; x = 3, 4, 5, 5'. Regions in close proximity with the primers are amplified.

(c) The A/T content of the pUCneo plasmid

ChIP assay showed the enrichment of replication initiation protein, MCM4 and DNA pol α/primase in region 3 (pUCneo GAA_{57} and pUCneo) and region 4 (pUCneo $GAA_{57,}$ pUCneo CTT_{57} and pUCneo). These two regions are associated with the A/T rich regions. No enrichment was observed in region 5 that has lower A/T rich content as compared to region 3 and 4 for all three plasmids.

As compared to the control pUCneo plasmid, pUCneo GAA_{57} showed enrichment of both MCM4 and DNA pol α/primase in region 5', whereas pUCneo CTT_{57} showed enrichment of MCM4 in region 5'. Consistent with our expectation, the presence of polypurine/polypyrimidine GAA sequences recruits replication fork proteins to DNA in proximity to the repeated sequences.

DNA replication stalling within tetranucleotide microsatellites

DNA repeats of different motifs are associated with different strength of replication stalling. Herein, we analyzed the progression of mammalian replication fork through tetranucleotide microsatellites of several motifs cloned in SV40-based plasmid pUCneo. To study the progression of replication fork, plasmids with microsatellite repeats were transiently transfected into either COS-1 or 293A mammalian cells, followed by the isolation of replication intermediates from the cell culture. Intermediates were digested with AflIII, which generated Y-shaped structures with the repeat positioned one-third from the AflIII fragment end (Fig. 5B and Fig. 5C). Replication intermediates were analyzed with two-dimensional neutral agarose gel electrophoresis and followed by the southern blot hybridization. As opposed to 293A cell lines that lack T-antigen, COS-1 cells have T-antigen that could initiate replication at the SV40 origin. 293 A cells were used to model the situation when the chromatin is not fully assembled and when the first replication cycle occurs, whereas COS-1 cells were used to study subsequent replication cycles [10].

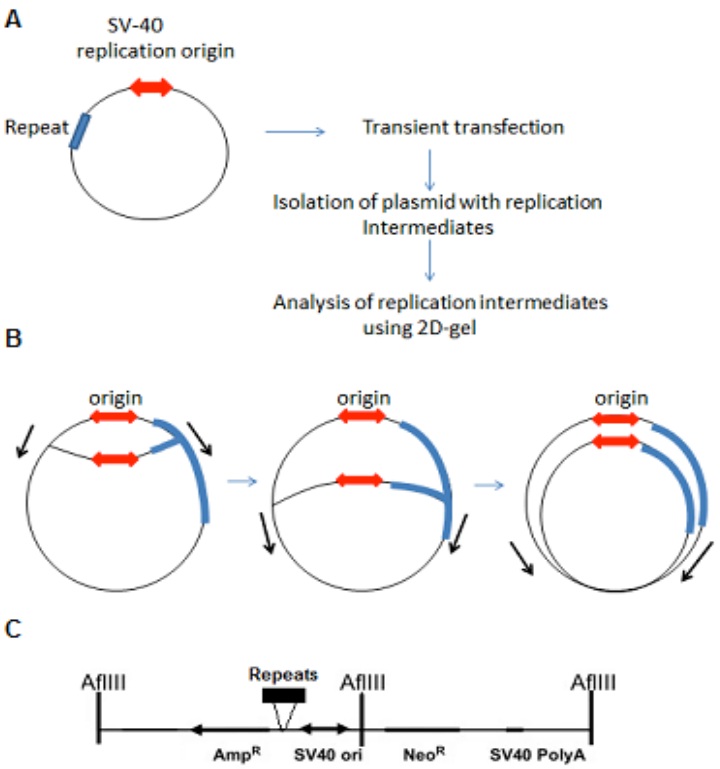


Figure 5: (A) Schematic representation of the experimental procedure for the analysis of replication fork progression. (B) Blue lines indicate regions complementary to the hybridization probe. Replication intermediates forming the Y-arc are shown in the figure as y-shaped blue lines. (C) The positions of AflIII digestion sites and repeats in the SV40-based plasmid.

In the first dimensional gel electrophoresis, replication intermediates were first separated based on their mass. The replication intermediates were then resolved in the second dimensional gel electrophoresis based on their mass and shape (Fig. 6). Digestion with AflIII restriction enzyme would generate Y arc since the replication origin is located outside of the probed region (Fig. 5B). Replication stalling at the repeat would generate a Y-arc with a bulge. The bulge forms because replication stalling would lead to the accumulation of replication intermediates of a particular shape. By hybridizing the fragment of interest with a radiolabeled probe, we could detect the radioactive signal and perform a quantitative analysis of replication stalling. The intensity of the bulge is proportional to the strength of replication stalling.

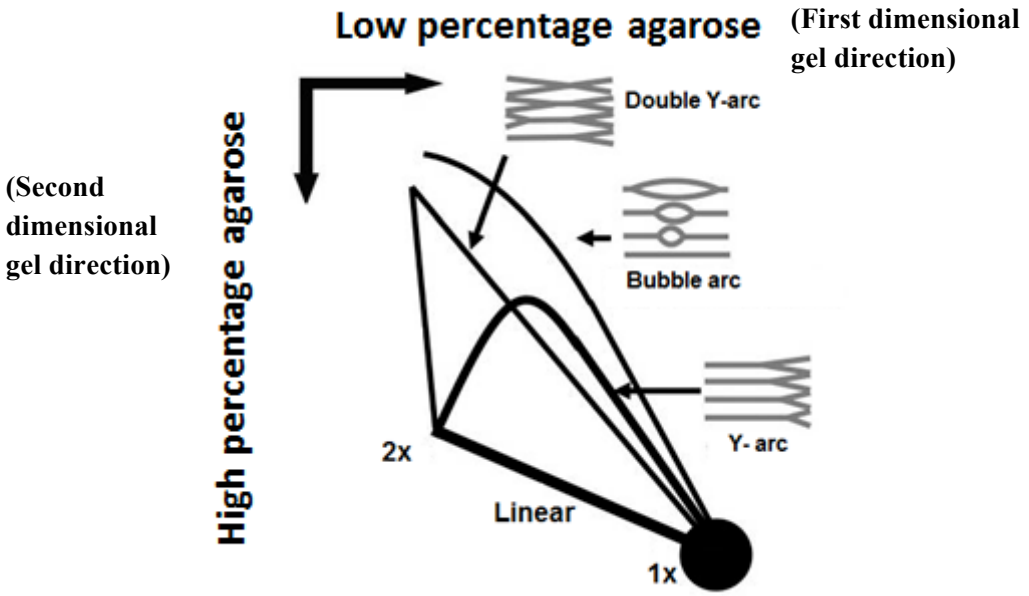


Figure 6: Schematic representation of a 2D-gel electrophoresis of replication intermediates. Replication stalling at the repeat would generate a Y-arc with a bulge.

A) Replication stalling in subsequent replication cycle for self-complementary tetranucleotides that can potentially form hairpin structure

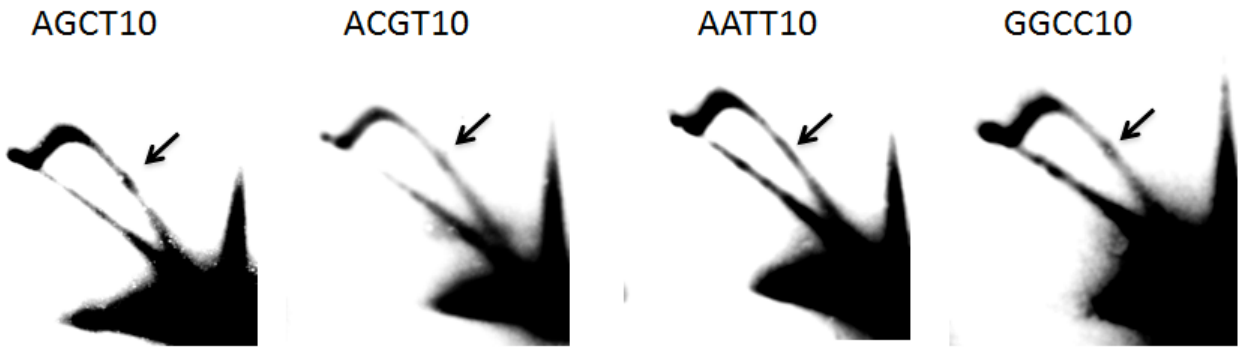


Figure 7: 2D gels of COS-1 cells replication intermediates in subsequent cycle. Replication stalling is indicated by the bulge along the Y-arc (arrows).

The bulge shown on the Y-arc of the replication intermediates of four different motifs, $AGCT_{10}$, $ACGT_{10}$, $AATT_{10}$ and $GGCC_{10}$, suggested replication stalling in mammalian COS-1 cells during third and further replication cycles (Fig. 7). The motifs that we analyzed here were prone to forming self-complementarity hairpin structures. Consistent with our expectation, these motifs caused replication stalling.

B) No significant replication stalling in first replication cycle for self-complementarity tetranucleotides that can potentially form hairpin structure

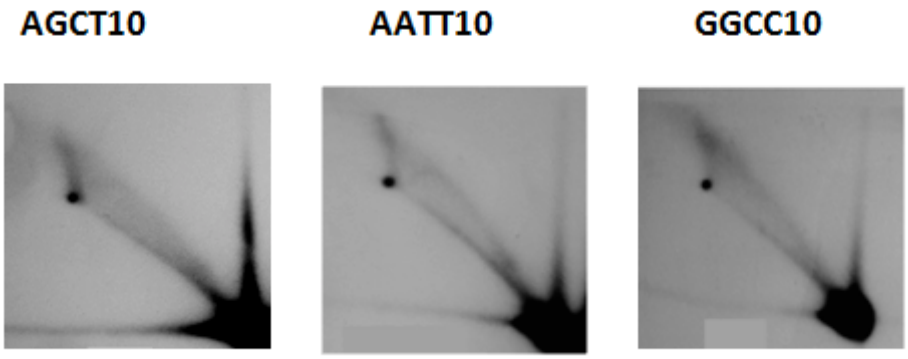


Figure 8: 2D gels of 293A cells replication intermediates in first replication cycle. No significant replication stalling was observed in these three motifs

We also analyzed the replication fork progression of $AGCT_{10}$, $AATT_{10}$ and $GGCC_{10}$ in the first replication cycle. No significant bulge was observed on the Y-arc, which indicates no significant replication stalling (Fig. 8). This result is compared to the tetranucleotide mutability data from the human-chimpanzee microsatellite evolution analysis that will be discussed in Figure 13. Overall, $AGCT_{10}$, $AATT_{10}$ and $GGCC_{10}$ that showed significant replication stalling in subsequent replication cycles gave no stalling effect in the first replication cycle at the repeat length that we had analyzed.

C) No significant replication stalling in third and further replication cycles for tetranucleotides that can potentially form triplex structure

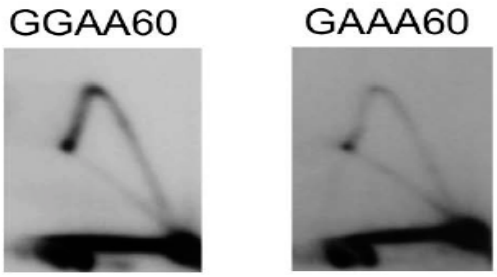


Figure 9: 2D gels of COS-1 cells replication intermediates in subsequent replication cycle. No significant replication stalling was observed. Data is available on https://honors.libraries.psu.edu/paper/13978/

D) Significant replication stalling in first replication cycle for tetranucleotides that can potentially form triplex structure

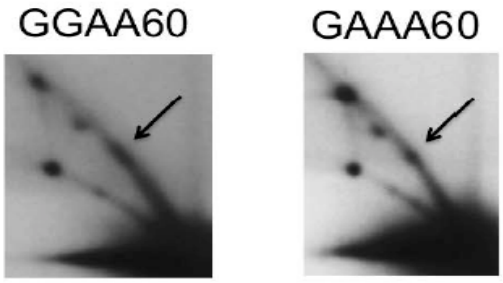


Figure 10: 2D ges of 293 A cells replication intermediates in first replication cycle. The bulge on the Y arc as shown by the arrows indicated the accumulation of signals of replication intermediates, where replication stalling occurred. Data is available on https://honors.libraries.psu.edu/paper/13978/

Different from the previous four tetranucleotide motifs that we have analyzed, $GGAA_{60}$ and $GAAA_{60}$ caused significant replication stalling only in the first replication cycle, and none in subsequent replication cycles. These two motifs can potentially form triplex structures that stalled only the first replication cycle and caused the accumulation of signals of replication intermediates of the same shape, resulting in a bulge at the one-third position of the Y-arc as indicated by arrows in Figure 10.

Replication stalling in microsatellites is dependent on the length of repeats and motifs.

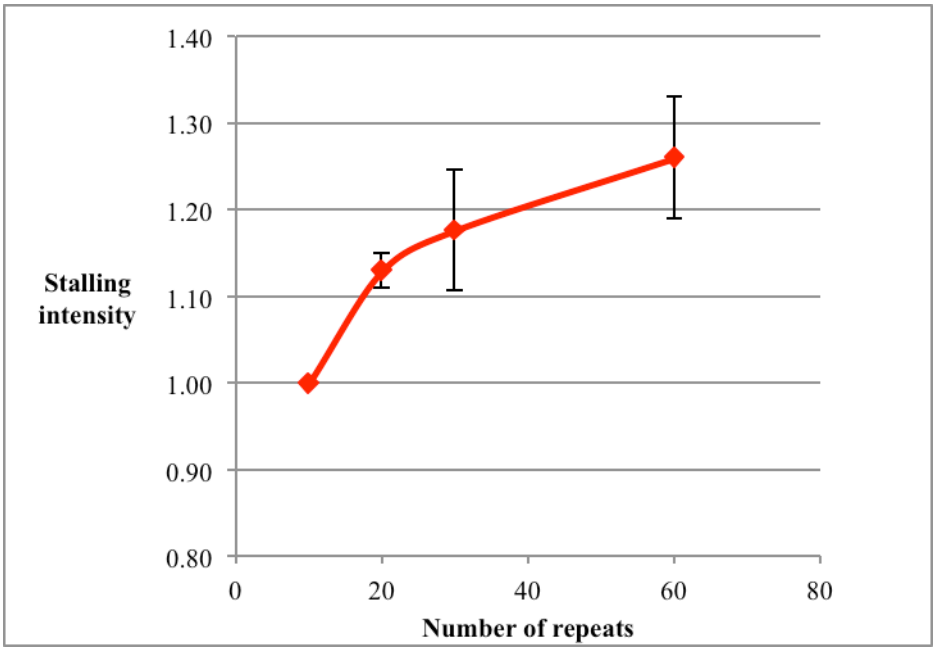


Figure 11: Quantitation of replication stalling in differing lengths of CTTT repeats. Stalling intensity is determined by measuring the intensity of radioactivity in the bulge, relative to the intensity of the arc

The strength of replication stalling is quantified by the measurement of the amount of radioactive signals in the bulge. Y-arcs were analyzed by using phosphorimager and the stalling intensity is determined by the ratio of radioactivity of the bulge in the Y-arc to that of the smooth replication arc. Replication stalling is not visible with $(CTTT)_{10}$: stalling strength equals 1, which means that the arc had the same intensity with the repeats as without them. However, the intensity increases as the number of repeats increases, as indicated in Fig. 11. The data showed that the higher the number of CTTT repeats, the greater the stalling intensity. Hence, the strength of replication stalling promoted by tetranucleotides is length dependent.

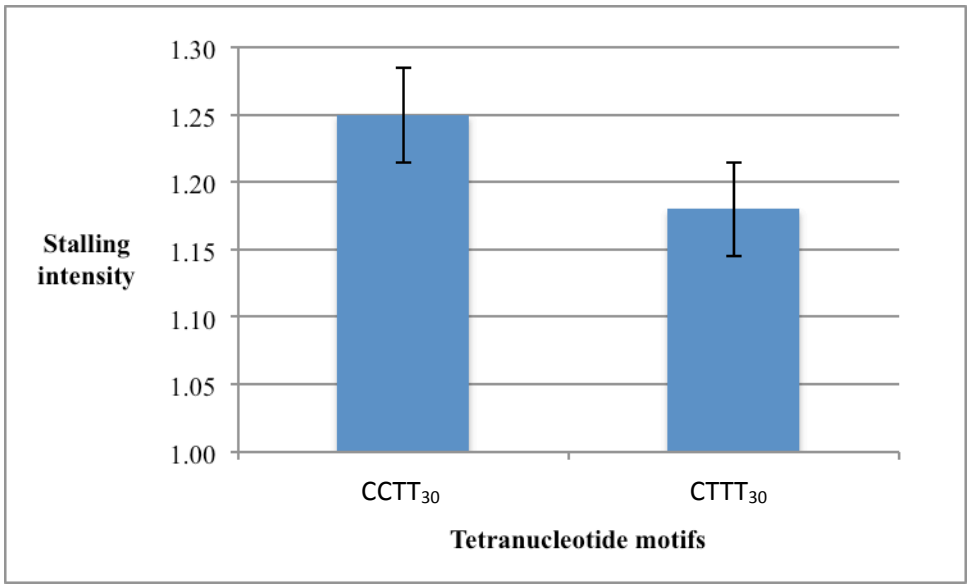


Figure 12: Quantitation of replication stalling in $CCTT_{30}$ and $CTTT_{30}$ repeats

We also analyzed the stalling intensity of different tetranucleotide motifs to determine whether the replication stalling is dependent only on the length, or also the motif itself. The stalling intensity is again determined by the ratio of radioactivity of the bulge in the Y-arc to that

of the smooth replication arc. This chart showed that CCTT caused stronger replication stalling than CTTT at a given length (30 repeats). Hence, we have confirmed that not all tetranucleotide repeats show the same strength of replication stalling. Stalling intensity is both motif and length dependent.

Discussion

Possible mechanism of unscheduled replication promoted by the presence of GAA repeats

Unscheduled replication has been previously detected at genomic regions that are free of nucleosomes. Nucleosome-free DNA occasionally occurs in the human genome, especially during the early zygotic divisions when the chromatin is still restructuring [10]. Replication is initiated there about 10-20 times during S-phase [3]. Nucleosome-free regions are also likely to occur in the hypomethylated regions in cancerous cells [36]. The overreplication of these short regions of DNA is most likely to be caused by the repeated loading of replication fork protein complex, Mcm2-7 complex, at the initiation site with a few cycles of ATP hydrolysis by ORC and Cdc6. The repeated loading of Mcm 2-7 complex results in the replication fork running into the previous ones and causes the expulsion of short and newly synthesized DNA fragments [16]. These fragments can then induce recombination with other genomic regions, which could eventually lead to genomic instability.

The repetitive nature of microsatellites causes them to be more prone to DNA slippage, which could eventually lead to high mutation rates. GAA repeats were proven to be able to form triplexes in vitro and in bacterial cells [24]. Triplex structures are accompanied by single stranded DNA regions that can serve as a loading dock in recruiting replication initiation proteins to initiate the unscheduled replication mode [24]. By using the method of chromatin immunoprecipitation, we have shown that GAA repeats recruited the following replication initiation proteins: ORC2, MCM4, and Pol α/primase (Fig. 5) to the DNA in proximity to the repeats. We therefore hypothesize that the efficiency of unscheduled replication initiation increases at DNA in proximity to the GAA repeats due to the presence of single-stranded regions

that could serve as a loading dock for replication initiation protein. The unscheduled replication mode that occurs during the first 24 hours post transfection is initiated at random positions, and the presence of single-stranded regions could increase the rate of this replication. It has been previously shown that the first replication cycle is not stimulated by DNA breaks or damaged DNA [10]. The increased rate of replication initiation may result in genomic instability that would ultimately lead to expansion events, which can make the symptoms of Friedreich's ataxia even worse in following generations. Active replication in the presence of strong secondary structures has been shown to activate DNA damage checkpoints and promote DNA breakage. When this DNA damage occurs during cell division, the risk of genomic instability greatly increases. [10].

Triplex structure formation in DNA containing such a weak chromatin structure is far more probable, as DNA breathing and single strandedness form more readily in the absence of stable nucleosomes. When taken as a whole, these observations support the hypothesis that unscheduled replication in the presence of a strong triplex structure caused by the GAA repeat causes instability events such as large scale expansions. These expansions can then go on to potentially cause a phenotypic expression of Friedreich's ataxia. Understanding the molecular mechanism of unscheduled replication at Friedreich's ataxia GAA repeats is important as it can shed light on the genomic instability in the patients.

We have also taken into account the A/T content of the four regions that we analyzed. The high A/T rich content is often associated with low internal stability, which could induce the destabilization of DNA double helix. It was previously shown that the unscheduled replication mode could take place in cancerous cells and more than half of the chromosomal translocations

in these cancerous cells are associated with A/T rich regions. Region 3 and 4 are associated with significantly higher A/T content as compared to region 5 and 5'. Hence, these two regions (region 3 and 4) are more likely to recruit the replication fork protein due to their increased nature of flexibility and instability in maintaining the double helix structure. Consistent with our expectation, we have observed enrichment of the replication initiation protein to these two regions even in the absence of GAA or CTT repeated sequences [10].

On the other hand, region 5 and 5' have lower A/T content as compared to region 3 and 4. Hence, no enrichment was observed in these regions in the control pUCneo plasmid. However, in the presence of CTT_{57} or GAA_{57} repeats, the DNAs in proximity to the repeats (region 5') have shown to recruit replication initiation protein MCM4 and DNA polymerase α /primase. Since the plasmids were isolated shortly after transfection, their chromatin was not fully assembled. Overall, the results that we obtained are consistent with our hypothesis: unscheduled replication is initiated at the GAA repeat in the absence of a complete and mature chromatin structure through the recruitment of three different replication initiation proteins—DNA polymerase α, origin recognition complex (ORC) and mini chromosome maintenance (MCM) to the DNA in proximity of the GAA repeat.

Tetranucleotides that can potentially form triplex structure cause significant replication stalling in first replication cycle

Our previous data on the effects of GAA repeats on DNA replication has shown that only the first replication cycle was significantly stalled by the presence of the GAA repeats, whereas the subsequent replication cycles was only mildly affected by it [11]. This clearly indicated that there is a difference between the first and subsequent replication cycles that were affected differently by the same trinucleotide motif. The mechanism of the first replication cycle in mammalian cells is still not understood. However, it is known that during fertilization, maternal and paternal chromatins are intensively remodeled and the chromatin structure of the fertilized egg remains incomplete even after the first zygotic division [17]. Herein, we compared the replication stalling at tetranucleotides that can potentially form triplex structure like the GAA repeats in the first and subsequent replication cycle. We showed that $GAAA_{60}$ caused significant replication stalling in first replication cycle and none in subsequent replication cycles. Similarly, $GGAA_{60}$ caused significant stalling in the first replication cycle and only mild stalling in subsequent cycles. Replication stalling at the subsequent cycle at $GGAA_{60}$ may also be associated with its weak hairpin-forming potential.

Tetranucleotides that can potentially form hairpin structure cause significant replication stalling in subsequent replication cycle

The strength of replication stalling depends on some of the microsatellite's features: the sequence motif, the number of nucleotide repeats and the total length of the repeated stretch. We analyzed the replication fork progression through tetranucleotides of different motifs: $AGCT_{10,}$ $ACGT_{10}$, $AATT_{10}$ and $GGCC_{10,}$ in first and subsequent replication cycles. The self-complementarity of these DNA sequences suggests the potential of these four microsatellites to fold into hairpins structures, which would then result in the replication stalling in the mammalian cells. However, the same tetranucleotides that showed significant replication stalling in the subsequent replication cycle gave no stalling in the first replication cycle.

This is different from the GAAA, GGAA and GAA repeats that can form triplex structure and stall replication only in the first replication cycle and little to none in the subsequent cycle [22]. The main difference between the first replication cycle and subsequent replication cycles is the chromatin state. The triplex structure formation requires the unwinding of a large stretch of DNA to form a stable secondary structure. Thus, a triplex structure is usually not formed in a tightly packed chromatin due to the high nucleation energy barrier. Instead, it is formed during the first replication cycle where the chromatin structure is yet to mature and is still loosely packed [10]. Hence, we proposed that the loose chromatin state in the first replication cycle allows the unwinding of the double helix for the formation of a triplex that would otherwise be impossible to form in subsequent replication cycles.

Hairpin structure is stabilized by the strong hydrophobic interactions of the bases in the hairpin loop that results in negative free energy and exclusion of water from within the loop [29].

The formation of a hairpin requires a shorter stretch of DNA for slippage. DNA slippage is a three-stage process. Firstly, DNA polymerase pauses at the base of the hairpin. It was previously shown that secondary structures are the hot spots for the pausing of DNA polymerase both *in vitro* and *in vivo* [30, 45, 46, 47]. Next, it has been detected that in prokaryotic cells, DNA polymerase dissociates upon pausing because the β-subunit of DNA polymerase III holoenzyme is not able to pass through the secondary structures [45]. Thirdly, replicative DNA polymerase can be replaced by 'translesion polymerases' (TLPs) to bypass the stalling. Eventually, the TLPs are displaced by the replicative polymerase upon the bypass of the block and replication will resume. The single-stranded loop can be deleted in the next round of replication [33]. This shows how replication stalling can lead to deletions and other types of genomic instability in mammalian cells.

Correlation between different tetranucleotides and mutability level

Previous results from Kelkar et al., 2008 have shown the correlation between tetranucleotides of different motifs and different lengths with their mutability level in human-chimpanzee evolution (Fig. 13) [22]. Polypurine/polypyrimidine tetranucleotides, such as repeats of AAGG and AAAG, were shown to have high mutability level. These two microsatellites can potentially form DNA triplex, which lead to replication stalling (Frank 1995). For instance, $(AAGG)_6$ has been demonstrated to form triplex DNA structures in vitro [18]. The stalling intensity showed in Figure 9 and Figure 10 correlated with the high mutability level of AAAG and AAGG repeats. In addition, consistent with the human-chimpanzee data analysis, the $(AATT)_n$ microsatellite that we have tested caused only a very small stalling in the first replication cycle and significant stalling effect on the subsequent replication cycles. As shown in

Figure 13, tetranucleotides of the same motif but different lengths showed different mutability level. All of the motifs were shown to have increased mutability with each increase in repeat number. This is consistent with the result that we obtained, at least up to the length analyzed: the higher the number of CTTT repeats, the greater the stalling intensity,

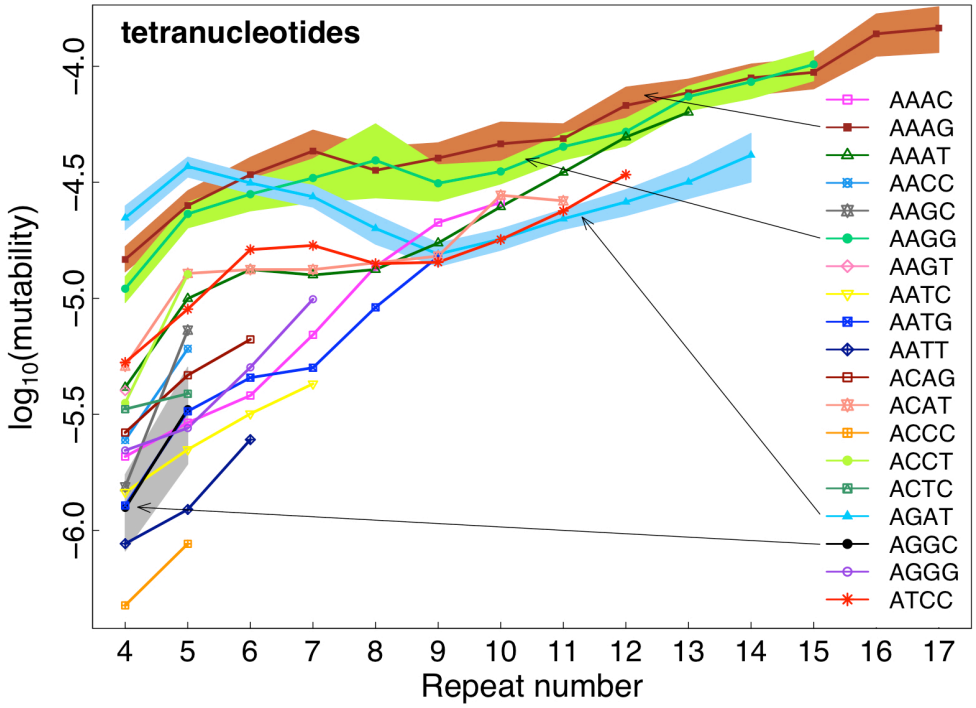


Figure 13: Correlation between the length and motifs on mutability. AATT tetranucleotide has relatively low mutability level as compared to other motifs.

In summary, the triplex structure at homopurine/homopyrimidine repeats can have a dual effect on the replication: First, it makes the unscheduled replication much more intense, and second, the triplex structures would nearly block the unscheduled replication once it has started. However, hairpin-forming, self-complementary microsatellites have almost no effect on the first replication cycle, but strongly affect the replication in subsequent cycles and in the presence of nucleosomes.

Overall, the tetranucleotide repeats that affect the first and not the subsequent cycles were also extremely unstable in primate evolution. It is possible that DNA triplex formation in early embryogenesis leads to genomic instability. However, more future studies and research are necessary to directly show the effect of those repeats on chromosomal replication in their natural environment.

Bibliography

1. Baptiste B. A., Ananda G., Strubczewski N., Lutzkanin A., Khoo S. J., Srikanth A., Kim N., Makova., K. D., Krasilnikova M. M., Eckert K. A. 2013. Mature microsatellites: mechanisms underlying dinucleotide microsatellite mutational biases in human cells. G3 (Bethesda). 3: 451-463.
2. Bit-Avragim N., Perrot A., Schöls L., Hardt C., Kreuz F.R., Zühlke C., Bubel S., Laccone F., Vogel H., Dietz R., Osterziel K.J. 2001. The GAA repeat expansion in intron 1 of the frataxin gene is related to the severity of cariac manifestation in patients with Friedreich's ataxia. J. Mol. Med. 78:626-632.
3. Blow J.J., Gillespie P. J. 2008. Replication licensing and cancer—a fatal entanglement. Nat. Rev. Cancer. 8: 799-806.
4. Bochman M.L., Paeschke K., Zakian V. A. 2012. DNA secondary structures: stability and function of G-quadruplex structures. 13: 770-780.
5. Bouffler S., Silver A., Cox R. 1993. The role of DNA repeats and associated secondary structures in genomic instability and neoplasia. Bioessays. 15 (6): 409-412.
6. Campuzano,V., Montermini,L., Lutz,Y., Cova,L., Hindelang,C., Jiralerspong,S., Trottier,Y., Kish,S., Faucheux,B., Trouillas,P. et al. (1997) Frataxin is reduced in Friedreich ataxia patients and is associated with mitochondrial membranes. Hum. Mol. Genet. 6: 1771–1780
7. Campuzano V, Montermini L, Moltò MD, Pianese L, Cossée M, Cavalcanti F, Monros E, Rodius F, Duclos F, Monticelli A,Zara F, Cañizares J, Koutnikova H, Bidichandani SI, Gellera C, Brice A, Trouillas P, De Michele G, Filla A, De Frutos R,Palau F, Patel PI, Di Donato S, Mandel JL, Cocozza S, Koenig M, Pandolfo M. 1996. Friedreich's Ataxia:

Autosomal Recessive Disease Caused by an Intronic GAA Triplet Repeat Expansion. Science. 271 (5254): 1423-1427.

8. Campuzano V., Montermini L., Lutz Y., Cova L., Hindelang C., Jiralerspong S., Trottier Y., Kish S.J., Faucheux B., Trouillas P., Authier F.J., Durr A., Mandel J.L., Vescovi A., Pandolfo M., Koenig M. 1997. Frataxin is reduced in Friedreich ataxia patients and is associated with mitochondrial membranes. Hum. Mol. Genet. 6(11):1771-1780.
9. Chan P. P., Glazer P. M. 1996. Triplex DNA: fundamentals, advances, and potential applications for gene therapy. J. Mol. Med. 75: 267-282.
10. Chandok G.S., Kapoor K.K., Brick R.M., Sidorova J.M., Krasilnikova M.M. 2011. A distinct firstreplication cycle of DNA introduced in mammalian cells. Nucleic Acids Res. 39(6): 2103-2115.
11. Chandok G.S., Patel M.P., Mirkin S.M., Krasilnikova M.M. 2012. Effects of Friedreich's ataxia GAA repeats on DNA replication in mammalian cells. Nucleic Acid Res. 40 (9): 3964-3974.
12. Cummings C.J., Zoghbi H.Y. 2000. Trinucleotide Repeats: Mechanisms and Pathophysiology. Annu. Rev. Genomics Hum. Genet. 1:281-328.
13. Daee D. L., Mertz T., Collins N., Lahue R. S. 2007. Post-replication repair inhibits CAG·CTG repeat expansions in *Saccharomyces cerevisiae.* Mol Cell Biol. 27: 102-110.
14. Frank-Kamenetskii M.D. 1995. Triplex DNA structures. Annu. Rev. Biochem. 64 (1): 65-95.
15. Glickman B.W., Ripley L. S. 1984. Structural intermediates of deletion mutatgenesis: A role for palindromic DNA. Proc. Natl. Acad. Sci. 81: 512-516.

16. Gomez M., Anteguera F. 2008. Overreplication of short DNA regions during S phase in human cells. Genes & Development. 22: 375-385.
17. Gottesfeld J.M., Pandolfo M. 2009. Development of histone deacetylase inhibitors as therapeutics for neurological disease. Future Neurol. 4: 775-784.
18. Hanvey J. C., Klysik J., Wells R. D. 1988. Influence of DNA sequence on the formation of non-right-handed helices in oligopurine.oligopyrimidine inserts in plasmids. J. Biol. Chem. 263 (15): 7386-7396.
19. Harding A.E. 1981. Friedreich's Ataxia: A clinical and genetic study of 90 families with an analysis of early dagnostic criteria and intrafamilial clustering of clinical features. Brain. 104: 589-620.
20. Hastings P.J.m Ira G., Lupski J.R. 2009. A microhomology-mediated break induced replication model of the origin of human copy number variation. PLoS Genet. 5(1): e1000327
21. Kato T., Inagaki H., Yamada K., Kogo H., Ohye T., Kowa H., Nagaoka K., Taniguchi M., Emanuel B., Kurahashi H. 2006. Genetic variation affects de novo translocation frequency. Science. 311: 971
22. Kelkar Y.D. 2008-01. The genome-wide determinants of human and chimpanzee microsatellite evolution. Genome research. 18(1): 30-38.
23. Kim J. C., Mirkin S. M. 2013. The balancing act of DNA repeat expansions. Curr. Opin. Genet. Dev. 23: 280-288.
24. Krasilnikova, M. M., Mirkin S.M. 2004. Replication stalling at Friedreich's ataxia (GAA)n repeats in vivo. Molecular and cellular biology, 24(6): 2286-95.

25. Krasilnikova M.M. 2012. Complexes between two GAA repeats within DNA introduced into Cos-1 cells. Mol. Genet. Element. 2(6): 267-271.
26. Koole W., Schafer H.S., Agami R., Haaften G., Tijsterman M. 2013. A versatile microsatellite instability reporter system in human cells. Nucl. Acids Res. 1-9.
27. Koutnikova H., Campuzano V., Foury F., Dolle P., Cazzalini O., Koenig M. 1997. Studies of human, mouse and yeast homologues indicate a mitochondrial function for frataxin. Nat. Genet. 15(4): 345-351.
28. Kozlowski P., Mezer M., Krzyzosiak W.J. 2010. Trinucleotide repeats in human genome and exome. Nucl. Acids Res. 38 (12): 4027-4039.
29. Kuznetsov S. V., Shen Y., Benight A. S., Ansari A. 2011. A semiflexible polymer model applied to loop formation in DNA hairpins. Biophys. J. 81: 2864-2875.
30. LaDuca R. J., Fay P. J., Chuang C., McHenry C. S., Bambara R. A. 1983. Site-specific pausing of deoxyribonucleic acid synthesis catalyzed by four forms of *Escherichia coli* DNA polymerase III. Biochemistry. 22: 5177-5188.
31. Li Y., Korol A. B., Fahima T., Beiles A., Nevo E. 2002. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Mol. Ecol. 11: 2453-2465.
32. Li Y., Korol A. B., Fahima T., Nevo E. 2003. Microsatellites within genes: structure, function, and evolution. Mol Biol Evol. 21 (6): 991-1007.
33. McMurray C.T. 2010. Mechanisms of trinucleotide repeat instability during human development. Nat. Rev. Genet. 11(11): 786-799.
34. Orr H.T., Zoghbi H.Y. 2007. Trinucleotide Repeat Disorders. Annu. Rev. Neurosci. 30: 575-621

35. Pandolfo M. 2002. The molecular basis of Friedreich ataxia. Adv. Exp. Med. Biol. 516: 99–118.

36. Portela A., Liz J., Nogales V., Setien F., Villanueva A., Esteller M. 2013. DNA methylation determines nucleosome occupancy in the 5'-CpG islands of tumor suppressor genes. Oncogene. 32: 5421-5428.

37. Pragna I. P., Grazia Isaya. 2001. Friedreich Ataxia: From GAA triplet-repeat expansion to frataxin deficiency. American J. Hum. Genet. 69 (1): 15-24.

38. Rajeswari M.R.2012. DNA triplex structures in neurodegenerative disorder Friedreich's ataxia. J. Biosci. 37(3) 519-532.

39. Sakamoto N., Chastain P.D., Parniewski P., Ohshima K., Pandolfo M., Griffith J.D., Wells R.D. 1999. Sticky DNA: Self-association properties of long GAA·TTC repeats in R·R·Y tripex structures from Friedreich's Ataxia. Molecular Cell. 3: 465-475.

40. Schöls L., Amoiridis G., Przuntek H., Frank G., Epplen J. T., Epplen C. 1997. Friedreich's ataxia Revision of the phenotype according to molecular genetics. Braom/ 120: 2131-2140.

41. Shapiro J. A., von Sternberg R. 2004. Why repetitive DNA is essential to genome function. Biol. Rev. 80: 227-250.

42. Sharma R., Bhatti S., Gomez M., Clark R.M., Murray C., Ashizawa T., Bidichandani S.I. 2002. The GAA triplet-repeat sequence in Friedreich ataxia shows a high level of somatic instability in vivo, with a significant predilection for large contractions. Hum. Mol. Genet. 11(18):2175-2187.

43. Shibata Y., Kumar P., Layer R., Willcox S., Gagan J.R., Griffith J.D., Dutta A. 2012. Extrachromosomal MicroDNAs and Chromosomal Microdeletions in Normal Tissues. Science. 336 (6077):82-86.
44. Spinaci M., Seren E., Mattioli M. 2004. Maternal chromatin remodeling during maturation and after fertilization in mouse oocytes. Mol. Reprod. Dev. 69: 215-221.
45. Viguera E., Canceill D., Ehrlich S. D. 2001. Replication slippage involves DNA polymerase pausing and dissociation. The EMBO Journal. 20 (10): 2587-2595.
46. Weaver D. T., De Pamphilis M.L. 1982. Specific sequences in native DNA that arrest synthesis by DNA polymerase α. J. Biol. Chem. 257: 2075-2086.
47. Weaver D. T, DePamphilis M. L. 1984. The role of palindromic and non-palindromic sequences in arresting DNA synthesis *in vitro* and *in vivo*. J. Mol. Biol. 180: 961-986.
48. Wells R. D. 2007. Non-B DNA conformations, mutagenesis and disease. Trends Biochem. Sci. 32: 271-278.
49. Zhao J., Bacolla A., Wang G. Vasquez K.M. 2010. Non-B DNA structure-induced genetic instability and evolution. Cell Mol. Life Sci. 67(1):43-62

ACADEMIC VITA

Teo, Yee Voan
2620 Plaza Drive,
State College, Pennsylvania 16801.
yvonne_tyv@yahoo.com

Education

- B.S. in Microbiology, The Pennsylvania State University
  - Minor in Biochemistry and Molecular Biology — August 2011 – May 2014

Academic Achievements

- Scholar of Schreyer's Honors College, The Pennsylvania State University
- Consecutive Dean's list from Fall 2011 to Fall 2013

Related Experiences

- **Undergraduate independent research**, **Biochemistry and Molecular Biology Department** — May 2012 – May 2014
  - Performed research in analyzing the effect of microsatellite repeats on neurological diseases and the mechanism of Friedreich's Ataxia disease
  - Learned important laboratory research skills such as gel electrophoresis, Southern blot, DNA isolation
  - Learned to maintain an organized and detailed lab notebook
- **Laboratory assistant, Biochemistry and Molecular Biology Department** — September 2012-August 2013
  - Performed technical and non-technical procedures such as preparation of media and autoclave to assist the research work of graduate students
- **Teaching assistant, Biochemistry and Molecular Biology Department** — August 2013 – May 2014
  - Taught Elementary Biochemistry Laboratory BMB 212
  - Prepared quizzes and graded laboratory reports
  - Gained skills in tutoring students and conducting class

Fellowships and awards

- **2013 Undergraduate Summer Discovery Grant**
  - Performed full-time summer research in analyzing the mechanism of Friedreich's Ataxia disease
- **Malaysian Public Service Department scholarship**
  - Received full scholarship for undergraduate study in the United States

Poster Presentations

- The 2013 Undergraduate Exhibition — 2013
- 4th Annual Retreat for the Center for Medical Genomics — 2013

Skills

- **Laboratory**
  - Fluorescence in situ hybridization
  - PCR
  - Chromatin immunoprecipitation
  - Tissue culture
  - DNA isolation and purification
  - Southern blot
  - Gel electrophoresis
  - Media preparation
- **Languages**
  - English, Mandarin, Malay (Full professional proficiency)
- **Computer**
  - proficient in Microsoft Word, Excel and Power Point

Volunteer Experiences

- **Volunteer staff at 32nd American Society for Virology Conference** — July 2013
  - Helped in coordinating afternoon and evening workshops
- **Volunteer at Kuala Lumpur Relay for Life 2010, Malaysia** — 2010
  - Helped in coordinating an overnight team cancer event