DNA Polymerase Navigation through Difficult-to-Replicate Sequences to Prevent Genome Instability

Kristin Eckert, Ph.D.
Professor of Pathology and Biochemistry & Molecular Biology

Penn State University

NIH DNA Repair Interest Group
December 21, 2021

GENOME INSTABILITY

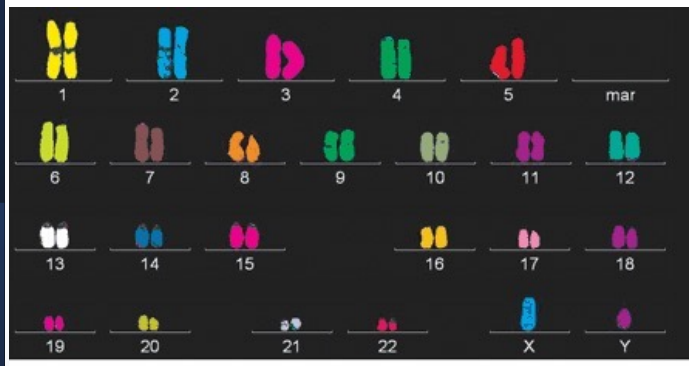
Genetic variants created randomly throughout the genome fuel cancer cell evolution

Types of Genetic Variation in Tumors

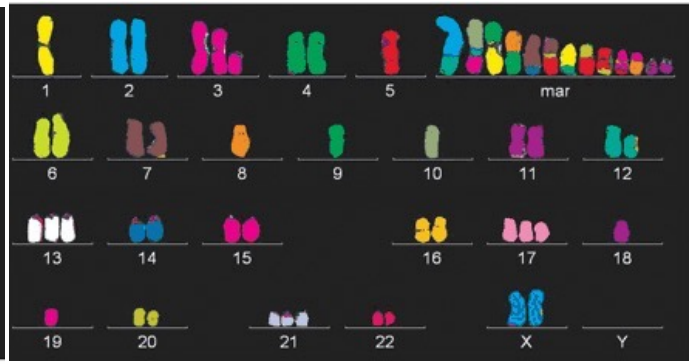
Small scale changes:

p53 in skin cancer	TGFβ RII in colon cancer
Met Asn Arg Arg	Glu Lys Lys Lys Pro Gly
ATG AA**C** **C**GG AGG	GAA AAA AAA **AA**G CCT GGT
↓	Δ
ATG AA<u>T</u> <u>T</u>GG AGG	GAA AAA AAA GCC TGG TGA
Met Asn Trp Arg	Glu Lys Lys Ala Trp *Stop*

Large scale changes (structural variation):



Normal cell: Diploid

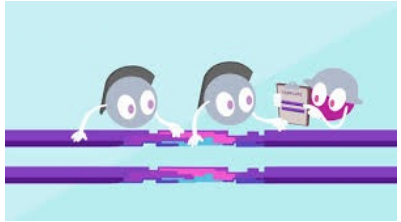


Cancer cell: Aneuploid; marker chromosomes

Cancer etiology: chemical carcinogenesis

Cell environment (inside or outside)

Damaged DNA



DNA Repair

-OR-

Cell Death

Prevents the accumulation of variant cells with mutations

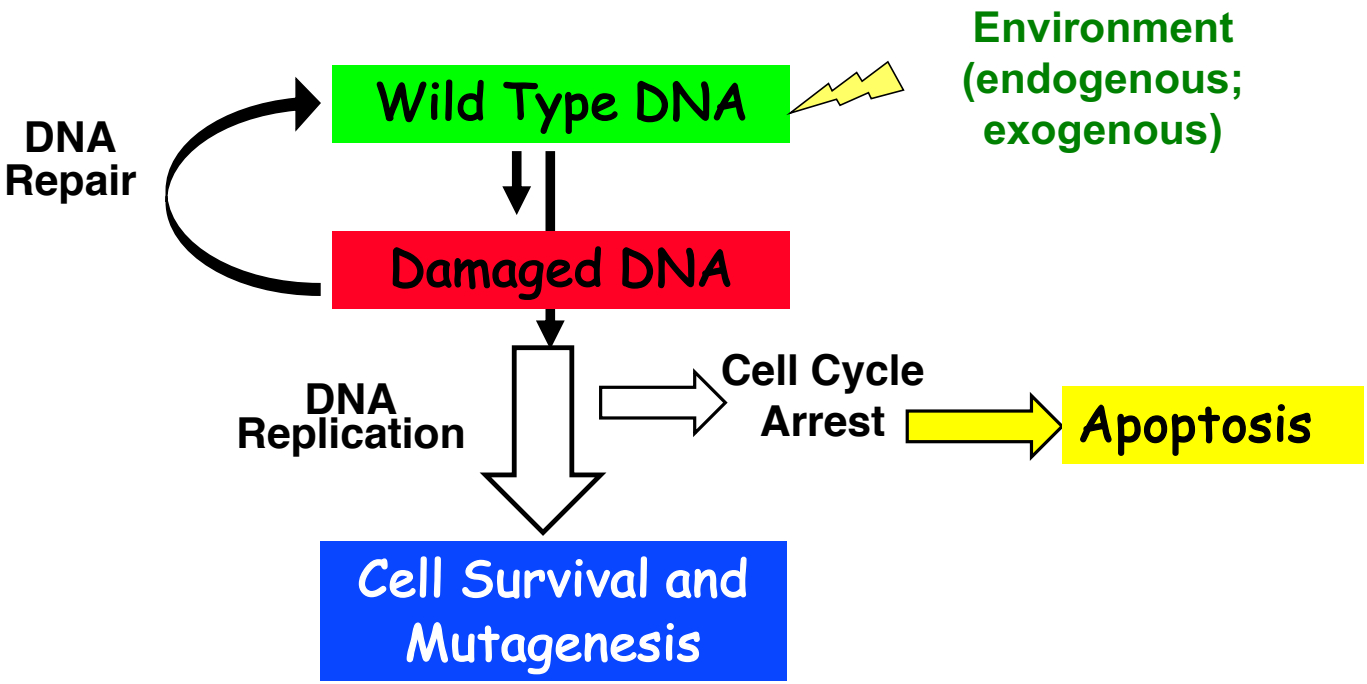
Cell Survival

-and-

Mutation

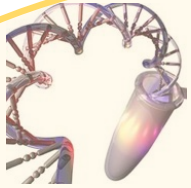
Creates Genetic Variants

Making Mutations is an Active Cellular Process

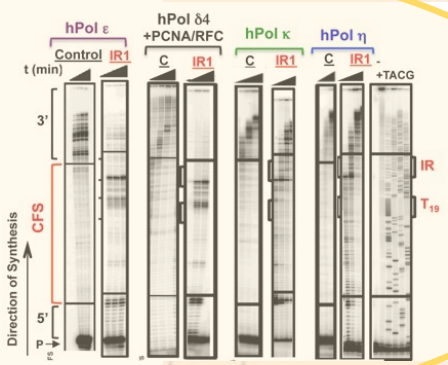


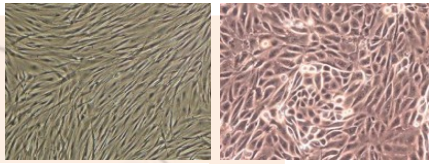
DNA polymerases are key determinants of "replication errors" during tumor cell evolution

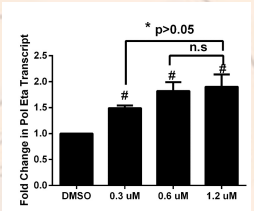
Eckert lab research: Experimental Approaches

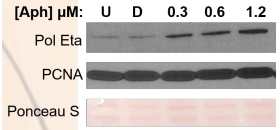


In vitro
DNA polymerase biochemistry

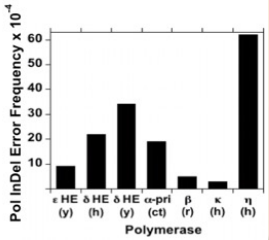


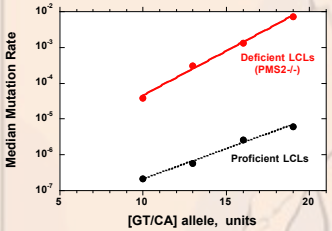


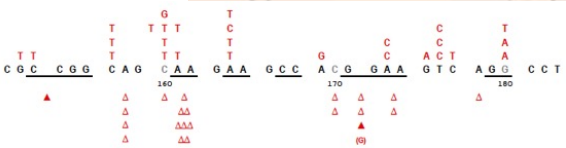




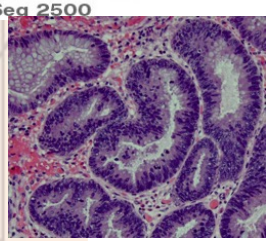
***Ex vivo* cell culture**



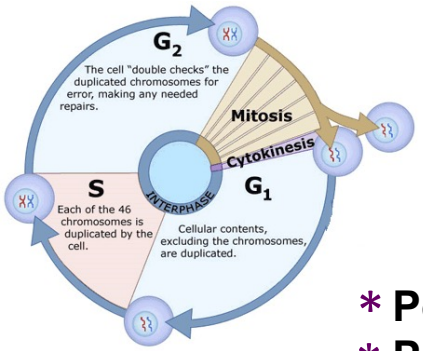




Molecular Genetics



Genomics



Multiple DNA polymerases are required to maintain human genome stability

Replication

* Pol alpha (α, *POLA*)
* Pol delta (δ, *POLD1*)
* Pol epsilon (ε, *POLE*)

Repair

* Pol beta (β, *POLB*)
  Pol lambda (λ, *POLL*)
  Pol mu (μ, *POLM*)
  Pol theta (θ, *POLQ*)

* Studied in Eckert lab

Specialized Synthesis

* Pol zeta (ζ, *POLZ; REV3*)
* Pol eta (η, *POLH; XPV; Rad30*)
* Pol kappa (κ, *POLK; DINB1*)
* Pol iota (ι, *POLI*)
  Rev 1 (dCMP transferase)
  Pol nu (ν, *POL N*)
* PrimPol (*CCDC111*)

Mitochondrial

Pol gamma (γ, *POLG*)

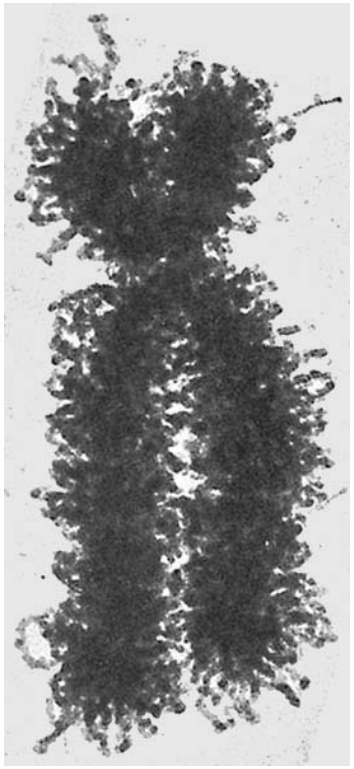
DNA polymerase activities must be orchestrated to ensure complete genome duplication before mitosis

What is the sequence composition of the human genome?

~60% of the human genome sequence is repetitive DNA

Repetitive Sequences are found throughout Genomes

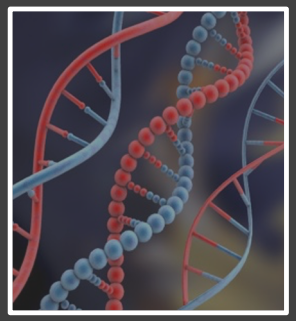
Telomeres	$[TTAGGG]_n$
Rare fragile sites	$[CGG/GCC]_n$ $[CAG/GTC]_n$
Common fragile sites (CFS)*	$[A/T]_n$ $[AT/TA]_n$
	Quasi-palindromes
Mobile Elements*	G4 motifs



*** Locations of recurrent structural variation in cancer cells**

What is the DNA structure within the human genome?

Random Sequences



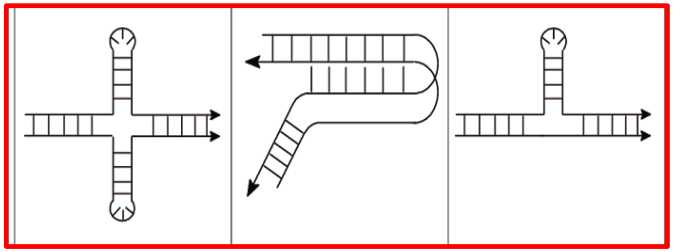
B-DNA

Repetitive sequences: Non B DNA

Cruciforms $(AT)_n$

Triplex $(TC)_n$ $(AAG)_n$

Hairpins $(AT)_n$ $(CAG)_n$

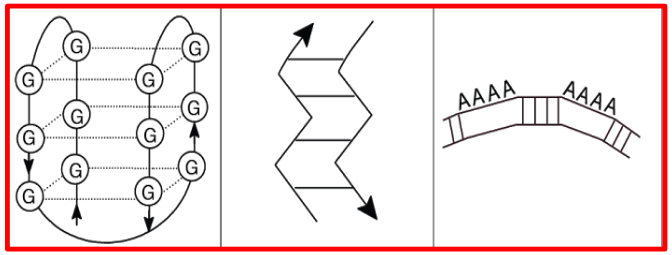


- Impact human genome mutagenesis
- Can be difficult to replicate

G4 $(TTAGGG)_n$

Z-DNA $(GT)_n$ $(GC)_n$

Bent DNA $(A)_n$ $(T)_n$



DiToRS: Difficult-To-Replicate Sequences

DNA polymerases—in the driver's seat

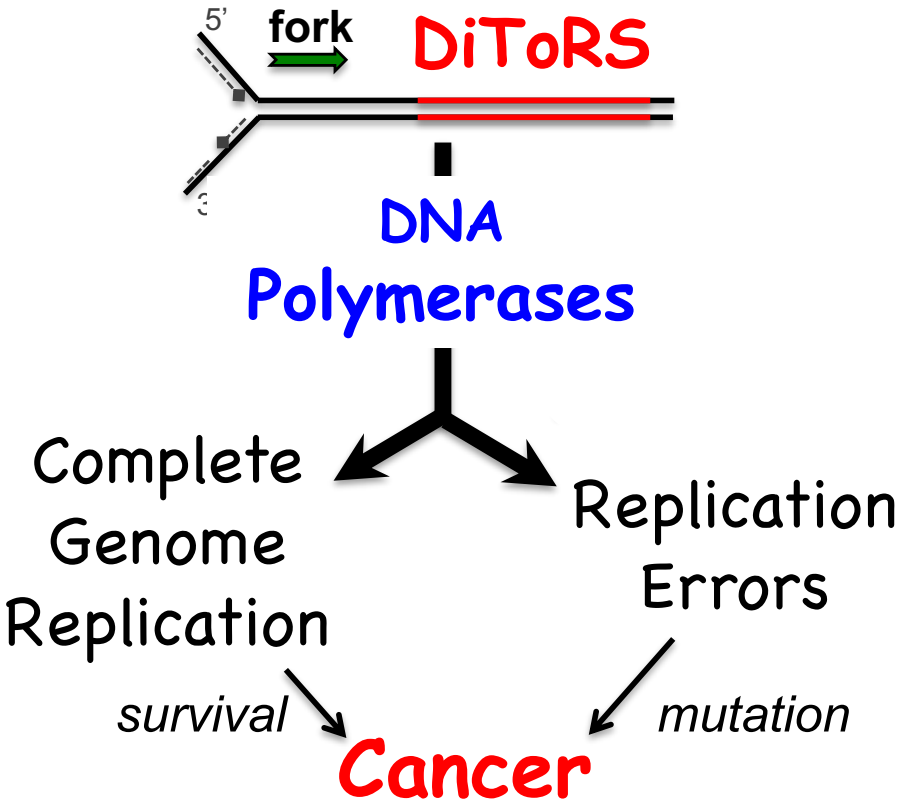
DNA polymerase inhibition and errors during replication

Chromosome breakage and mutation

What constitutes a "DiToRS" ?

What is the complete landscape of DiToRS in the human genome ?

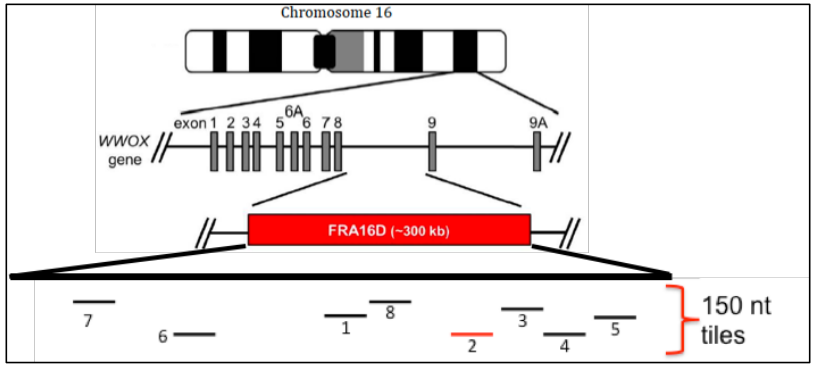
When do DiToRS become threats to genome stability in cancer cells ?



Defining DiToRS: Using FRA16D as a model locus

FRA16D locus:

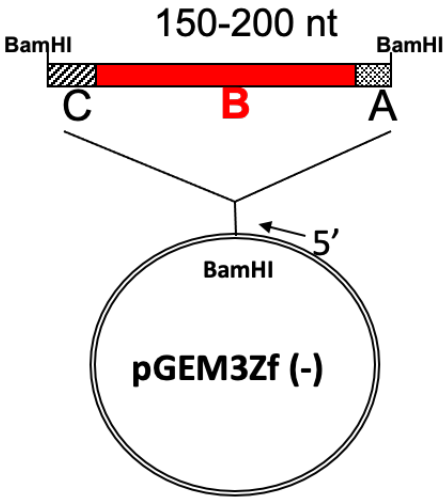
- One of the most "fragile" Common Fragile Sites
- ~300 kb in length
- Resides within an intron of the WWOX tumor suppressor gene
- Site of recurrent deletions in tumor cells



Template	CFS	Repeat Elements*
AT repeat 1	FRA16D	$[A/T]_{28}$, $[AT/TA]_{23i}$
AT repeat 2	FRA16D	$[A/T]_{9}$, $[AT/TA]_{8}$
AT repeat 4	FRA16D	$[AT/TA]_{9}$
Quasipalindrome 1	FRA16D	$[A/T]_{19}$, QP_{36}
Quasipalindrome 2	FRA16D	QP_{37}
Quasipalindrome 3	FRA16D	QP_{29}

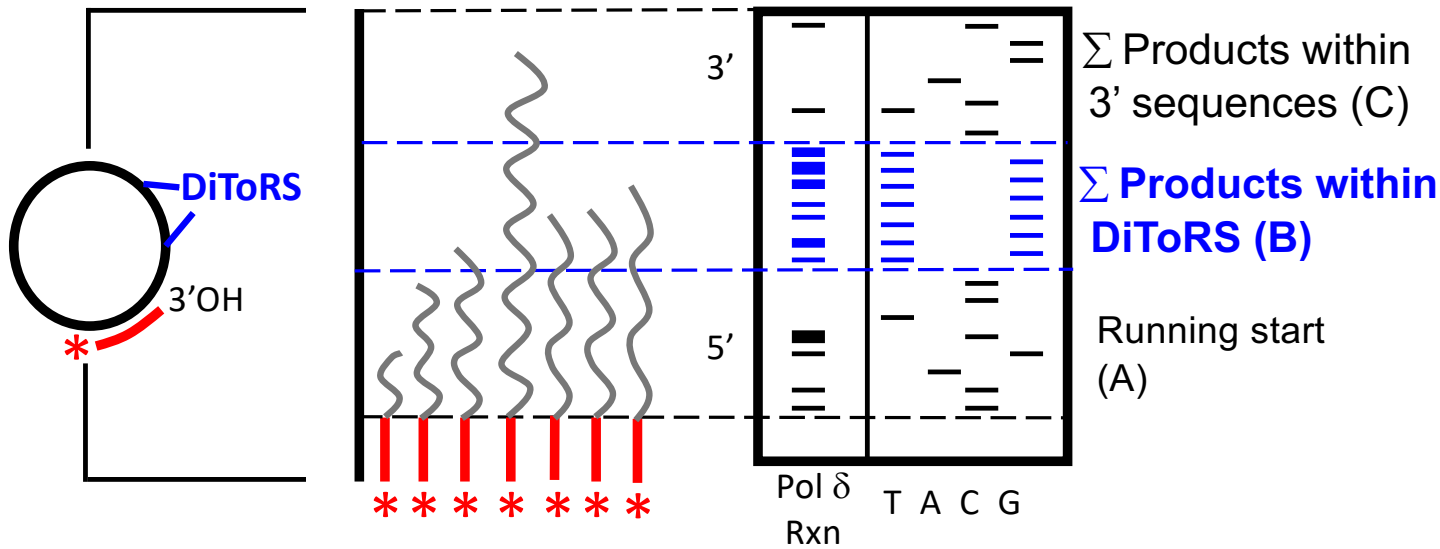
*** Sequences include known breakpoints observed in tumor cells**

Experimental Approach to Define DiToRS: Quantitative Assay with Nucleotide Resolution



A = constant proximal
B = DiToRS
C = constant distal

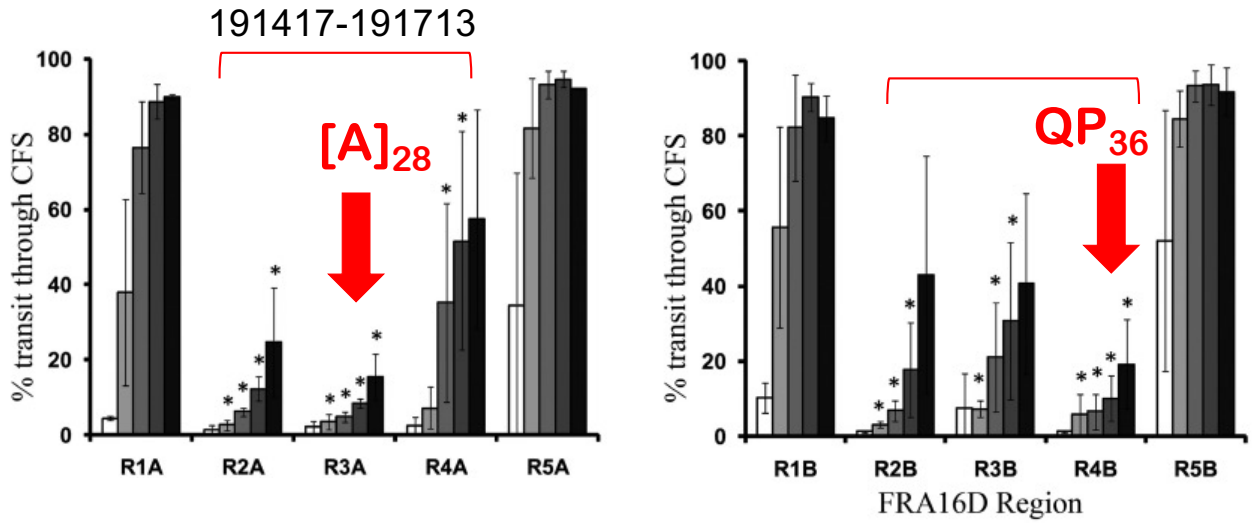
Control FRA16D sequence:
76 % A+T content
No repeat elements



* ^{32}P radioactive label; amount of radioactivity corresponds to number of DNA molecules

Shah et al (2010) Nucl. Acids Res.38: 1149.

Replicative Human Pol δ displays slowed elongation within Repetitive FRA16D Regions

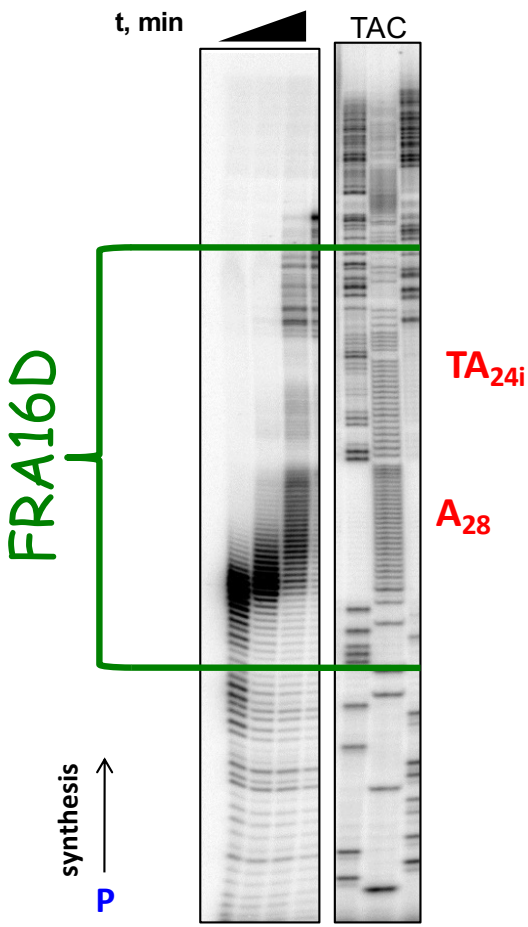


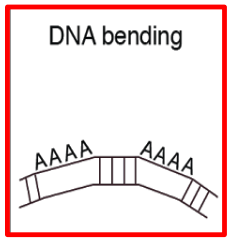
- Slowed DNA synthesis within specific sequences
- Strand biases observed
- Similar results for Pol α-primase

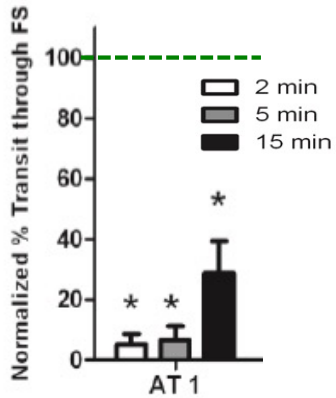
Shah et al (2010) Nucl. Acids Res.38: 1149.

Human Pol δ Synthesis Pausing within A/T Microsatellites

AT-1 (Flex5)

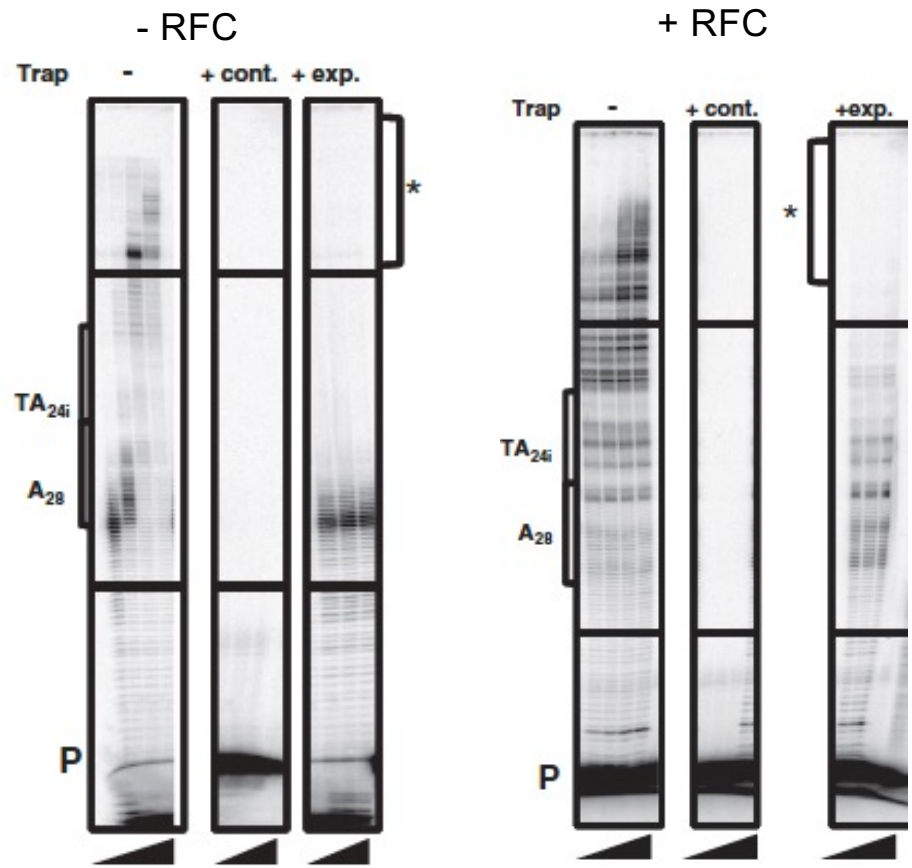






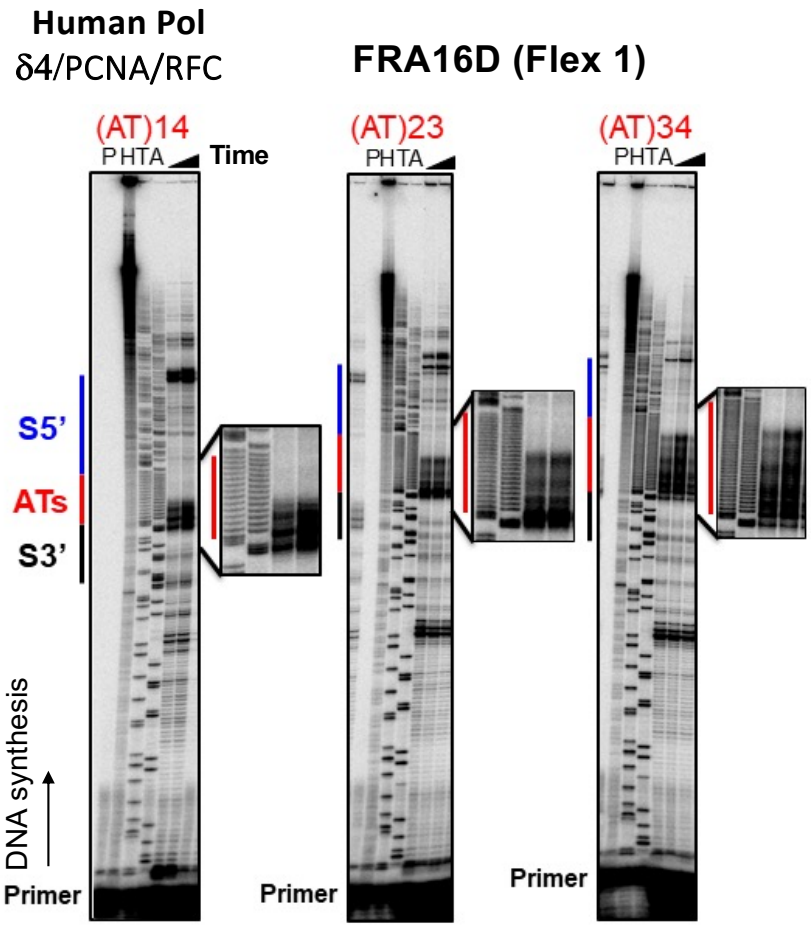
* $p < 0.05$ relative to control sequence

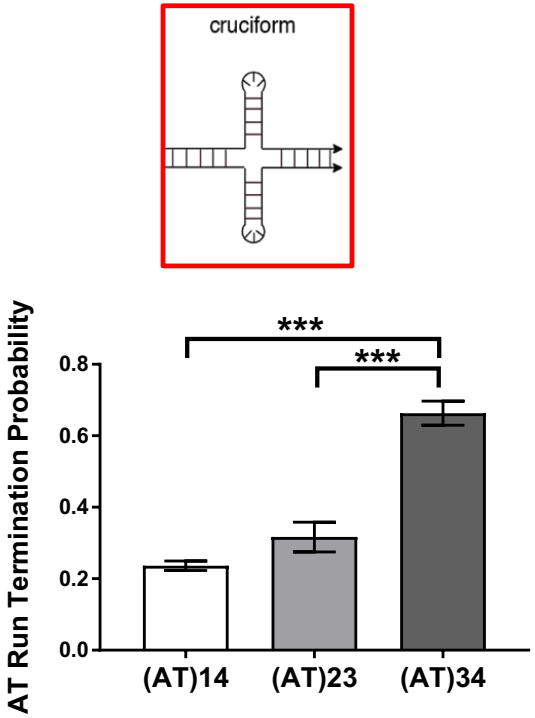
Pol δ dissociates within FRA16D sequence



Walsh et al. (2013) J. Mol. Biol. 425: 232.

Non-B DNA Structure Formation at [AT] repeats slows Pol δ Synthesis

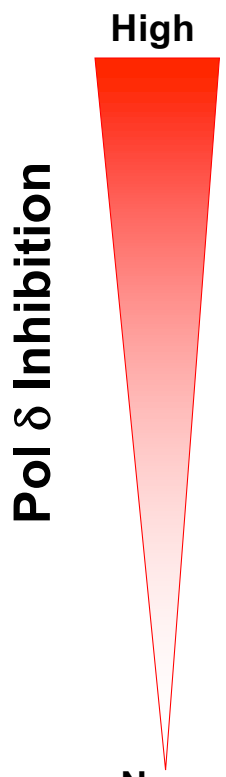




*** p<0.001, one-way ANOVA, Tukey's multiple comparisons

Kaushal, S. et al. (2019) Cell Reports 27:1151-64

Defining DiToRS for Replicative Polymerases (to date)



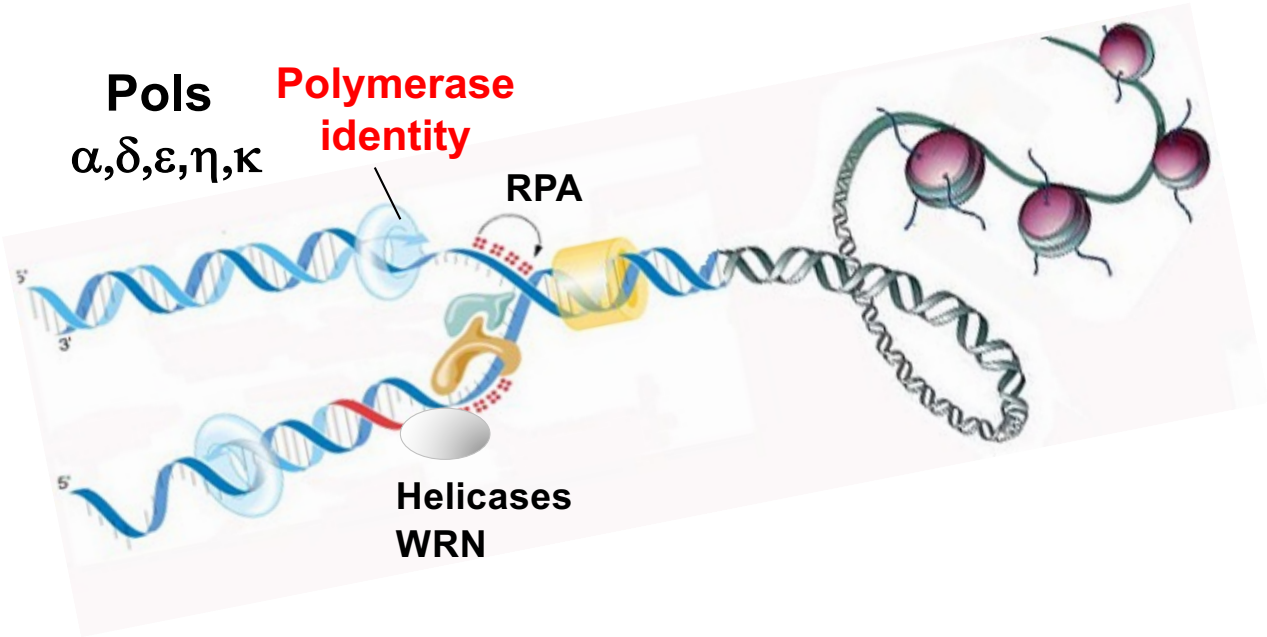
AT-rich Microsatellites		Quasi-Palindromes	G-rich
$[A/T]_{28}$	$[AT/TA]_{34}$	34 nt stem (82% homol.) 29 nt stem (76% homol.)	$[CAGAGG]_{15}$
	$[AT/TA]_{25}$ $[AT/TA]_{24i}$	20 nt stem (100% homol.)	
$[A/T]_{22}$ $[A/T]_{19}$			
$[A/T]_{9}$	$[AT/TA]_{8}$		

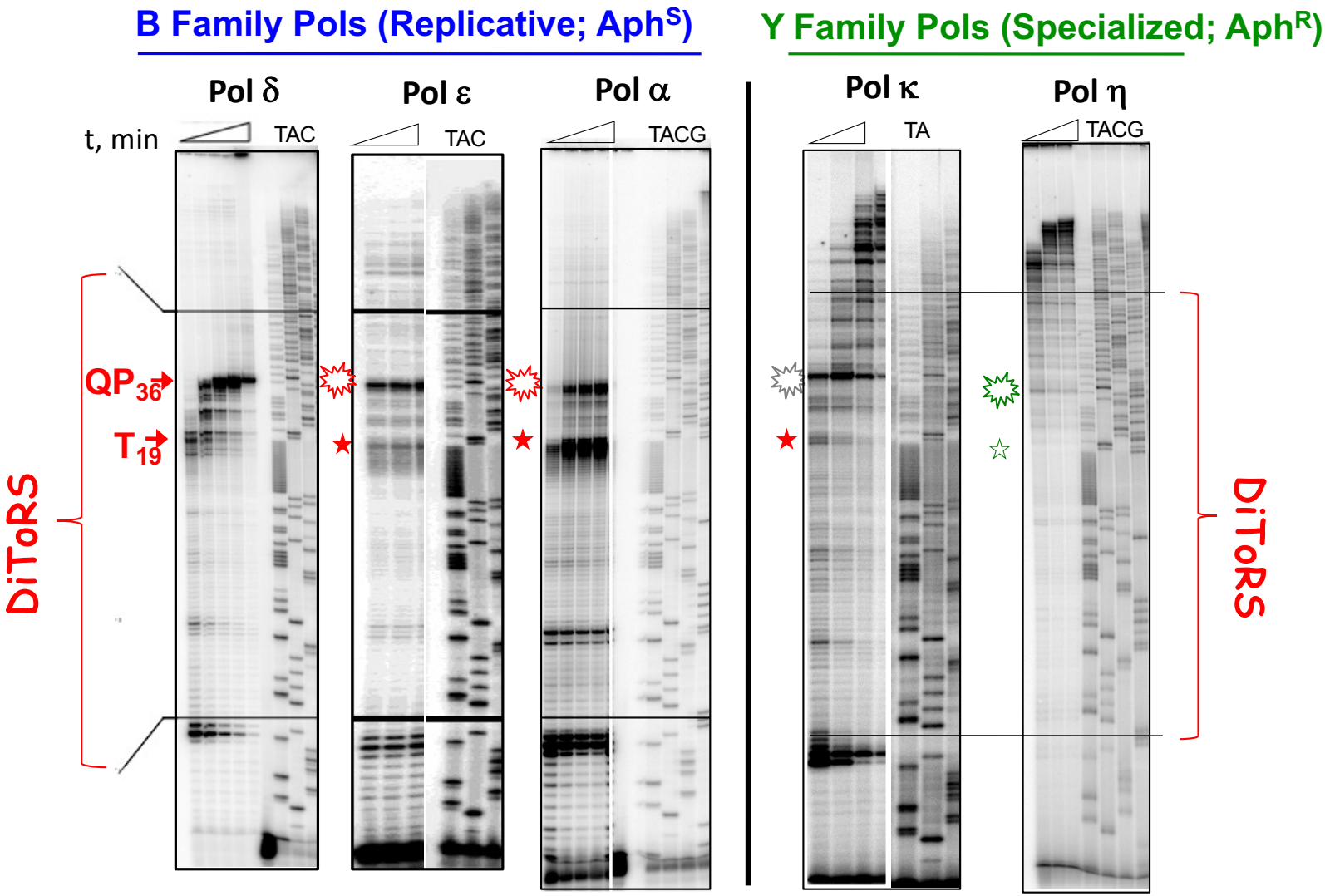
The extent of human Pol δ inhibition within repetitive sequences with non-B DNA potential is length-dependent

Shah et al (2010) ; Walsh et al. (2013) ; Shastri et al (2018); Kaushal, S. et al. (2019)

Getting Through DiToRS: Help for Replicative Polymerases

- Replicative polymerases-inhibition
- Specialized polymerases-efficient synthesis




B Family Pols (Replicative; Aph^S)
Y Family Pols (Specialized; Aph^R)
Pol δ
Pol ε
Pol α
Pol κ
Pol η
t, min
TAC
TAC
TACG
TA
TACG
QP$_{36}$
T$_{19}$
DiToRS
DiToRS

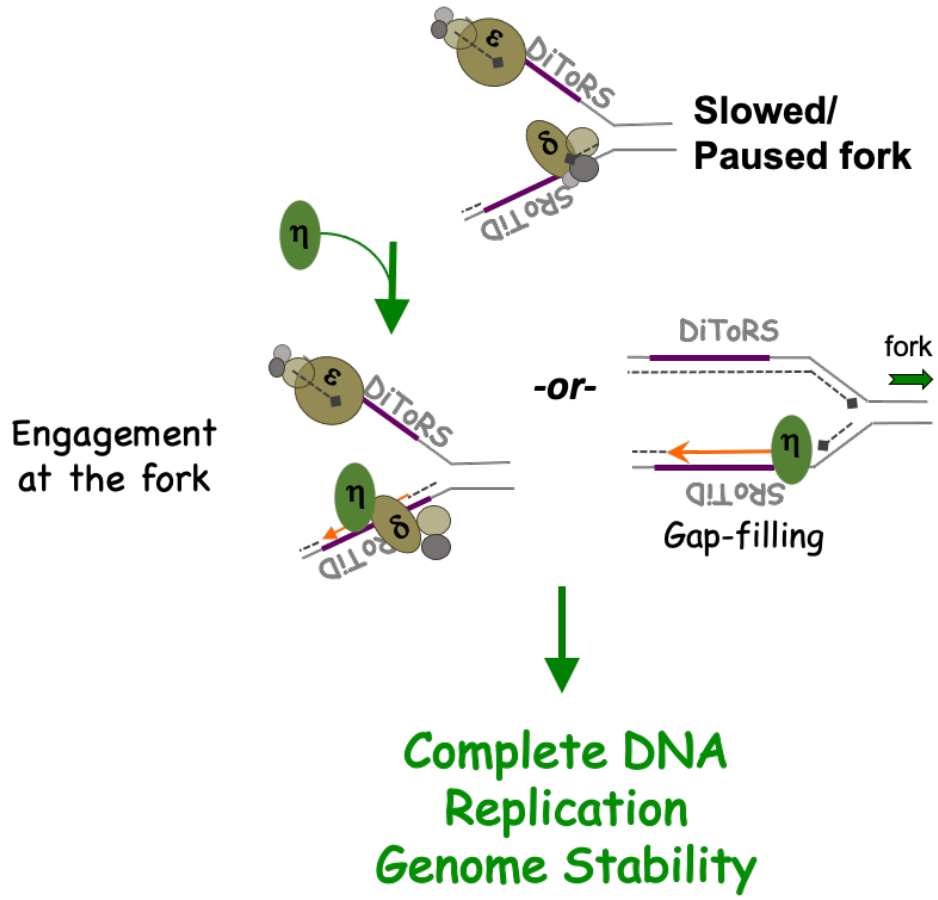
Pols η and κ display efficient synthesis through FRA16D repeat elements

Barnes et al. (2017) DNA Repair.

Walsh et al. (2013) J. Mol. Biol.

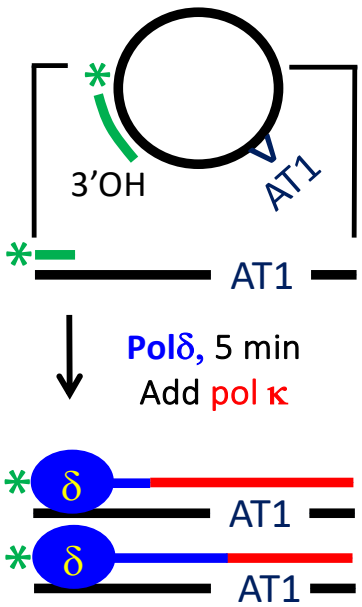
Bergoglio et al. (2013) J. Cell Biol.

Are multiple polymerases needed to complete DiToRS replication?



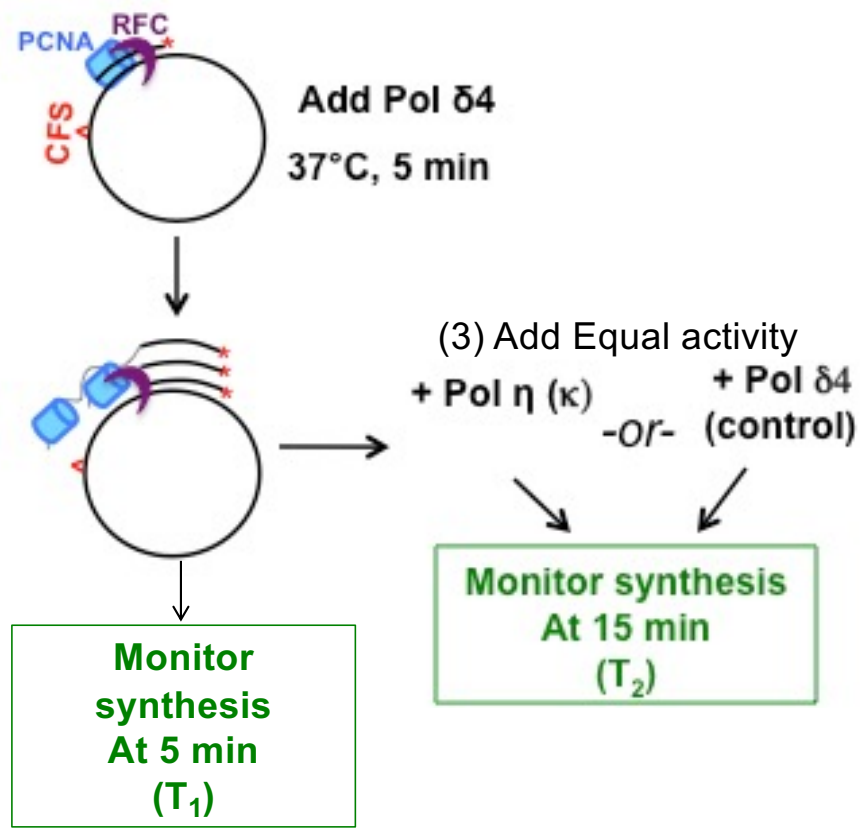
Complete DNA Replication
Genome Stability

Biochemical assay to measure polymerase exchange at DiToRS



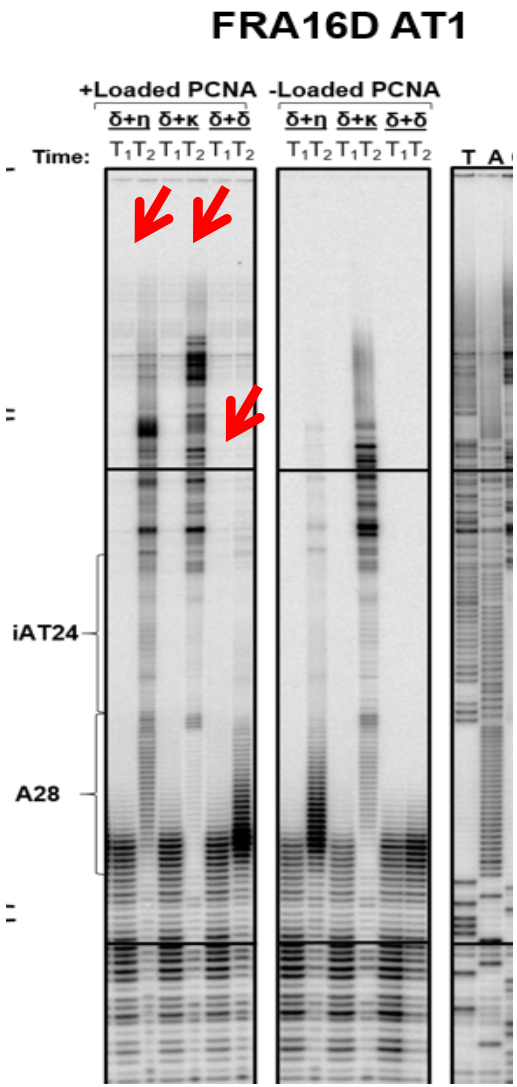
(1) RFC loading of PCNA onto primed ssDNA templates

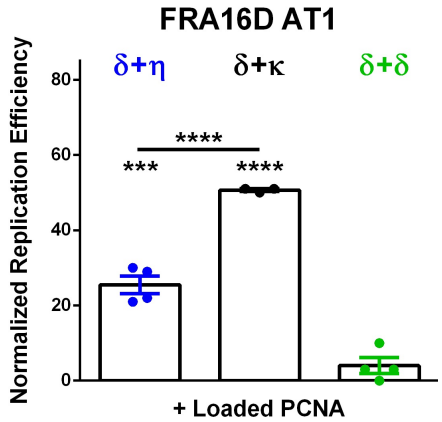
(2) Pol δ HE formation and initial synthesis



Barnes et al (2017) DNA Repair 57: 1-11.

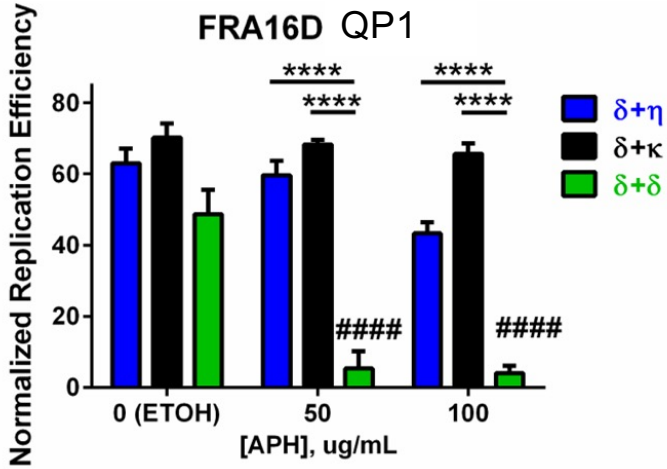
Specialized Pols Efficiently Exchange with Replicative Pol δ to Complete DiToRS Replication





mean ± SEM, ANOVA
Significance relative to δ+δ:
*** =p < 0.001
****= p < 0.0001

Dual polymerase reactions with equal activity



Pols η or κ ensure complete DiToRS synthesis in the presence of aphidicolin

(mean ± SEM) for N > 3 independent experiments.
#, statistical significance relative to EtOH
*, statistical significance relative to δ+δ.

Barnes et al (2017) DNA Repair

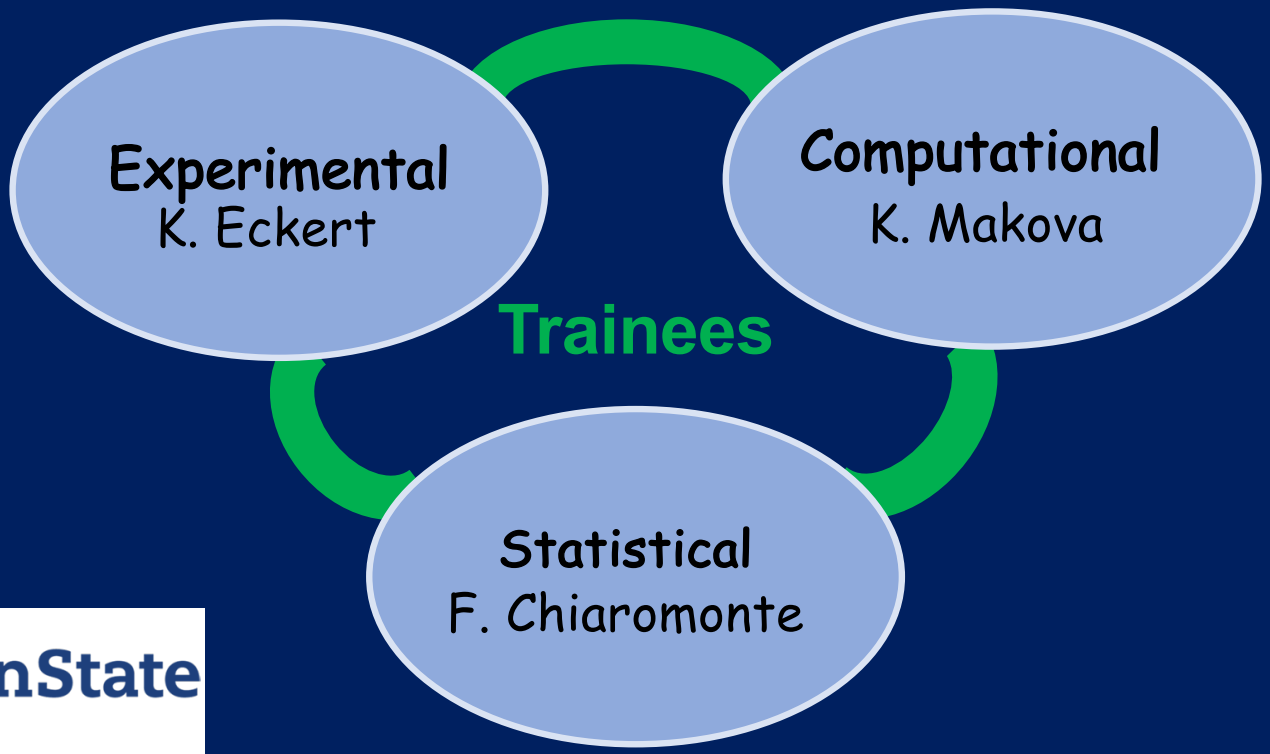
What is the Impact of RPA on Pol δ synthesis through DiToRS?

Kristin Eckert
Penn State-Hershey
College of Medicine

Mark Hedglin
Penn State-University Park
College of Science
Department of Chemistry

What is the complete landscape of DiToRS in the human genome?

Integrative approach to analyzing repetitive sequences

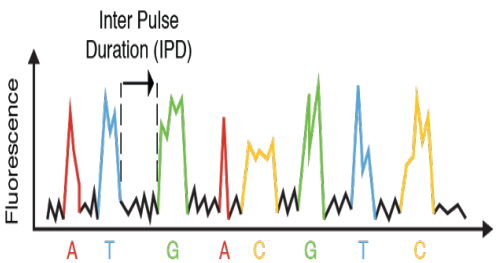


PennState

Identifying DiToRS Genome-Wide

Wilfried Guiblet

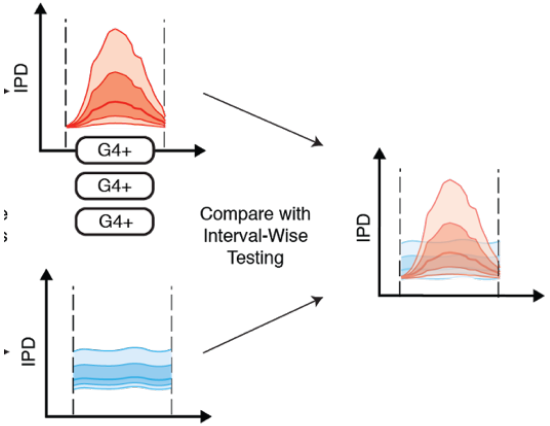
PacBio IPD data used as surrogate for polymerase pausing



- IPD curve shapes differ, depending on motif sequence
- All **G4 motifs** have significantly higher IPDs, compared to motif-free controls
- **Other non-B motifs with altered polymerization kinetics**: Z-DNA, A-phased repeats, mirror repeats
- Sequencing error rates significantly increased for G4 vs motif-free windows:
  - Base substitutions- 1.79 fold
  - Single base deletions-1.49 fold

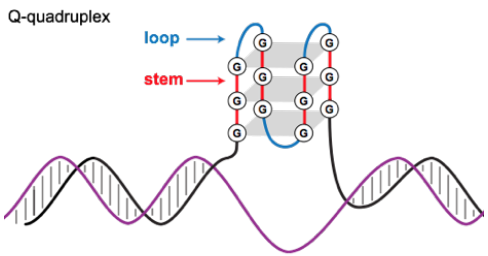
Marzia Cremona

Interval-Wise Testing (IWT) for statistical differences in curve distributions



Wilfried Guiblet, Marzia Cremona et al. (2018) Genome Research 28: 1767-1778

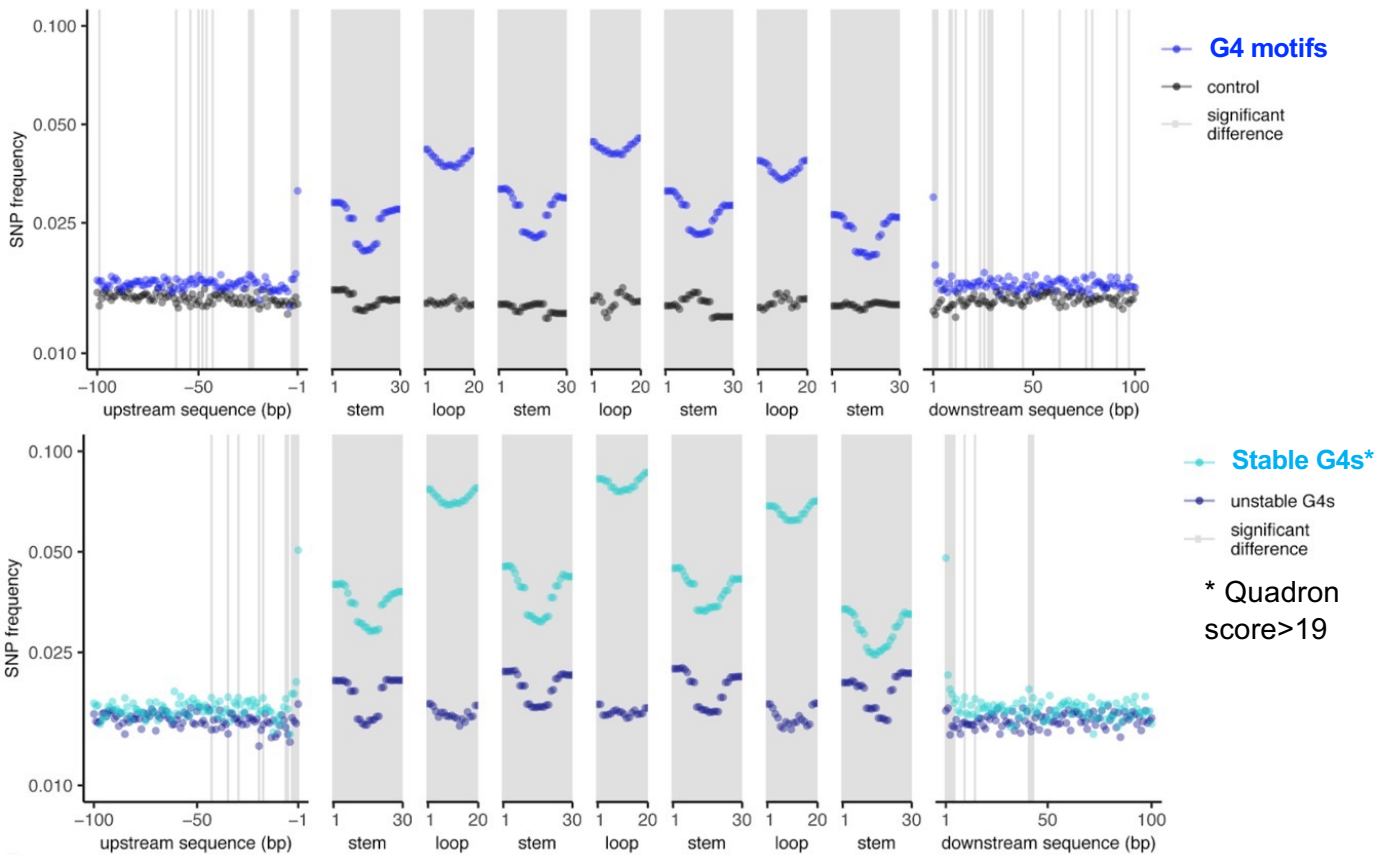
Nucleotide Substitution Frequencies Vary at Stable G4 motifs Genome-wide



4 stems; 3 loops

- G4 loci annotated with Quadron (hg19), using the non-coding, non-repetitive subgenome
- SNP frequency determined from the Simons Genome Diversity Project (~30X depth, 279 individuals)

Guiblet, Cremona et al (2021) Nucleic Acids Research 49:1497-1516

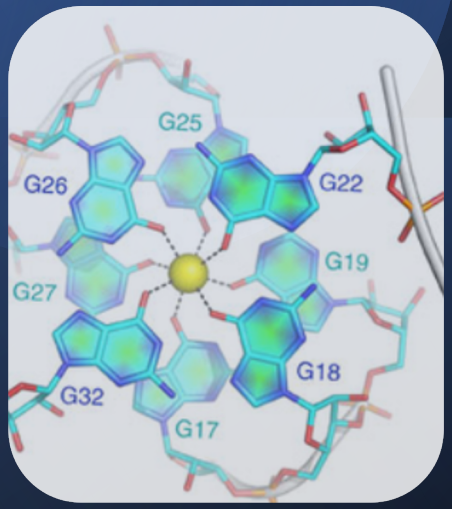


* Quadron score>19

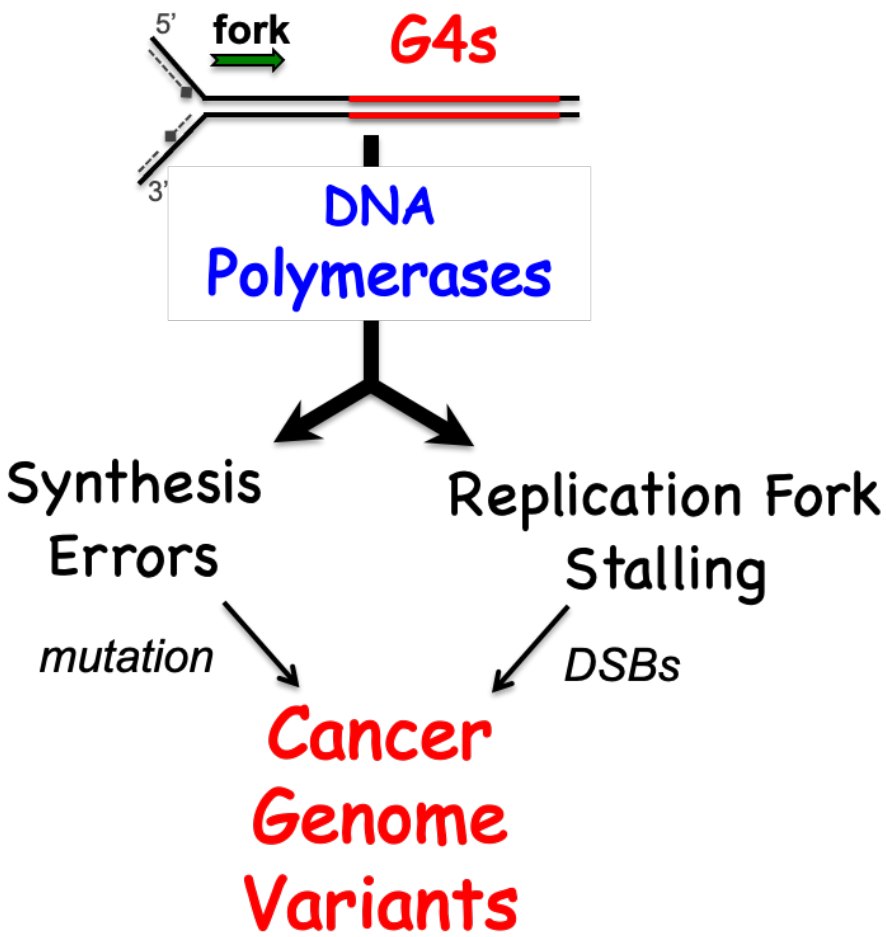
*Sahakyan et al. (2017) Machine learning model for sequence-driven DNA G-quadruplex formation. Sci. Rep. , 7 , 14535.

Do G4s affect human DNA polymerase fidelity?

Do G4 sequences inhibit replicative DNA polymerase holoenzymes ?



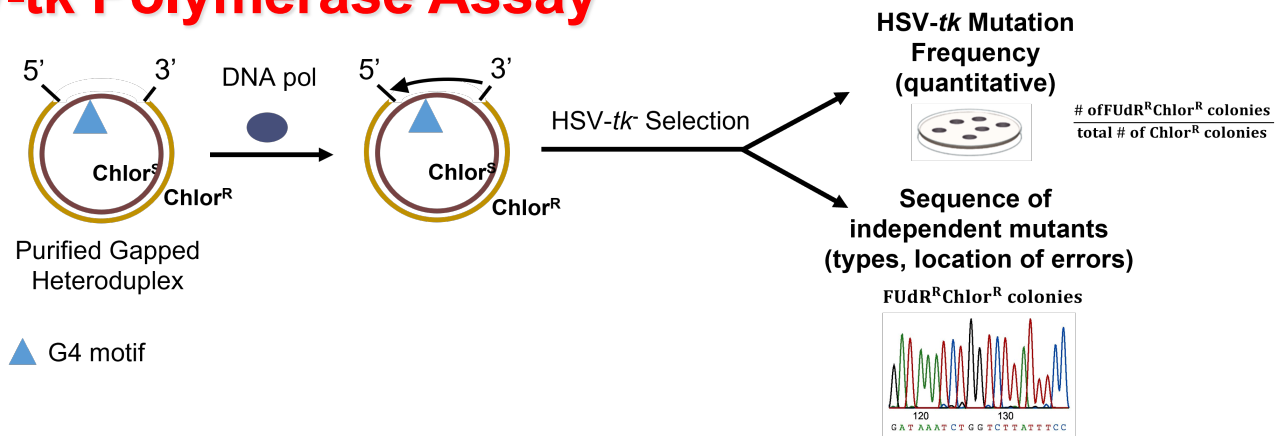
Current research:



Do G4s Increase Human DNA Polymerase Errors?

In vitro HSV-tk Polymerase Assay

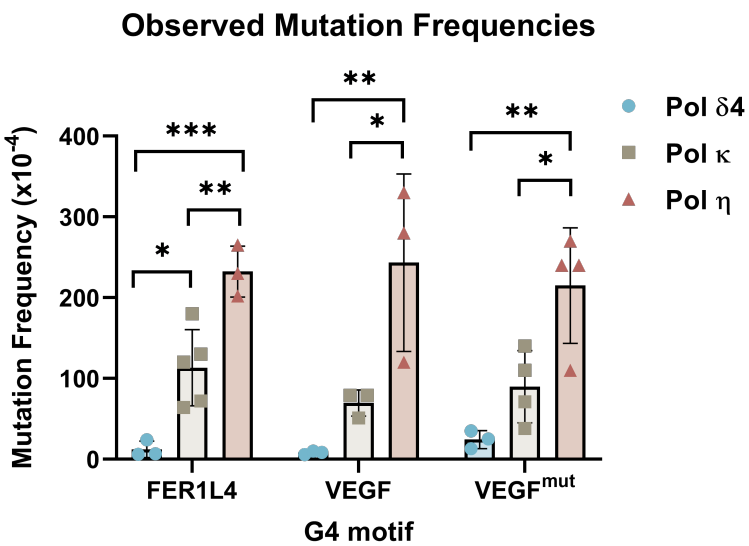
MaryElizabeth Stein



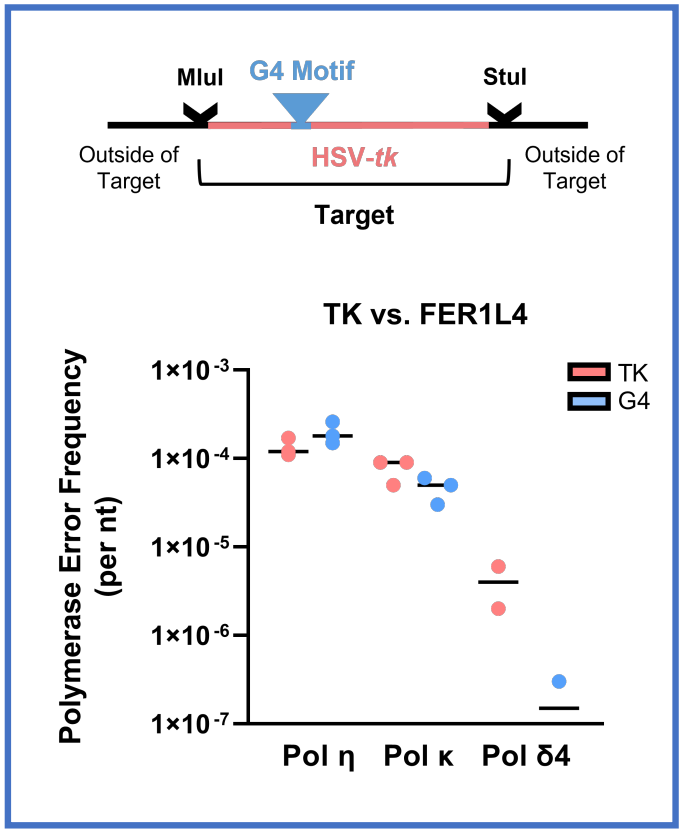
G4 Motif	ID	Length (nts)	Location	G4 Type*	T_m (°C)
5' GGGcgaaGGGGcgagccaGGGGtaaGGGG 3'	FER1L4	29	intron	antiparallel	75.6
5'GGGGcGGGccGGGGGcGGGGtcccggcGGGG3'	VEGF	31	promoter	parallel	83.5
5' GGGGcGGGccGGGGGcGGGG 3'	VEGF^{mut}	20	n.a.	parallel	85.5

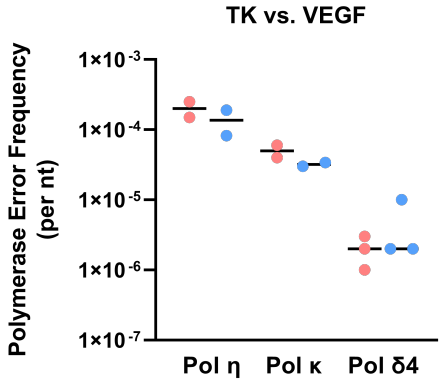
* Determined in polymerase reaction buffer using circular dichroism

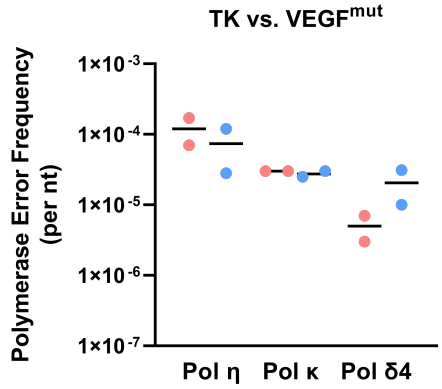
Polymerase fidelity hierarchy is maintained during G4 synthesis



Mean ± SD, One-Way ANOVA, Tukey test for multiple comparisons *p<0.05, **p<0.01, ***p<0.001



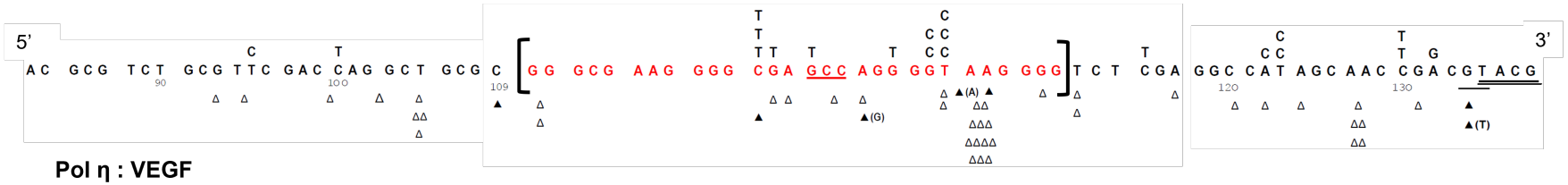




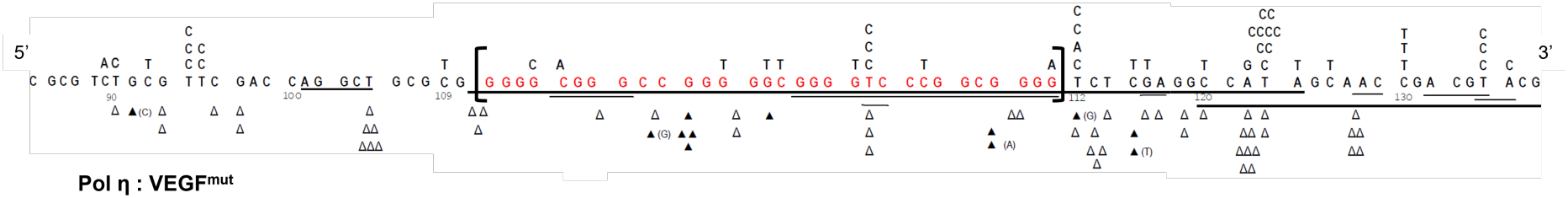
M. Stein, in preparation

G4 motifs influence errors made in flanking sequences

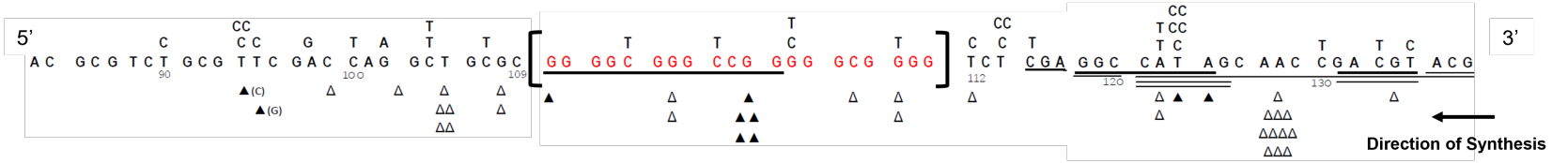
Pol η : FER1L4



Pol η : VEGF



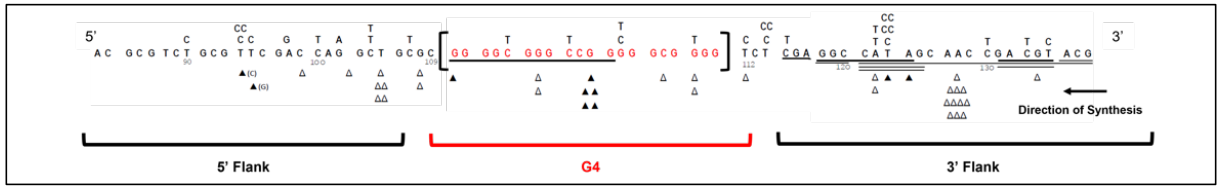
Pol η : VEGF^{mut}

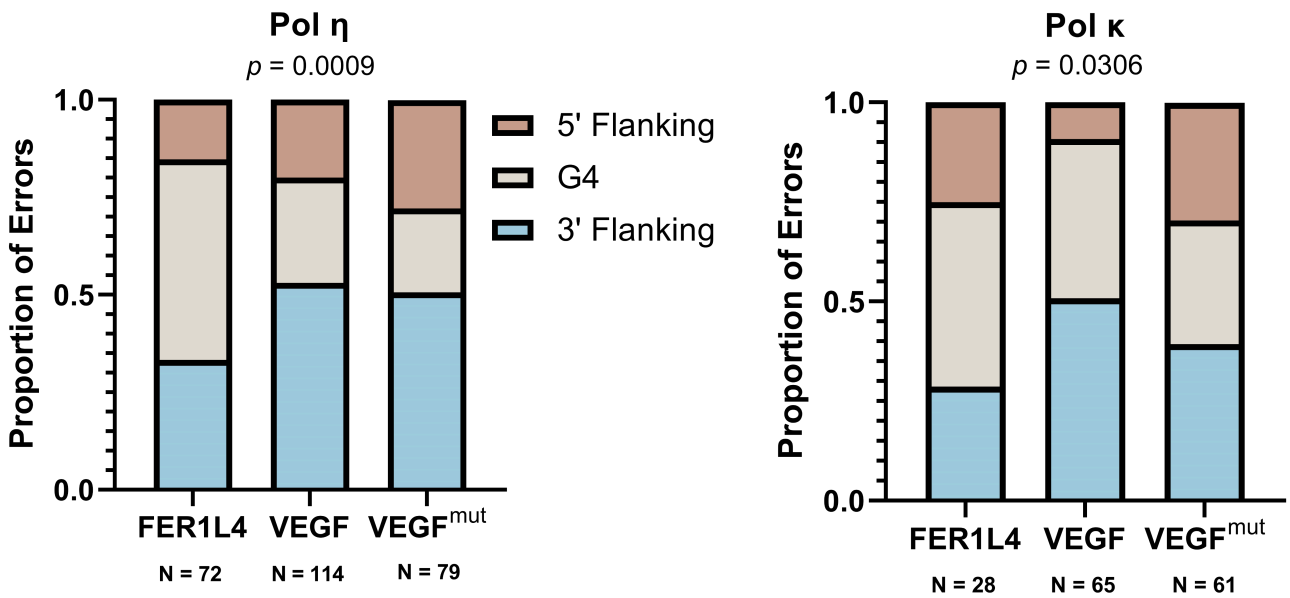


5' Flank **G4** **3' Flank**

Flanking window size based on average G4 length.
Δ = Deletion event, ▲ = Insertion event, Underline = deletions >1nt, substituted bases above sequence.

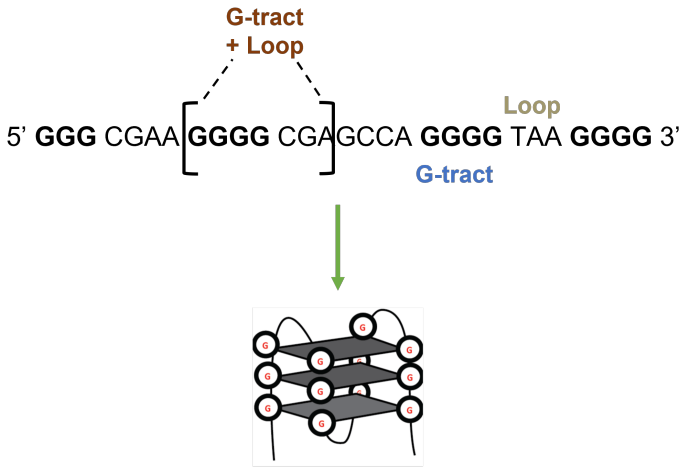
G4 motifs influence errors made in flanking sequences

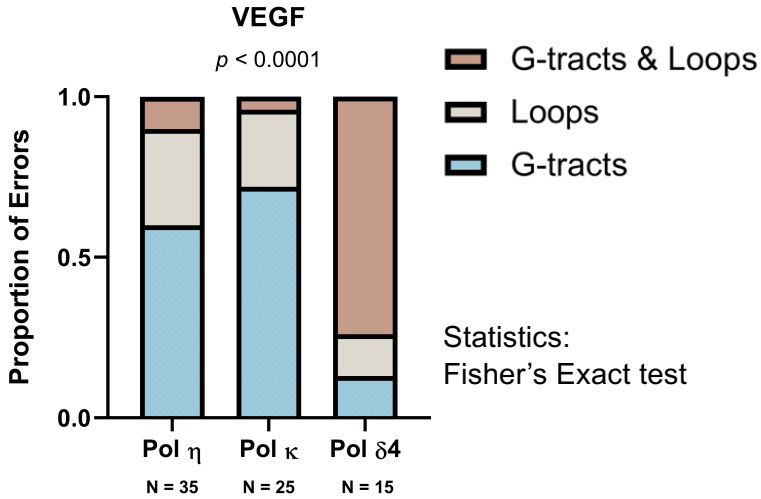




Chi-square analysis

Breadth of DNA pol errors involving G4 motifs

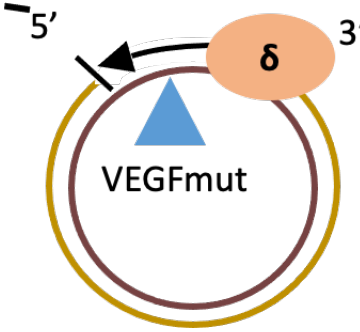




Insertion mutations

tgcgc **G** [ggggcgggccggg **GGCGGGGTCCC** ggcgggg] tctcga

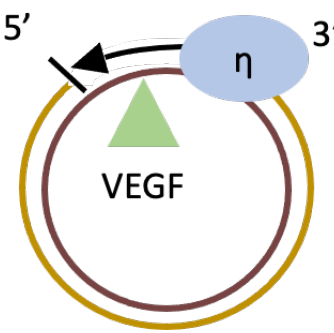
- Unique to stable G4s
- Expand G4 motif

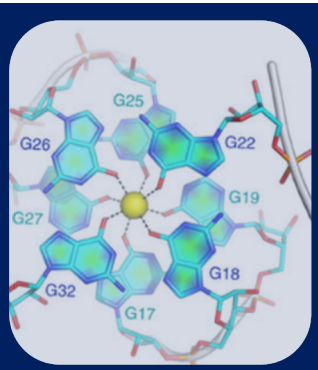


Deletion mutations

ggctgc [**31 base del.**] gccata

- Most frequent in stable G4s

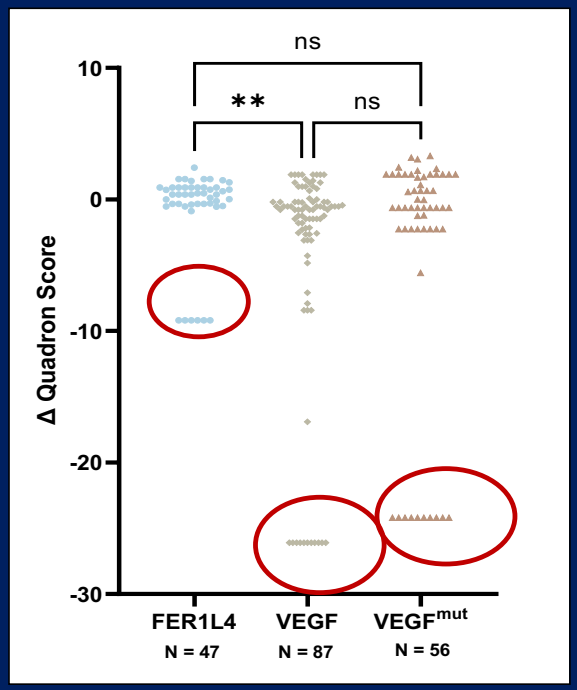




Integrative approach to analyzing DiToRS

What are the potential biological consequences of polymerase errors within G4s?

Matthias Weissensteiner



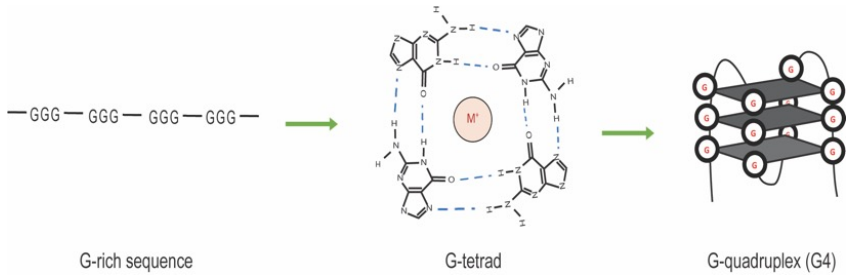
Quadron Score = likelihood of G4 formation:

- Potential quadruplex sequence (PQS) identification (Quadparser and G4seq)
- Features of G4 sequence and flanking sequence (e.g., loop lengths, which G-tract number)

Δ Quadron Score = mutant score – WT score
Kruskal-Wallis test, ** $p < 0.01$; N = total number of mutants

Some errors completely abolished G4 potential

Do G4s Inhibit Human Pol δ?



Suzanne Hile

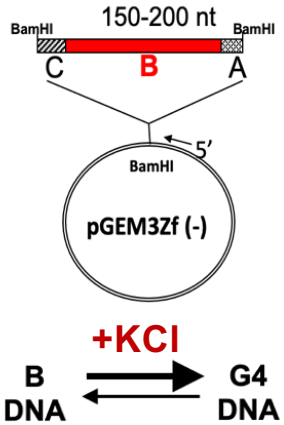
Motifs Examined:

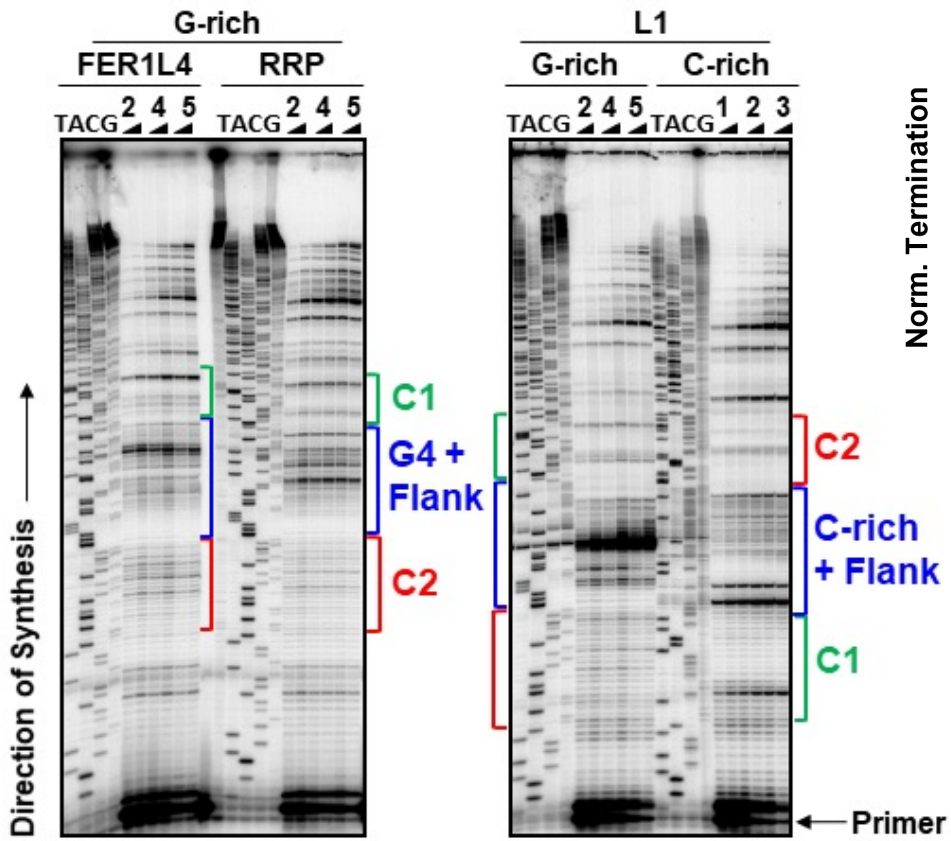
Name	G4 Motif (5'-3')	Quadron Score*	Fold (CD)[a]	T_M (°C)
FER1L4	GGGCGAAGGGGCGAGCCAGGGGTAAGGGG	17.8	Antiparallel	75.6
RRP	GGGCAGGGGCTCCCTGGGCTGGG	11.7	Antiparallel	67.3
L1	GGGGTGGGGGGAGGGGGGAGGG	**22.0**	**Parallel**	**>90**

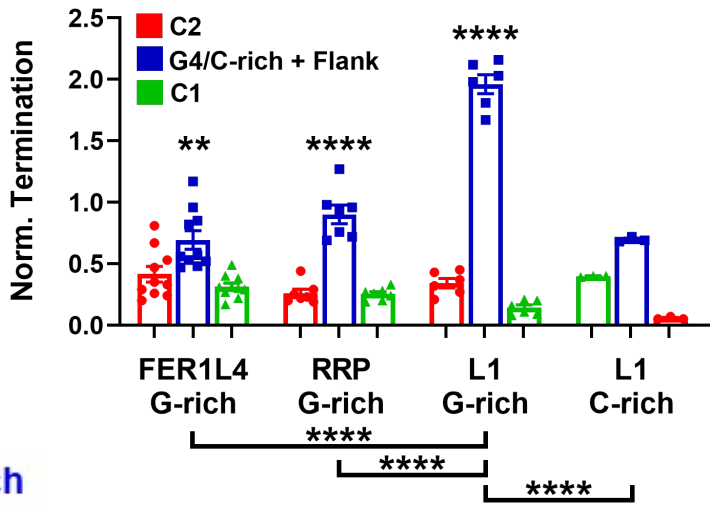
[a] Circular Dichroism and thermal melting performed in polymerase reaction buffer (+150 mM KCl)
All motifs form intrastrand G4s

- **Motifs selected from genome-wide study**
- **Vary in sequence composition and stability**

Pausing by Human Polδ4/PCNA/RFC Increases with G4 Stability





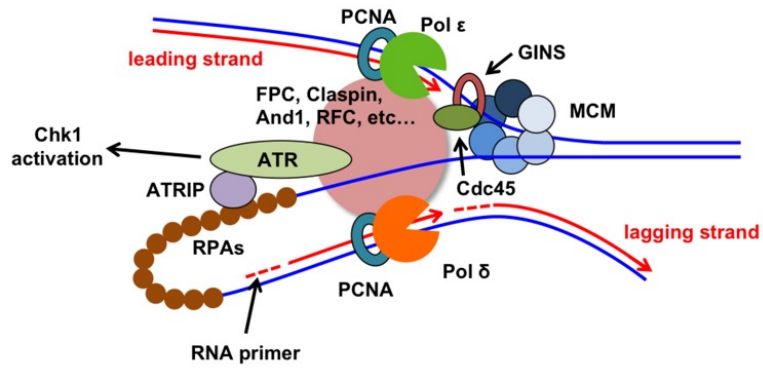


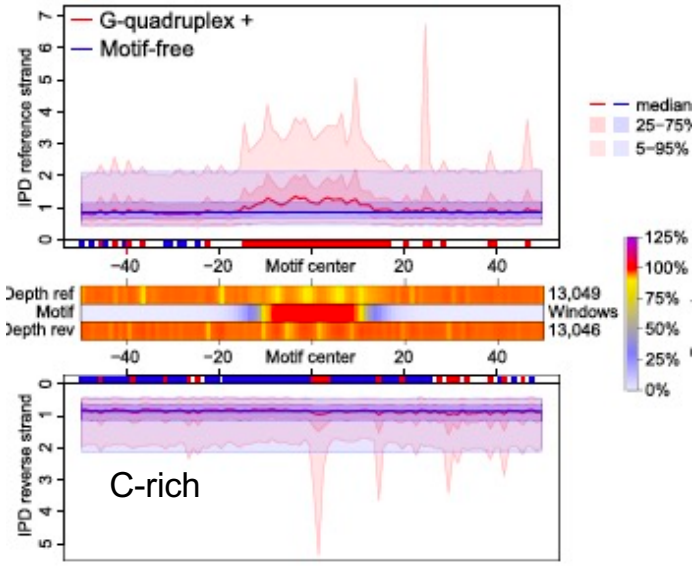
Data from 4-10 reactions
** p< 0.01, ****, p< 0.0001 (Anova)

S.Hile, unpublished

Does DNA synthesis remain coupled on complementary G-rich and C-rich strands?

G4 motifs: Strand biased IPD
G strand > C strand





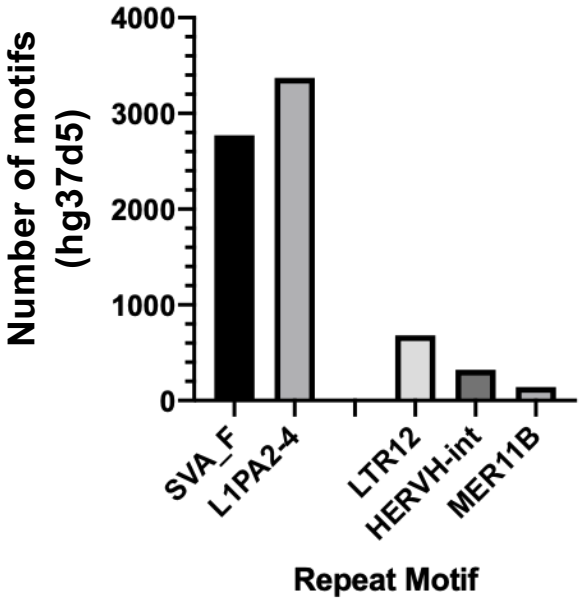
Guiblet and Cremona 2018

G4 Motif Choice: Abundance and Function

Name	G4 Motif (5'-3')	Quadron Score*	Fold (CD)[a]
FER1L4	GGGCGAAGGGGCGAGCCAGGGGTAAGGGG **1 motif in human genome**	17.8	Antiparallel
RRP	GGGCAGGGGCTCCCTGGGCTGGG **3 motifs in human genome**	11.7	Antiparallel
L1	GGGGTGGGGGGAGGGGGGAGGG **2,937 occurrences within L1 PA2-4 elements**	**22.0**	Parallel
SVA	GGGAGGGAGGTGGGGGGG **2,681 occurrences within SVA_F elements**	**32.6**	Parallel
OGRE[a]	GGGGGATGGGGTTGGAATGGGGGCGGG	n.d.	n.d.
Random	GAGCTGAGTGGAGGCGTGAGCG	n.a.	n.a.

[a] Origin G-rich element (mouse genome), ori2 sequence
Prorok et al. (2019) Nature Communications. PMID: 31332171

Mobile elements are a highly abundant source of stable G4s in the human genome

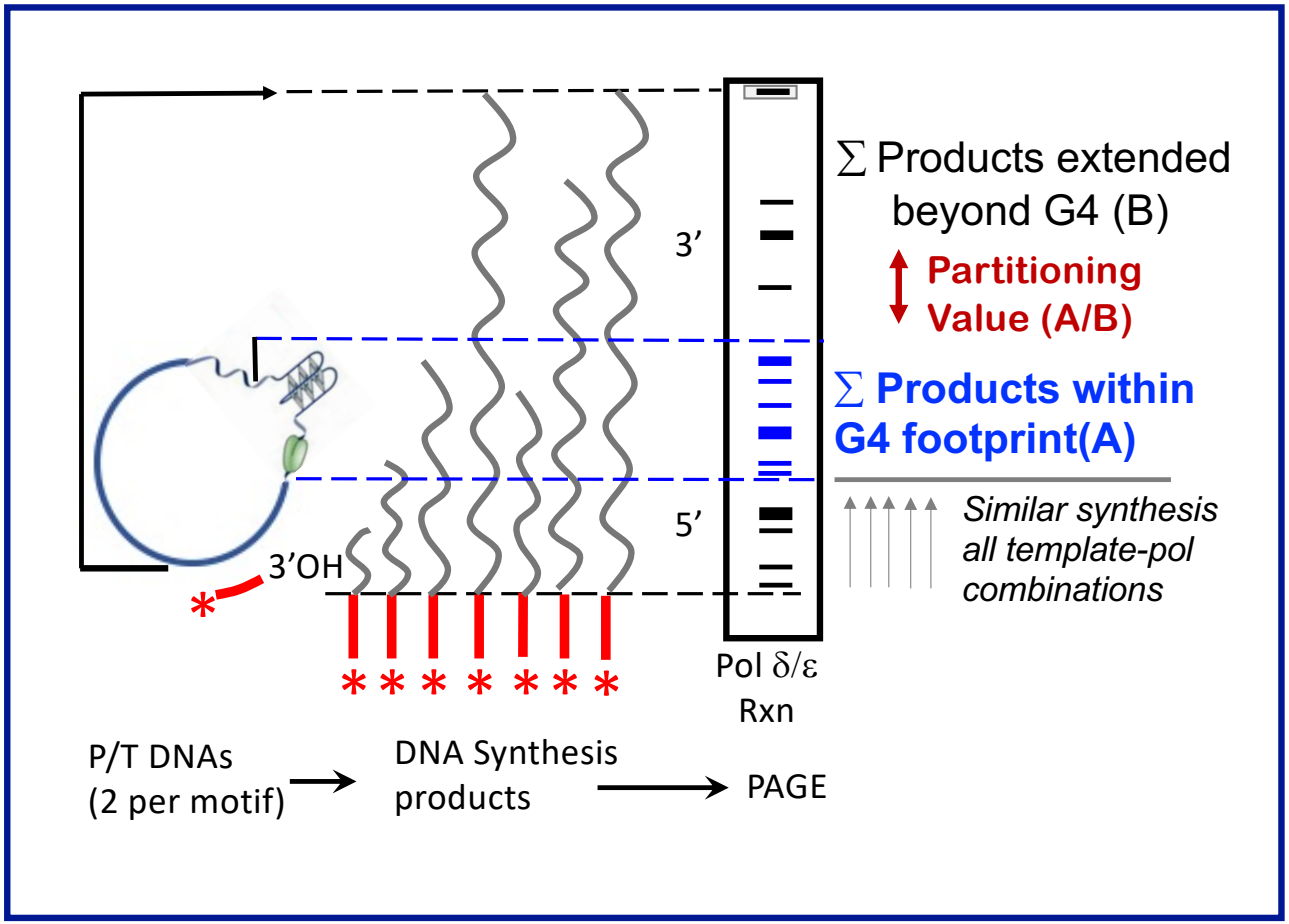


M. Weissensteiner

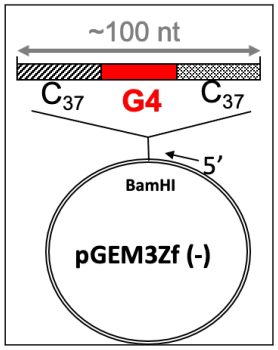
Joe Dahl
(NIEHS)

yeast Pol δ and Pol ε holoenzymes

Approach: Compare Pol δ and Pol ε synthesis on complementary strands of G4 motifs



Suzanne Hile



Perspective: How are G4 motifs maintained in the human genome, so as to retain structure/function?

MUTATION

Human polymerase error frequencies within G4s are not elevated, relative to flanking control sequences

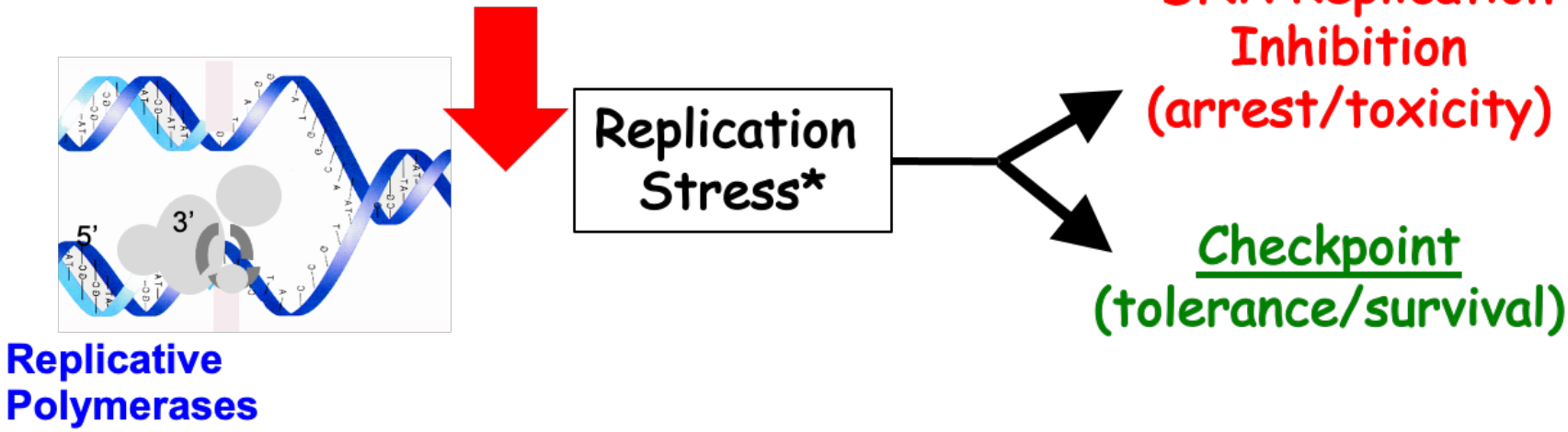
- Presence of G4 increases polymerase errors within 3' flanking regions
- Polymerase errors within the G4 motifs can increase or decrease predicted G4 stability
- Increased germline SNP frequency could be due to G4 effects on other mutational pathways
  - Inhibition of mismatch repair
  - Increased DNA damage

REPLICATION

Some, but not all, G4 motifs impede replicative holoenzyme synthesis in a strand-dependent manner.

- Increased pausing on G-rich strand observed for both Pol ε and Pol δ .
  - G4 motifs in mobile elements and an OGRE
- Replication uncoupling at G4s is greatest when in the leading strand conformation
- Assistance of other replication proteins will be crucial to replicate G4s in the genome
  - Specialized polymerases (Pols η, κ)
  - RPA
  - Specialized helicases (FANCJ, WRN)

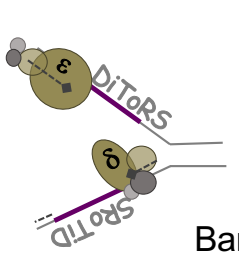
When do DiToRS become threats to genome stability?



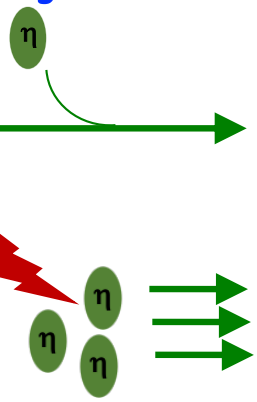
Replication Stress and DiToRS threats

- Ribonucleotide reductase inhibition (low [dNTPs]
- RPA exhaustion
- Specialized polymerase deficiency (e.g., Pol η)
- Specialized helicase deficiency(e.g., WRN)
- Oncogene activation

Specialized Polymerases Ensure Complete DiToRS Replication



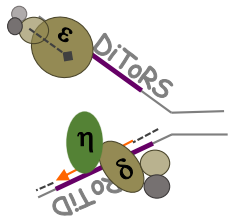
Slowed/ Paused fork

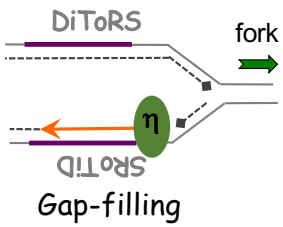


Barnes et al (2018) Cancer Research

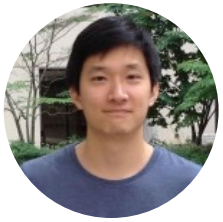
Aph
HU

Engagement at the fork *-or-* **Fork Repriming**

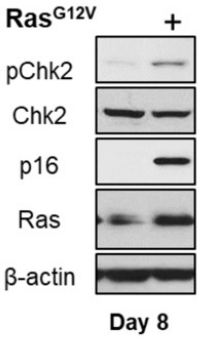


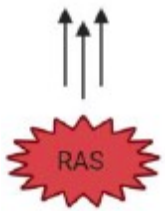


Complete DNA Replication Genome Stability



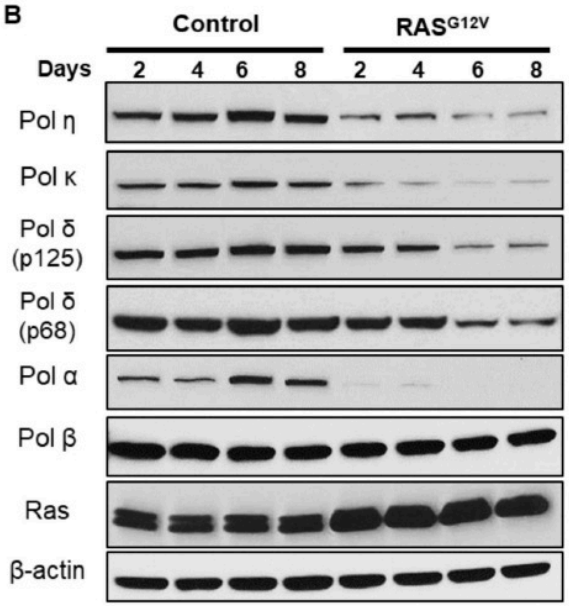
Bryan Tsao



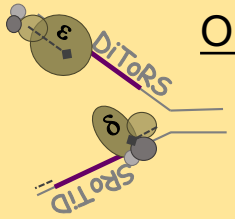


hTert-BJ5a cells

Tsao et al (2021) Plos One

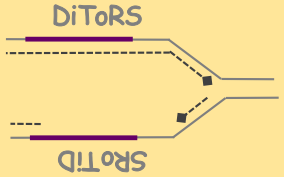


Oncogene Overexpression



Increased fork stalling
Polymerase degradation

Incomplete genome duplication



Increased vulnerability to genome instability?

The Penn State Team

Eckert Laboratory

- Suzanne Hile
- MaryElizabeth Stein
- Bryan Tsao, Ph.D.
- former grad students

Makova/Chiaromonte Group

- Wilfried Guiblet
- Marzia Cremona
- Matthias Weissensteiner

Hedglin Laboratory (RPA studies)

- Mark Hedglin
- Kara Pytko

Collaborators

Purified proteins

- Marietta Lee/Sufang Zhang (hPolδ)
- Linda Bloom (RFC)
- Joe Dahl (yPolδ, Polε)

G4 biophysical analyses

- Iva Kejnovska
- Eduard Kejnovsky

Funding

NIH NATIONAL CANCER INSTITUTE

NIH National Institute of General Medical Sciences

JAKE GITTLEN Memorial Golf Tournament
Benefitting Gittlen Cancer Research