Human Molecular Genetics, 2020, Vol. 29, No. 8 1292–1309

doi: 10.1093/hmg/ddaa043
Advance Access Publication Date: 19 March 2020
General Article

GENERAL ARTICLE

Mitochondrial genetic variation is enriched in G-quadruplex regions that stall DNA synthesis *in vitro*

Thomas J. Butler[1,†], Katrina N. Estep[2,†], Joshua A. Sommers[2,†], Robert W. Maul[3], Ann Zenobia Moore[1], Stefania Bandinelli[4], Francesco Cucca[5], Marcus A. Tuke[6], Andrew R. Wood[6], Sanjay Kumar Bharti[2], Daniel F. Bogenhagen[7], Elena Yakubovskaya[7], Miguel Garcia-Diaz[7], Thomas A. Guilliam[8], Alicia K. Byrd[9], Kevin D. Raney[9], Aidan J. Doherty[8], Luigi Ferrucci[1], David Schlessinger[10], Jun Ding[1,*] and Robert M. Brosh Jr[2,*]

[1]Translational Gerontology Branch, National Institute on Aging, Baltimore, MD 21224, USA, [2]Laboratory of Molecular Gerontology, National Institute on Aging, Baltimore, MD 21224, USA, [3]Laboratory of Molecular Biology and Immunology, National Institute on Aging, Baltimore, MD 21224, USA, [4]Geriatric Unit, Azienda Sanitaria di Firenze, Florence 50121-50145, Italy, [5]Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche, Monserrato 09042, Italy, [6]Genetics of Complex Traits, University of Exeter Medical School, Exeter EX1 2LU, UK, [7]Department of Pharmacological Sciences, Stony Brook University, Stony Brook, NY 11794-8651, USA, [8]Genome Damage and Stability Centre, School of Life Sciences, University of Sussex, Brighton BN1 9RQ, UK, [9]Department of Biochemistry and Molecular Biology, University of Arkansas for Medical Sciences, Little Rock, AR 72205, USA and [10]Laboratory of Genetics and Genomics, National Institute on Aging, Baltimore, MD 21224, USA

*To whom correspondence should be addressed. Tel: +1 4105588578; Email: broshr@mail.nih.gov; Tel: +1 4105588297; Email: jun.ding@nih.gov

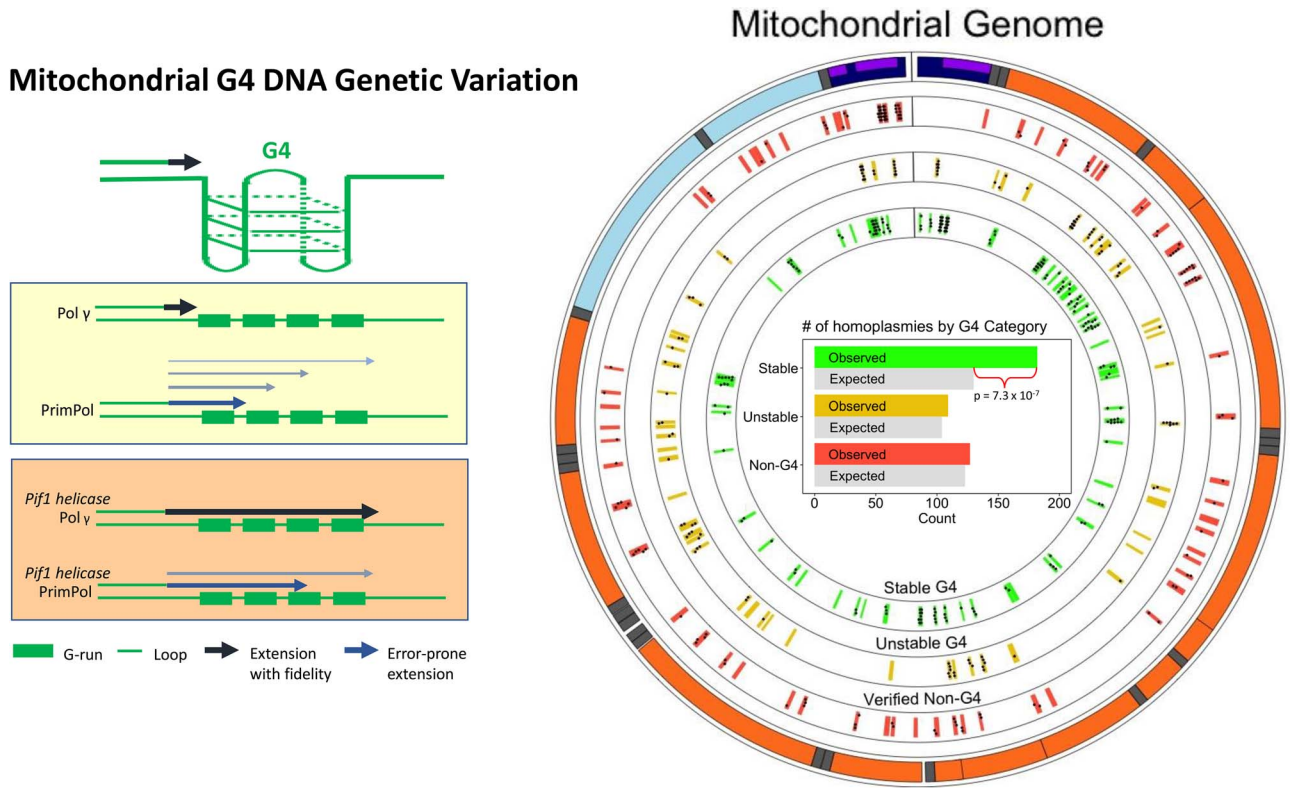
Abstract

As the powerhouses of the eukaryotic cell, mitochondria must maintain their genomes which encode proteins essential for energy production. Mitochondria are characterized by guanine-rich DNA sequences that spontaneously form unusual three-dimensional structures known as G-quadruplexes (G4). G4 structures can be problematic for the essential processes of DNA replication and transcription because they deter normal progression of the enzymatic-driven processes. In this study, we addressed the hypothesis that mitochondrial G4 is a source of mutagenesis leading to base-pair substitutions. Our computational analysis of 2757 individual genomes from two Italian population cohorts (SardiNIA and InCHIANTI) revealed a statistically significant enrichment of mitochondrial mutations within sequences corresponding to stable G4 DNA structures. Guided by the computational analysis results, we designed biochemical reconstitution experiments and demonstrated that DNA synthesis by two known mitochondrial DNA polymerases (Pol γ, PrimPol) *in vitro* was strongly blocked by representative stable G4 mitochondrial DNA structures, which could be overcome in a specific manner by the ATP-dependent G4-resolving helicase Pif1. However, error-prone DNA synthesis by PrimPol using the G4 template sequence persisted even in the presence of Pif1. Altogether, our results suggest that genetic variation is enriched in G-quadruplex regions that impede mitochondrial DNA replication.

[†]These authors contributed equally to this work.

Received: November 13, 2019. **Revised:** January 27, 2020. **Accepted:** March 18, 2020

Graphical Abstract



Introduction

As the site of oxidative phosphorylation, mitochondria (mt) are indispensable for proper cell function. Therefore, it is no surprise that the 16 kb circular mt genome must be faithfully replicated to maintain genomic stability and prevent deleterious mutations that can lead to mt diseases. When mutations do occur, they may become fixed in every copy of the mt genome in a cell, a property referred to as homoplasmy, or may be present only in a certain percentage of copies, termed heteroplasmy. Understanding how mt variants arise and assessing their biological consequence are a crucial step in identifying risk factors for mt disease.

A potential source of mt point mutations is polymerase stalling during DNA replication. In humans, mt replication is carried out by several nuclear-encoded proteins that comprise the mt replisome; key factors include the DNA polymerase gamma (Pol γ), the DNA helicase Twinkle and the mt single-strand binding protein (mtSSB) (1). In addition to Pol γ, several other DNA polymerases have been implicated in mitochondrial DNA metabolism, including PrimPol (2,3), Pol zeta (4), Pol theta (5) and Pol beta (6). Other factors that bind mtDNA, including the mt transcription factor A (TFAM), may also play a role in replication (7). These proteins are highly specialized to carry out accurate replication but are sensitive to secondary structures in the DNA template. One such structure that is likely to pose a significant roadblock for the mt replication machinery is the G-quadruplex (G4), a non-canonical secondary structure stabilized by Hoogsteen bonds that can occur in guanine (G)-rich regions of DNA (8). G-quadruplex structures have been shown to stall the nuclear replication machinery (9–11), leading to fork arrest and potential collapse (12,13). In terms of mtDNA metabolism, evidence suggests that hybrid RNA/DNA G-quadruplex structures stimulate transcription termination to facilitate replication initiation (14,15). However, the relevance of G4s to mutagenesis in the mt genome is poorly understood. Previous research has shown that DNA double-stranded breaks tend to occur proximal to predicted G4-forming sequences (16,17), but to our knowledge there is currently no study assessing possible accumulation of base substitutions associated with replisome pausing at G4 structures.

Despite the wealth of information regarding the biochemical mechanisms of nuclear tolerance of G4s (8), little is known about how G4s affect mt replication or how the mt replisome copes with these difficult-to-replicate sequences. Recent studies showed that translesion polymerase Rev1 disrupts and replicates G4 (18,19) and the primase-polymerase PrimPol is required for replicative tolerance of G4s in vertebrate nuclei (20). PrimPol has been shown to extend primers as well as bypass photolesions and common oxidative lesions (8-oxoguanine, abasic sites) (3,21,22). Interestingly, PrimPol shows some preference for incorporating the correct nucleotide (dCMP) opposite 8-oxoguanine to an extent that is dependent on the divalent cation, Mg^{2+} versus Mn^{2+} (22). Additionally, PrimPol can function as a primase, catalyzing synthesis of *de novo* primers downstream of replication block sites, and is localized to mitochondria (3,23), making it a potential candidate for G4 replication.

We have combined bioinformatic analyses of mt genome sequence data from two population cohorts with biochemical approaches to study the association between mt variants and G4 structures. The two cohorts, SardiNIA (https://sardinia.nia.nih.gov/) and InCHIANTI (http://inchiantistudy.net/wp/), provide

Table 1. Statistical enrichment of homoplasmies and heteroplasmies in stable G4, but not unstable G4 or verified non-G4 categories

	Stable G4	Unstable G4	Verified non-G4	Stable G4	Unstable G4	Verified non-G4
Homoplasmies	SardiNIA (no. of total variants: 829)			InCHIANTI (no. of total variants: 1313)		
No. of homoplasmies in G4 category (%)[a]	182 (21.9%)	109 (13.1%)	127 (15.3%)	267 (20.3%)	187 (14.2%)	198 (15.1%)
Proportion of bases in G4 category	15.6%	12.5%	14.8%	15.6%	12.5%	14.8%
Expected no. of homoplasmies under null	130	104	123	205	164	194
P-value for homoplasmy enrichment	7.3×10^{-7}	0.28	0.32	2.7×10^{-6}	0.030	0.38
Bonferroni *P*-value[b]	2.2×10^{-6}	0.84	0.96	8.1×10^{-6}	0.090	1.0
Heteroplasmies	SardiNIA (no. of total variants: 697)			InCHIANTI (no. of total variants: 433)		
No. of heteroplasmies in G4 category (%)[a]	138 (19.8%)	85 (12.2%)	113 (16.2%)	82 (18.9%)	57 (13.2%)	64 (14.8%)
Proportion of bases in G4 category	15.6%	12.5%	14.8%	15.6%	12.5%	14.8%
Expected no. of heteroplasmies under null	109	87	103	68	54	64
P-value for heteroplasmy enrichment	1.5×10^{-3}	0.58	0.14	0.028	0.32	0.48
Bonferroni *P*-value[b]	4.4×10^{-3}	1.0	0.42	0.084	0.95	1.0

[a]To capture potential mutational instability adjacent but not within G-quadruplex structures, each of the sequences as defined by Bedrat *et al.* (Nucleic Acids Res. 2016) was extended to include five bases upstream and five bases downstream.
[b]Bonferroni correction is done to adjust for multiple tests, which is 3 in this case.

sequence information to study the location and frequency of mt variants. SardiNIA is a genetically isolated founder population, and InCHIANTI provides a more mixed Italian cohort. The cohorts allowed the study of both frequent and rare mt homoplasmies (variants affecting all mtDNA copies within a cell compared to the reference sequence) and heteroplasmies (the mixture of two or more alleles). We used biophysically categorized G4 sequences to co-localize mt variants in each cohort to the mtDNA sequences forming G4 *in vitro*. Of the three categories of G4s—stable G4, unstable G4 and verified non-G4—our analyses showed a statistically significant enrichment of mt heteroplasmies and homoplasmies in stable G4 regions but not in unstable G4 or verified non-G4 regions. We used the bioinformatics results to guide our selection of two biologically relevant mtDNA sequences containing G4s for the *in vitro* biochemical experiments of mt G4 replication using purified recombinant mt replisome proteins. We present evidence that human mt Pol γ is potently blocked by G4s *in vitro*. We further showed that Twinkle, mtSSB and TFAM (an abundant mt dsDNA binding protein that also preferentially interacts with G4 DNA (24)) fail to stimulate synthesis through a G4 but a yeast version of the functionally conserved human Pif1 G4 helicase allows Pol γ to make a full-length product using the G4 DNA template. Lastly, we show that in the absence of other replisome proteins, PrimPol is better able to replicate G4s than Pol γ, albeit in an error-prone manner that persists even when Pif1 helicase facilitates DNA synthesis, suggesting that PrimPol may play a role in G4 replication *in vivo* but in an error-prone fashion. Our combined results demonstrate that G4s represent a potent block to Pol γ and PrimPol, supporting the notion that failure to replicate G4 structures contributes to polymerase stalling and the introduction of variants in the mt genome.

Results

Mitochondrial single nucleotide variants are statistically enriched in stable G4 sequences

To investigate whether mtDNA variants are enriched in regions that form G-quadruplex structures, we looked at homoplasmic and heteroplasmic single nucleotide variants (SNVs) in the three different G4 stability categories from the G4Hunter paper [stable G4 (conclusive evidence for G4 formation), unstable G4 (evidence both for and against G4 formation) and verified non-G4 (conclusive evidence that sequences do not form G4 structures) (25)]. mtDNA variants (homoplasmies and heteroplasmies) were identified from whole-genome sequences of 2757 SardiNIA and InCHIANTI project participants. Table 1 shows the number of unique homoplasmic or heteroplasmic sites identified in SardiNIA or InCHIANTI cohorts within a specific G4 category (e.g. stable G4). Homoplasmies were 40.0 and 30.2% more frequent than expected by chance in the stable G4 category of the SardiNIA and InCHIANTI cohorts, respectively (accounting for 21.9 and 20.3% of the homoplasmies identified in the two cohorts, respectively), with Bonferroni-corrected *P*-values of 2.2×10^{-6} and 8.1×10^{-6}, respectively (Table 1, see Supplemental Information S1 Results for an example of *P*-value calculation). By contrast, no significant enrichment was seen in the unstable G4 or verified non-G4 categories. We find consistent results when we consider only moderately common homoplasmies (50% or less of cohort) and rare variants (5% or less of cohort) (Supplementary Material, Table S1).

Heteroplasmies were also significantly enriched in the stable G4 category of SardiNIA (Bonferroni-corrected *P*-value: 4.4×10^{-3}, Table 1), with moderate enrichment in InCHIANTI (Bonferroni-corrected *P*-value: 0.084, Table 1). As the most frequent heteroplasmy was present in 2.9 and 6.5% of the SardiNIA and InCHIANTI cohorts, respectively, we did not stratify by moderate and rare variants for analysis of heteroplasmies. Thus, mtDNA variants were preferentially located within stable G4 sequences in both cohorts examined.

Figure 1 represents homoplasmic variants in the SardiNIA cohort co-localizing to stable G4, unstable G4 and verified non-G4 regions in three separate tracks. We note that 49 of the 57 stable G4 regions contain at least 1 homoplasmy identified from the SardiNIA cohort. Consistent with results from Table 1, after adjusting for region lengths, we saw ~30% more homoplasmies in stable G4 than in unstable or verified non-G4 regions. A similar pattern of enrichment was also observed in the InCHIANTI cohort (Supplementary Material, Fig. S1).

Enrichment of SNVs in loops, not G-runs, of stable G4 sequences

Given the enrichment of variants in stable G4 sequences, we next asked whether variants preferentially map to one of the

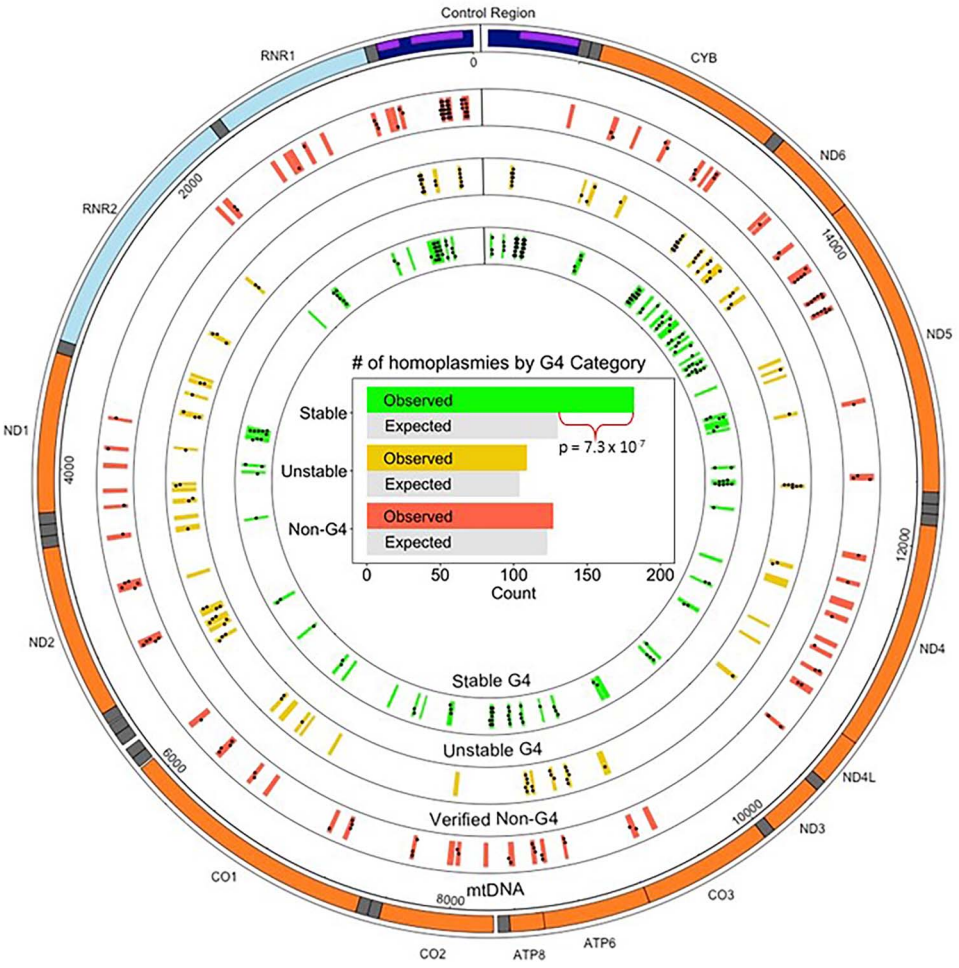


Figure 1. SardiNIA homoplasmic variants enriched in stable G4 regions, as compared to unstable and verified non-G4 regions. Visualization of circular mtDNA with the three G4 categories (stable G4, green; unstable G4, yellow; verified non-G4, red) and the locations of the SardiNIA homoplasmic variants that overlap with the specific G4 regions. The outermost track shows the locations of the 13 mtDNA genes, 2 rRNAs and 22 tRNAs along with the three hypervariable regions within the control region/D-loop, oriented counterclockwise. We note that SardiNIA homoplasmic variants overlap with stable G4 at a level 33% higher than unstable or verified non-G4 regions. The bar plot shows the number of SardiNIA homoplasmies localized in each of the three G4 categories (stable G4, green; unstable G4, yellow; verified non-G4, red) along with the expected number of homoplasmies in each G4 category (grey, see Supplementary Information for an explanation of how the expected number of variants in a category was calculated).

two features of a G4 structure: the G-runs (two or more consecutive guanines) or loops (any adenine, thymine, cytosine or single guanine). Indeed, more homoplasmies were observed in loop regions of stable G4 sequences than in loop regions of no G4 regions (definition of no G4 in Materials and Methods and Supplemental Information S1 Materials and Methods) (SardiNIA Bonferroni-corrected *P*-value: 1.2×10^{-7}; InCHIANTI Bonferroni-corrected *P*-value: 1.9×10^{-9}, Table 2). In contrast, we observed no enrichment of homoplasmies in G-runs of stable G4 sequences (Bonferroni-corrected *P*-value was 1.0 in both SardiNIA and InCHIANTI, Table 2). We found homoplasmic enrichment in loops but not G-runs of unstable G4 regions in the InCHIANTI cohort (Bonferroni-corrected *P*-value: 0.023, Table 2) with the same trend but without statistical significance in SardiNIA (unadjusted *P*-value: 0.071; Bonferroni-corrected *P*-value: 0.43, Table 2). Heteroplasmies followed a similar pattern: stable G4 sequences showed enrichment of variants in loops (SardiNIA Bonferroni-corrected *P*-value: 2.1×10^{-5}; InCHIANTI Bonferroni-corrected *P*-value: 1.7×10^{-4}, Table 3) but not in G-runs. Unstable G4 and verified non-G4 sequences showed no evidence of heteroplasmic enrichment in either cohort.

Suggestive evidence of pyrimidine bias in loops with homoplasmic variants

Research indicates that short loops enriched in pyrimidines demonstrate a higher thermodynamic stability *in vitro* than longer loops or those with a purine bias (26–29) and that thermodynamic stability affects genomic instability in the presence of G4s (28). Loops containing a homoplasmic variant from the SardiNIA cohort demonstrated a statistically significant pyrimidine bias (*P*-value 0.042; Supplementary Material, Table S2), with a similar trend that was not statistically significant in InCHIANTI (*P*-value 0.70; Supplementary Material, Table S2). However, this finding did not extend to heteroplasmies. We did not find an association between variant-per-base frequency and loop length in stable G4 regions. For a full discussion of the

Table 2. Statistical enrichment of homoplasmies in loops but not G-runs of stable G4 sequences in both SardiNIA and InCHIANTI

Homoplasmies	SardiNIA				InCHIANTI			
	Loops[a]		G-runs[a]		Loops[a]		G-runs[a]	
	G4 category	No G4	G4 category	No G4	G4 category	No G4	G4 category	No G4
Stable G4								
Observed/expected	98/57.6	483/523.4	43/40.6	59/61.4	151/93.3	791/848.7	60/55.8	80/84.2
P-value (chi-square test)	2.0×10^{-8}		0.64		3.2×10^{-10}		0.47	
Bonferroni *P*-value[b]	1.2×10^{-7}		1.0		1.9×10^{-9}		1.0	
Unstable G4								
Observed/expected	59/47.1	483/494.9	24/25.0	59/58.0	102/77.7	791/815.3	37/35.3	80/81.7
P-value (chi-square test)	0.071		0.81		3.9×10^{-3}		0.72	
Bonferroni *P*-value[b]	0.43		1.0		0.023		1.0	
Verified non-G4								
Observed/expected	51/50.4	483/483.6	16/22.2	59/52.8	97/83.9	791/804.1	30/32.6	80/77.4
P-value (chi-square test)	0.93		0.12		0.13		0.59	
Bonferroni *P*-value[b]	1.0		0.72		0.78		1.0	

[a]G-runs refer to two or more guanine residues in a row; Loops refer to guanine singletons along with adenine, cytosine, and thymine.
[b]Bonferroni correction is done to adjust for multiple tests, which is 6 in this case.

Table 3. Statistical enrichment of heteroplasmies in loops but not G-runs of stable G4 sequences in both SardiNIA and InCHIANTI

Heteroplasmies	SardiNIA				InCHIANTI			
	Loops[a]		G-runs[a]		Loops[a]		G-runs[a]	
	G4 category	No G4	G4 category	No G4	G4 category	No G4	G4 category	No G4
Stable G4								
Observed/expected	83/51.4	436/467.6	30/26.7	37/40.3	57/33.9	285/308.1	8/8.4	13/12.6
P-value (chi-square test)	3.5×10^{-6}		0.41		2.9×10^{-5}		0.87	
Bonferroni *P*-value[b]	2.1×10^{-5}		1.0		1.7×10^{-4}		1.0	
Unstable G4								
Observed/expected	51/42.4	436/444.6	16/16.0	37/37.0	35/27.8	285/292.2	4/5.1	13/11.9
P-value (chi-square test)	0.17		0.99		0.16		0.55	
Bonferroni *P*-value[b]	1.0		1.0		0.96		1.0	
Verified non-G4								
Observed/expected	46/45.5	436/436.5	13/14.8	37/35.2	32/29.9	285/287.1	3/4.7	13/11.3
P-value (chi-square test)	0.94		0.58		0.69		0.34	
Bonferroni *P*-value[b]	1.0		1.0		1.0		1.0	

[a]G-runs refer to two or more guanine residues in a row; Loops refer to guanine singletons along with adenine, cytosine, and thymine.
[b]Bonferroni correction is done to adjust for multiple tests, which is 6 in this case.

results characterizing loop variants, please see Supplemental Information S1 Results.

mtDNA hypervariable regions: CpG islands do not explain variant enrichment in stable G4 regions

We investigated two alternative explanations for the enrichment of homoplasmies and heteroplasmies in stable G4 regions—mtDNA hypervariable regions and CpG islands—but did not find meaningful contributions from either one (Supplementary Material, Tables S3 and S4; please see Supplemental Information S1 Results for a fuller discussion).

Variant spectrum does not differ across G4 categories

We attempted to characterize the variants found in the stable G4 regions, but found no significant differences among the various comparisons made. The homoplasmies and heteroplasmies from both cohorts did not show a significant difference regarding the transition/transversion ratio (Supplementary Material, Table S5) or the variant type (synonymous, non-synonymous, non-exonic, Supplementary Material, Table S6) comparing among the three G4 categories. Variants with different frequencies did not tend to map differentially to specific G4 categories (Supplementary Material, Table S7), nor did they map preferentially to G-runs or loops (Supplementary Material, Table S8). For a full discussion of the comparisons made and the supporting literature, see Supplemental. Information S1 Characterization of Variants Overall and Characterization of Variants in Loops sections.

Test for G4 action during DNA replication *in vitro*

The sum of the evidence *in vivo* suggests that G-quadruplexes tend to provoke mutations. A potential source of mutations would be caused by interference of mtDNA replication.

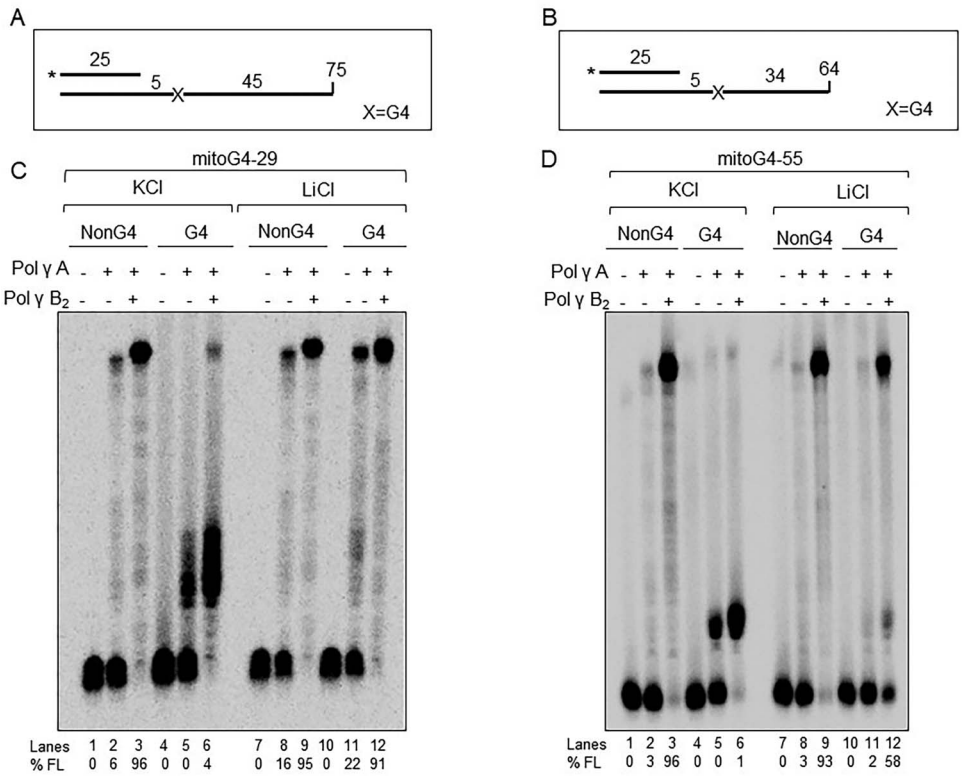


Figure 2. Pol γ arrest assays. (**A**, **B**) Schematic diagram of G4 oligonucleotide substrates. Templates are a 75-mer (A, mitoG4-29) or 64-mer (B, mitoG4-55). The G4 sequence begins at the +6 position (indicated as 'x5'). The G4-mer or 64-mer DNA product represents full extension of the 5′ ^{32}P-labeled (*) 25-mer primer to the end of the template. (**C**, **D**) Purified DNA Pol γ (Pol A) and its accessory subunit (Pol B) were incubated with G4 or non-G4 control templates for 10 min at 37°C in the presence of K^+ or Li^+ (50 mM) that favors or disfavors G4 formation, respectively.

Polymerase action would be modified at those sites—perhaps transiently slowed or stalled—and provoke misincorporation of nucleotides. To test for such effects, we have used standard *in vitro* replication systems with templates containing G4 constructs based on representative mtDNA sequences from the cohort data and enzymes putatively thought to be involved in the process.

Mitochondrial Pol γ stalls at G4 sites

To investigate how efficiently the mt Pol γ polymerizes through G4 sites, we used computational analysis results to design mtDNA templates (mitoG4-29 or mitoG4-55) that are capable of forming G4 DNA structures and had been determined to be sites of variation from the SardiNIA mt genome sequence data. We then performed *in vitro* DNA primer extension assays. DNA substrates were designed such that Pol γ would encounter the G4 structure at the 6th base after initiation of DNA synthesis off the primer. Control substrates contained a template lacking a G4 site but were otherwise identical in sequence and length. Substrates were prepared under conditions in which the G4 structure was stabilized (KCl) or destabilized (LiCl).

Under conditions in which the G4 structure was stabilized by the presence of K^+ ions, Pol γ holoenzyme (AB_2) extended the 25-mer end-labeled primer to a fully extended product on the non-G4 control DNA substrates [96% full-length (FL) product for either non-G4 substrate] (Fig. 2). However, Pol γ displayed significant polymerase arrest on both G4 DNA substrates in reaction mixtures containing K^+ (1%—4% FL product). We did observe some extension by Pol γ holoenzyme of up to five nucleotides (nt) past the predicted G4 site of mitoG4-29 template, but the amount of FL product was very low (4%). In contrast, Pol γ synthesis yielded a much greater fraction of fully extended product on the mitoG4-29 DNA substrate in the presence of Li^+ ions, particularly in the presence of the Pol γ B processivity subunit (91% FL product). Similarly, Pol γ reaction mixtures containing Li^+ enabled a much greater extension of the mitoG4-55 DNA substrate (58% FL product), consistent with the expected destabilization of the G4 structure by Li^+.

After observing strong blockage of Pol γ by stable G4 mtDNA templates, we looked at Pol γ replication of DNA templates classified as unstable G4 or verified non-G4 (Supplementary Material, Fig. S3). Results from primer extension assays with unstable G4 (UG4-53, UG4-85) or verified non-G4 (NG4-19, NG4-49) DNA templates in parallel with the stable G4 DNA templates (G4-29, G4-55) demonstrated efficient DNA synthesis on substrates harboring sequences corresponding to the unstable G4 (82% FL product, UG4-53; 78% FL product, UG4-85) or verified non-G4 (87% FL product, NG4-19; 84% FL product, NG4-49) under conditions of potent blockage by both stable G4s (0% FL product, mitoG4-29; 0% FL product, G4-55).

Effect of additional replisome proteins on Pol γ synthesis past G4 arrest site

We next tested purified recombinant mt proteins implicated in mtDNA replication (Twinkle, mtSSB) and the human TFAM which is known to preferentially bind G-quadruplex DNA (24) to assess if any of these proteins would enable Pol γ to synthesize past the G4 roadblock. Under conditions in which these mtDNA metabolic

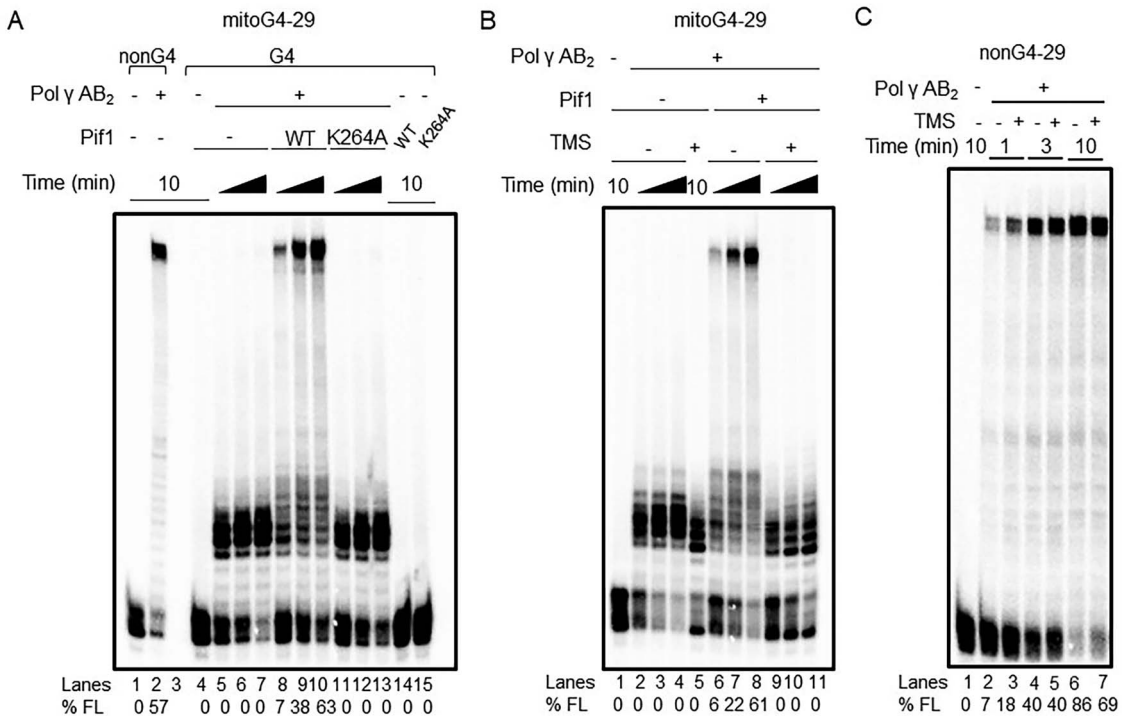


Figure 3. Pif1 stimulation of Pol γ primer extension past the G4 is dependent on its helicase activity. (**A**) A time course experiment was performed using the mitoG4-29 substrate (5 nM) in the presence or absence of Pol γ (15 nM Pol A, 30 nM Pol B), 26 nM Pif1 (WT or ATPase-dead K264A) and 4 mM ATP under the Pol γ reaction conditions described in Materials and Methods. Primer extension after 1, 3 and 10 min is shown. Activity of Pol γ on the non-G4-29 substrate is shown in lanes 1 and 2. (**B**) A kinetic experiment as described in Panel A except that the G4 DNA substrate was preincubated with 500 nM telomestatin (TMS) for 5 min prior to the addition of Pol γ and Pif1. (**C**) A time course experiment was performed on the non-G4 substrate (5 nM) preincubated with 500 nM TMS in the presence or absence of 15 nM Pol A, 30 nM Pol B, 26 nM Pif1-WT and 4 mM ATP. Primer extension after 1, 3 and 10 min is shown.

accessory factors bound the non-G4 and G4 DNA templates with similar affinity (Supplementary Material, Fig. S4 and Supplementary Material, Table S12), none of them stimulated synthesis by Pol γ past the stable G4 (Supplementary Material, Fig. S5). Reaction mixtures containing TFAM exhibited very poor synthesis by Pol γ from the primer end (irrespective of the presence of Twinkle and/or mtSSB), suggesting that complete binding of TFAM to the DNA substrate (Supplementary Material, Fig. S4) blocked access of the polymerase to the template (Supplementary Material, Fig. S5A and B, lanes 14, 16–18). To address this apparent inhibition of DNA synthesis, we titrated TFAM and observed inhibition of Pol γ synthesis in a concentration-dependent manner (1.9-fold reduction at 215 nM TFAM; 4.5-fold reduction at 860 nM TFAM) (Supplementary Material, Fig. S6A). Repeating primer extension experiments with a 16-fold lower concentration of TFAM (Supplementary Material, Fig. S6B) showed no inhibition of DNA synthesis (Supplementary Material, Fig. S6B). Despite the different experimental reaction conditions tested, no combination of the three proteins facilitated Pol γ synthesis through the G4 block.

Given limited extension by Pol γ on the stable G4 substrate even in the presence of additional replisome factors, we addressed whether Pif1, a known G4 resolvase (30) that localizes to mitochondria (31–34), may aid in this capacity *in vitro*. Pif1 stimulated Pol γ DNA synthesis past the G4 site under conditions in which both the polymerase and helicase were active, producing an increasing amount of fully extended product in a time-dependent manner (Fig. 3A), up to 63% after 10 min of incubation. We tested a catalytically inactive Walker A box mutant Pif1-K264A (35) for comparison and saw no full-length extension by Pol γ after 10 min, indicating Pif1's ability to resolve the G4 and allow for Pol γ synthesis is dependent upon Pif1's ATPase activity (Fig. 3A). Preincubation of the G4 substrate with the G-quadruplex ligand telomestatin (TMS) potently inhibited Pol γ DNA synthesis past the G4, even in the presence of Pif1 (Fig. 3B, no detectable FL product), suggesting that TMS-induced stabilization of the G4 prevented Pif1 from resolving the structure and enabling Pol γ synthesis past the G4 arrest site. In control experiments, TMS exerted little effect on Pol γ synthesis using the non-G4 DNA substrate as the template (Fig. 3C).

The ability of Pif1 to efficiently stimulate Pol γ DNA synthesis past the G4 block caused us to question if the effect of Pif1 is specific and if another G4-resolving helicase could substitute for Pif1. To address this issue, we tested purified recombinant FANCJ, a DNA helicase which we have shown to robustly resolve intramolecular and intermolecular G4 DNA substrates (36,37), in primer extension assays with the mitoG4-29 template. As shown in Fig. 4A, at all concentrations tested, FANCJ very poorly stimulated Pol γ DNA synthesis past the G4 block (~5% FL product) under conditions that Pif1 robustly stimulated Pol γ synthesis (~60% FL product). In contrast to the results with the G4 substrate, Pif1 partially inhibited Pol γ activity with the non-G4 substrate as evidenced by the reduced amount of FL product, whereas FANCJ enhanced Pol γ DNA synthesis on the control substrate, suggesting differential effects of the helicase proteins on Pol γ DNA synthesis.

To ensure that FANCJ was able to resolve the G4 structure formed by the DNA template under reaction conditions used for Pol γ primer extension assays, the G4 DNA template was incubated for 10 min with FANCJ in the presence of ATP and an oligonucleotide with complementary sequence to serve as a trap

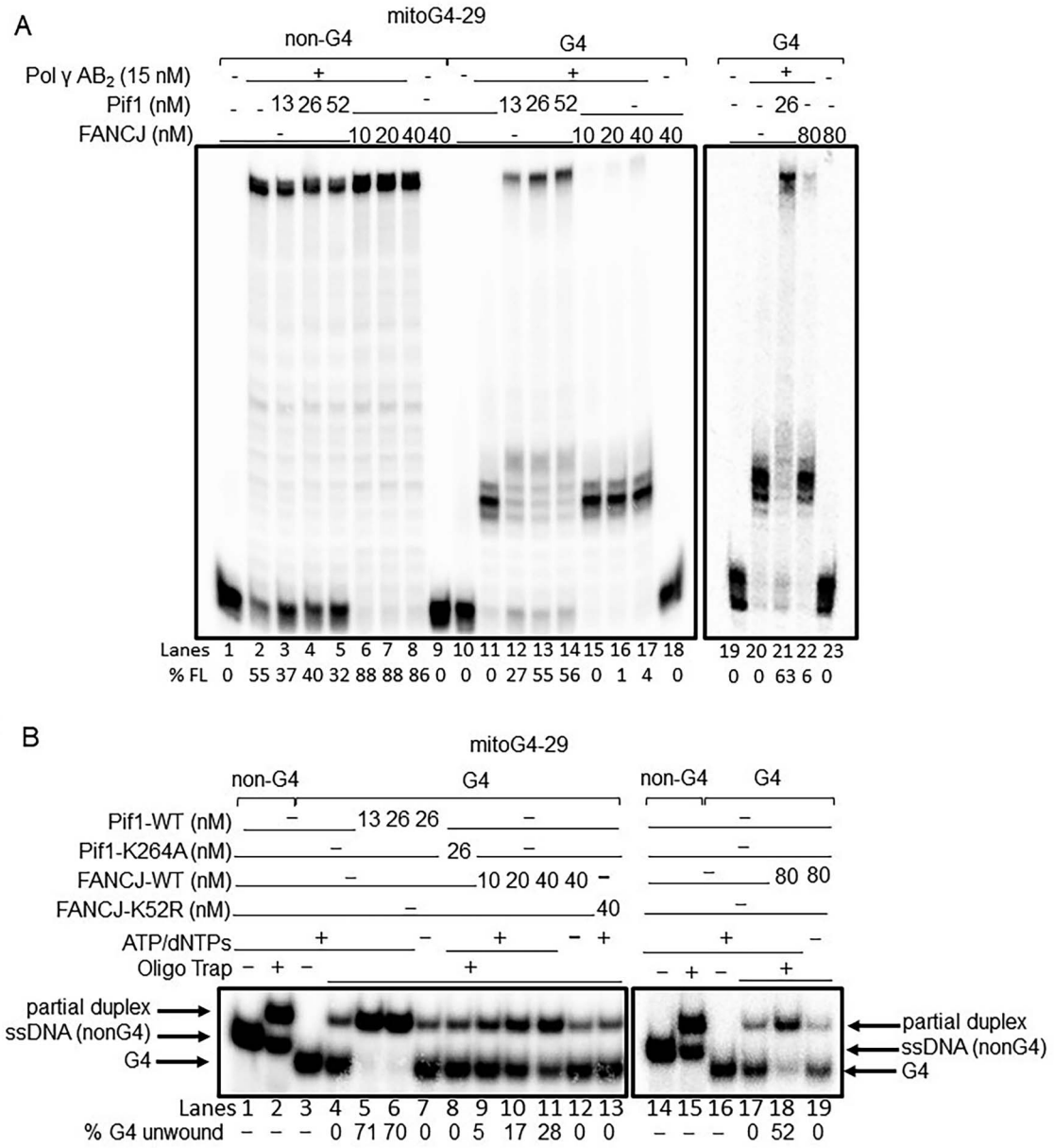


Figure 4. Pif1 specifically enhances DNA synthesis by Pol γ past the G4 in the DNA substrate. (**A**) Reaction mixtures included 15 nM Pol γ AB_2, 100 μM dNTPs and 4 mM ATP in the presence of varying concentrations of Pif1 and FANCJ. Primer extension on the mitoG4-29 and non-G4-29 DNA substrates after a 10 min incubation at 30ncubation at (**B**) Reaction mixtures included 100 μM dNTPs and 4 mM ATP in the presence of varying concentrations of Pif1 and FANCJ. ATPase-deficient Pif1 and FANCJ were included as controls. Percent mitoG4-29 or non-G4-29 DNA substrates unwound after a 10 min incubation at 30°C is shown by trapping unwound after template with a short ssDNA complementary to the G4 motif sequence when unwound. Background partial duplex from lanes 4 and 17 subtracted from total DNA unwound.

and reaction mixtures were resolved on native polyacrylamide gels (Fig. 4B). FANCJ unwound increasing percentages of the G4 substrate up to 52% in an ATP-dependent manner. A FANCJ-K562R ATPase mutant (37) failed to resolve the G4 substrate, indicating that the G4 resolvase activity was intrinsic to the purified recombinant FANCJ. These results indicate that Pif1 stimulates Pol γ DNA synthesis past the G4 block, whereas the G4-resolving helicase FANCJ does not.

PrimPol catalyzes synthesis beyond the Pol γ G4 arrest site

The domain architecture and catalytic activities of human PrimPol implicated in nuclear and mtDNA replication have been studied (38). We hypothesized that upon encountering a G4 site, PrimPol may be able to assist the mt replication machinery as a translesion synthesis polymerase, thereby allowing for faithful G4 replication. We first tested whether PrimPol can work alongside Pol γ to improve bypass of G4 lesions compared to Pol γ alone. We carried out primer extension assays under PrimPol conditions with 10 mM KCl present (contributed from the DNA substrate) in which Pol γ retains comparable activity. Using the same reaction mixture components and in the co-presence of 800 mM LiCl and incubation at 37°C for 30 min, both the mitoG4-29 and mitoG4-55 DNA sequences retained their G4 structure (Supplementary Material, Fig. S7). Importantly, retention of the G4 structures even in the presence of Li^+, which is known to prevent G4 from folding (39), attested to the

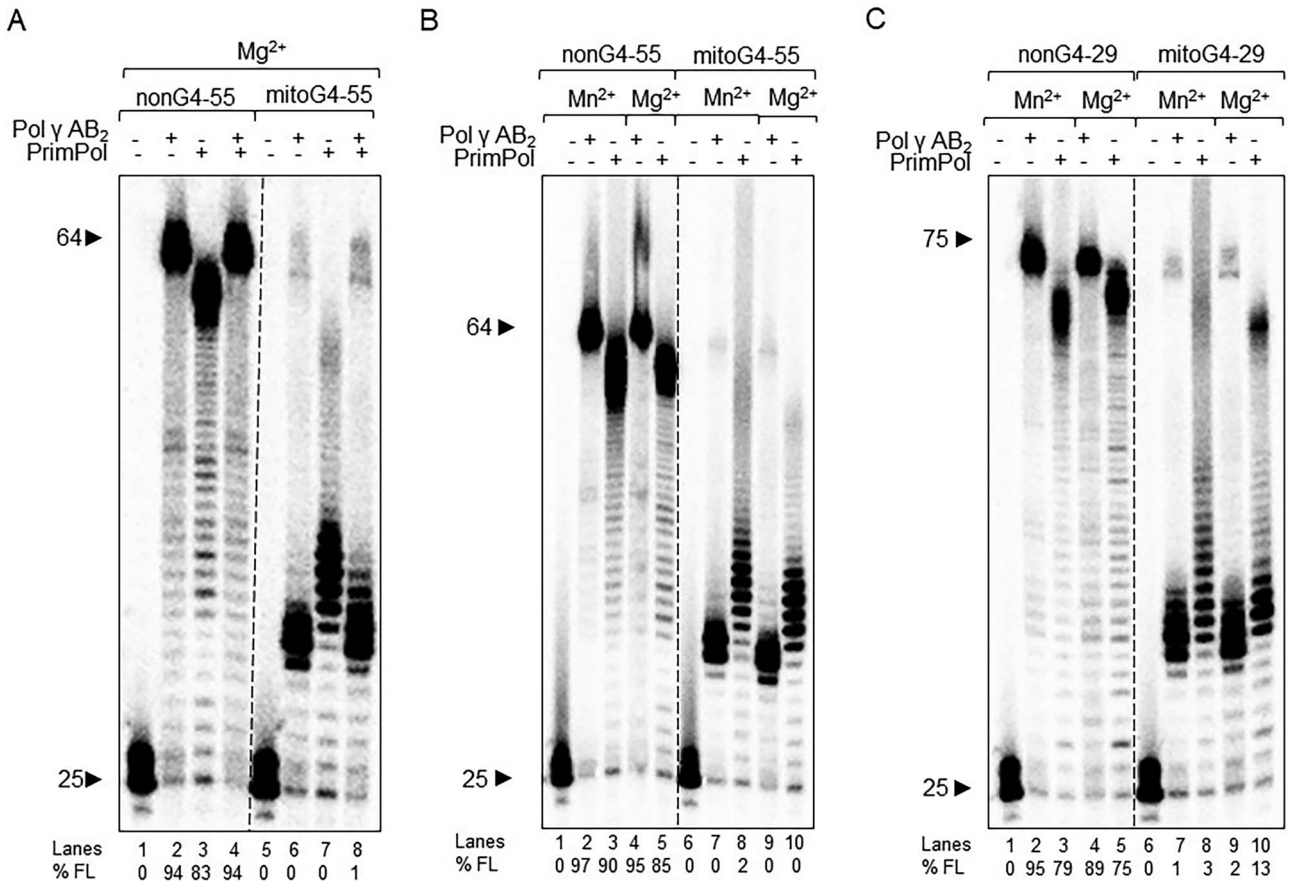


Figure 5. PrimPol catalyzes DNA synthesis beyond Pol γ G4 arrest site. Reaction mixtures for Pol γ AB_2 or PrimPol were identical and corresponded to the PrimPol reaction conditions as described in Materials and Methods; 10 mM KCl derived from the DNA substrate preparation was present in the reaction mixture. (**A**) PrimPol does not stimulate DNA synthesis by Pol γ. Primer extension in the presence of 5 nM substrate and 15 nM Pol γ AB_2, 200 nM PrimPol or both after 30 min at 37°C is shown. (**B**, **C**) Effect of Mn^{2+} and Mg^{2+} on DNA synthesis catalyzed by Pol γ or PrimPol. Reactions containing 15 nM Pol γ AB_2 or 200 nM PrimPol as indicated were conducted in the presence of 1 mM Mn^{2+} or 1 mM Mg^{2+} as described in Materials and Methods.

stability of the G4's formed by mitoG4-29 and mitoG4-55 DNA sequences.

As shown on the non-G4 DNA substrates, in contrast to Pol γ, PrimPol displayed difficulty catalyzing polynucleotide synthesis near the end of the DNA template (Fig. 5A, lanes 2 and 3), and did not significantly improve G4 bypass by Pol γ when both enzymes were present together (Fig. 5A, lane 8). PrimPol alone did, however, extend the substrate several nt further than Pol γ alone on the G4 substrate under identical reaction conditions (Fig. 5A, compare lanes 6 and 7). The observation that reactions with both polymerases produced DNA products resembling those produced by Pol γ alone is consistent with reports that Pol γ binds DNA with high affinity, whereas PrimPol exhibits weaker binding (40,41).

Given that under identical reaction conditions PrimPol was able to extend further than Pol γ after encountering the G4, we hypothesized that PrimPol may be better able to facilitate G4 synthesis in the presence of mt replication proteins. TFAM inhibited DNA synthesis by PrimPol in the 30 min reaction on both the control and G4 substrates (Supplementary Material, Fig. S8A), whereas MtSSB did not stimulate any G4 synthesis by PrimPol (Supplementary Material, Fig. S8A). Previously it was reported that Twinkle helicase stimulates primer extension catalyzed by PrimPol (42). In our experiments, 20 nM Twinkle demonstrated a modest inhibition of PrimPol primer extension on the control and G4 DNA substrates (Supplementary Material, Fig. S8B). However, using a 4-fold lower Twinkle concentration of 5 nM, we observed stimulation of PrimPol DNA synthesis on the control non-G4 DNA substrate but not on the G4 DNA substrates (Supplementary Material, Fig. S9).

It has been reported that the divalent cation Mn^{2+}, when substituted for Mg^{2+}, is able to stimulate PrimPol-catalyzed DNA synthesis (3,23). Therefore, we sought to test whether the presence of Mn^{2+} would allow PrimPol to better synthesize DNA past the G4, comparing the effect of Mn^{2+} to that observed in reaction mixtures containing Pol γ. Figure 5B shows that PrimPol-catalyzed DNA synthesis past the G4 on the mitoG4-55 substrate was not significantly different using Mn^{2+} as compared to Mg^{2+}. However, under identical reaction conditions, PrimPol catalyzed DNA synthesis better than Pol γ on both G4 substrates, as evident by the population of slower migrating, longer DNA products (Fig. 5B and C). PrimPol produced a small amount of mitoG4-29 fully extended product (13% FL product) in the presence of 1 mM Mg^{2+} (Fig. 5C) or 1 mM Mn^{2+} (data not shown). No fully extended product was observed for the mitoG4-55 substrate in the presence of either ion. Importantly, we noted that PrimPol underwent less severe polymerase arrest upon encountering the G4 compared to Pol γ. We observed Pol γ termination products starting six nucleotides beyond the primer (corresponding to the last position before the G4 structure). Upon

encountering the G4, extension by Pol γ was fully terminated within a range of 2–4 nt for both substrates, which can be equated to complete arrest within the first G-run. By contrast, PrimPol showed less potent polymerase arrest under identical conditions. PrimPol catalyzed synthesis well beyond the first G-run, extending through the interstitial loop and second G-run and into the second loop. Several additional extension products were observed beyond this position, indicating that a fraction of the DNA was extended beyond the G4 structure and, to a lesser extent, to a fully extended product.

The difference in concentration of the two polymerases (PrimPol at 200 nM, Pol γ at 15 nM) may contribute to the stronger replicative capacity of PrimPol. In polymerase titration experiments (Supplementary Material, Fig. S10), DNA synthesis by a Pol γ concentration of 120 nM using the G4 substrate as a template still showed a strong arrest at the first G-run (only 2% FL product) (Supplementary Material, Fig. S10A). For PrimPol reactions, DNA synthesis past the first G-run was evident at concentrations as low as 50 nM PrimPol (Supplementary Material, Fig. S10B). Although fully extended product by PrimPol with the G4 template was hardly detectable at lower concentrations, FL product for the non-G4 template was also significantly decreased in reaction mixtures containing PrimPol concentrations of 13–100 nM. Therefore, the experimental data suggest that even at more comparable polymerase concentrations, PrimPol shows a stronger ability to replicate past the G4 than Pol γ. Furthermore, our combined experiments to assess DNA synthesis by PrimPol using the G4 template confirmed that PrimPol catalyzes DNA replication further into the G4 structure than Pol γ, irrespective of cofactor ion choice.

As before for Pol γ, the deterrence of PrimPol DNA synthesis by mtDNA templates categorized as stable G4 DNA prompted us to test DNA templates classified as unstable G4 or verified non-G4 (Supplementary Material, Fig. S11). Results from primer extension assays with unstable G4 (UG4-53, UG4-85) or verified non-G4 (NG4-19, NG4-49) DNA templates run in parallel with the stable G4 DNA templates (mitoG4-29, mitoG4-55) demonstrated that PrimPol can efficiently perform DNA synthesis on substrates harboring sequences corresponding to the unstable G4 (72% FL product, UG4-53; 66% FL product, UG4-85) or verified non-G4 (87% FL product, NG4-19; 73% FL product, NG4-49) under conditions of potent blockage by both stable G4s (0% FL product, mitoG4-29; 0% FL product, mitoG4-55).

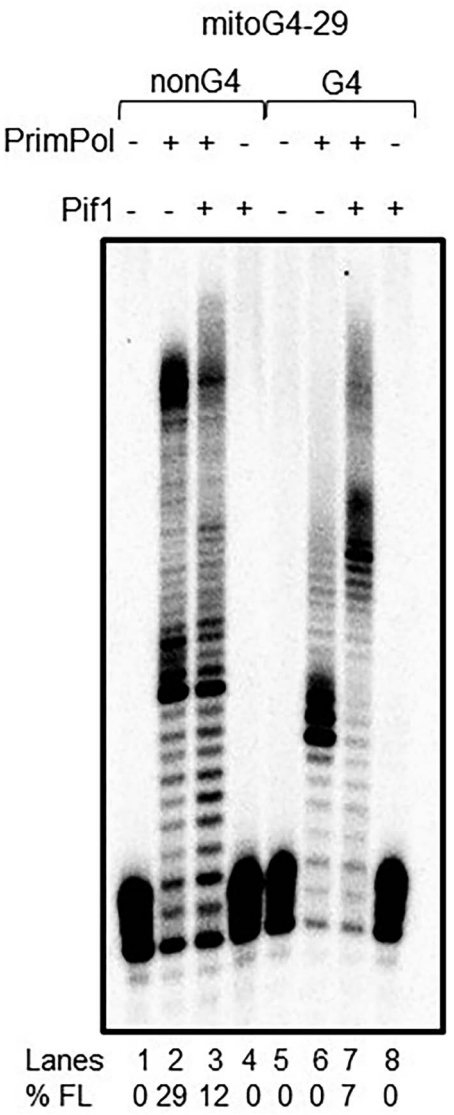
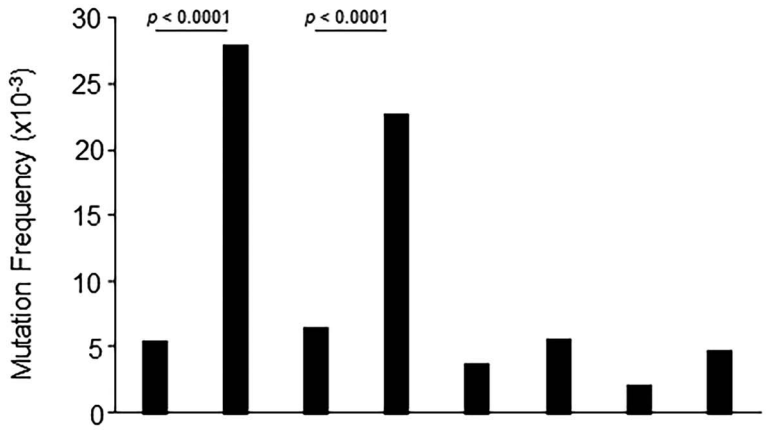


Figure 6. Pif1 facilitates PrimPol-catalyzed DNA synthesis past the G4. Reaction mixtures with PrimPol (200 nM) in the presence or absence of Pif1 (26 nM) were initiated with the addition of dNTPs and ATP as described in Materials and Methods except that the incubation temperature was 30°C, which was permissive of both PrimPol and Pif1 catalytic activity. Reactions were quenched with formamide loading dye and products resolved on 16% acrylamide 0.5XTBE denaturing gels.

Pif1 stimulates PrimPol to bypass the G4 in an error-prone manner

The ability of Pif1 to stimulate Pol γ to bypass the G4 raised the question if the G4-resolving helicase has a similar effect on PrimPol synthesis past G4s. As shown in Fig. 6, the strong block to DNA synthesis by PrimPol (0% FL) past the G4 was partly alleviated by the presence of Pif1, resulting in a shift to longer extended DNA, including 7% FL products. In contrast, Pif1 inhibited primer extension by PrimPol to some extent on the non-G4 DNA template, suggesting that the DNA-interacting Pif1 interferes with PrimPol-catalyzed DNA synthesis, similar to what was reported for single-stranded DNA binding proteins (43); however, Pif1 enables PrimPol to more efficiently perform DNA synthesis past the G4.

Pif1's enhancement of PrimPol's ability to bypass the G4 block led us to ask how this affected the fidelity of DNA synthesis. To do this, primer extension products were analyzed for sequence integrity. Analysis of PrimPol extension products in the absence of Pif1 showed the introduction of single base-pair substitutions, insertions (1–2 bp) and deletions (1–30 bp) (Supplementary Material, Fig. S12A and B). Quantification of mutated sequences is shown in Fig. 7. On a non-G4 substrate, PrimPol displayed a 5.6×10^{-3} mutation frequency, similar with a previous study (43). This error rate increased over 5-fold in the presence of a G4 structure. This high mutation frequency is retained even in the presence of Pif1 helicase (Fig. 7; Supplementary Material, Fig. S12C and D), suggesting the mutation rate is influenced by the intrinsic base-pair sequence itself in addition to the DNA structure; nonetheless, G4 poses a block to DNA synthesis by either PrimPol or Pol γ. Taken together, our *in vitro* data suggest that DNA replication by PrimPol can extend through stable G4 structures in the presence of Pif1 helicase; however, PrimPol bypass of the G4 aided by Pif1 helicase is still a mutagenic process.



PrimPol	+	+	+	+	-	-	-	-
Pol γ AB_2	-	-	-	-	+	+	+	+
Pif1	-	-	+	+	-	-	+	+
Substrate	NG4	G4	NG4	G4	NG4	G4	NG4	G4
Total seq.	63	125	176	103	110	482	52	79
Mutated seq.	9	37	17	31	11	17	4	9
Nuc. Synthesized	1798	1678	2905	1711	3417	2959	1730	1859
Mutations	10	47	19	39	13	17	4	9
Mutation Freq. (x10-3)	5.6	28.0	6.5	22.8	3.8	5.7	2.3	4.8

Figure 7. Faithful or mutagenic replication of G4 substrates. DNA polymerase fidelity was quantified as the percentage of mutated to total sequences derived from *in vitro* primer extension assays using the mitoG4-29 or sequence-related non-G4 DNA substrates. Statistical significance was calculated by Fisher exact test.

We next wanted to examine the influence of G-quadruplex structures on DNA replication fidelity by Pol γ. Analysis of primer extension through the G-quadruplex sequences shows that while DNA Pol γ is inefficient at extending through G-rich sequence (Fig. 2), the enzyme maintains similar replication fidelity to the non-G4 substrate when bypass does occur (Fig. 7; Supplementary Material, Fig. S13A and B). The mutation frequency of 3.8×10^{-3} is consistent with mutation frequencies from lacZ forward mutation assays (44,45). Furthermore, in the presence of Pif1 helicase, bypass efficiency of Pol γ is increased (Fig. 3) without compromising the enzyme's fidelity (Fig. 7; Supplementary Material, Fig. S13C and D).

Discussion

In this study, we combined bioinformatic and biochemical approaches to study the prevalence of G4s in mt, their association with mt variants in two large cohorts, and attempt to define features of the mechanism by which they are replicated. We tested three biophysically characterized categories of G4 sequences for variant enrichment, showing that stable G4 sequences have an enrichment of homoplasmies and heteroplasmies in the SardiNIA and InCHIANTI populations. What is the evidence that stable G4 and not another feature of mtDNA structure is the basis for the enrichment of mutations at those sites? Our analyses demonstrated that the enrichment of mt variants was only observed in stable G4s, but not in unstable or verified non-G4s, which can serve as suitable negative controls here. The rationale is that these three stability categories of G4s were selected as predicted G4 regions by G4Hunter using the same criteria (i.e. the same score threshold of 1.0 and window size of 25 were employed). They were classified into three different G4 categories only when six biophysical experimental techniques were used to test for evidence of G4 formation *in vitro*. Therefore, the three stability categories of G4s share similar genomic sequence elements (G-richness and G-skewness), and unstable or verified non-G4s can serve as negative controls for these DNA sequence elements also found in stable G4s.

We extended this analysis to show preferential enrichment of variation in the loop regions of stable G4s, consistent with the notion that tetrad guanines within the G4 structure are less prone to mutation or possibly shielded from DNA damage. Taken together, our computational analyses suggest that G4s in the mt genome are a potential source of mutagenesis. We also note that although stable G4 regions had 30–40% more variants than other regions in the mt genome, hundreds of variants were found and distributed across the mt genome in the two cohorts, and we believe most of them were natural variations.

Even though the complete protein composition of the mtDNA replisome is not known, biochemists have established *in vitro* conditions that are thought to reflect the action of major features *in vivo*. These are based primarily on the evidence that the major mt replicative polymerase is Pol γ which catalyzes DNA synthesis poorly on a G4 template. To the best of our knowledge, this and another very recently published study (46) are the first biochemical evidence that G4s act as a potent barrier to mt replication. Furthermore, Twinkle, mtSSB and TFAM were each unable to facilitate replication through stable G4s. Although TFAM was reported to bind G4 DNA with an affinity comparable to that of dsDNA (24), we found that it was unable to enhance the efficiency of Pol γ past the G4. Presumably TFAM bound the dsDNA region of the substrate, preventing productive

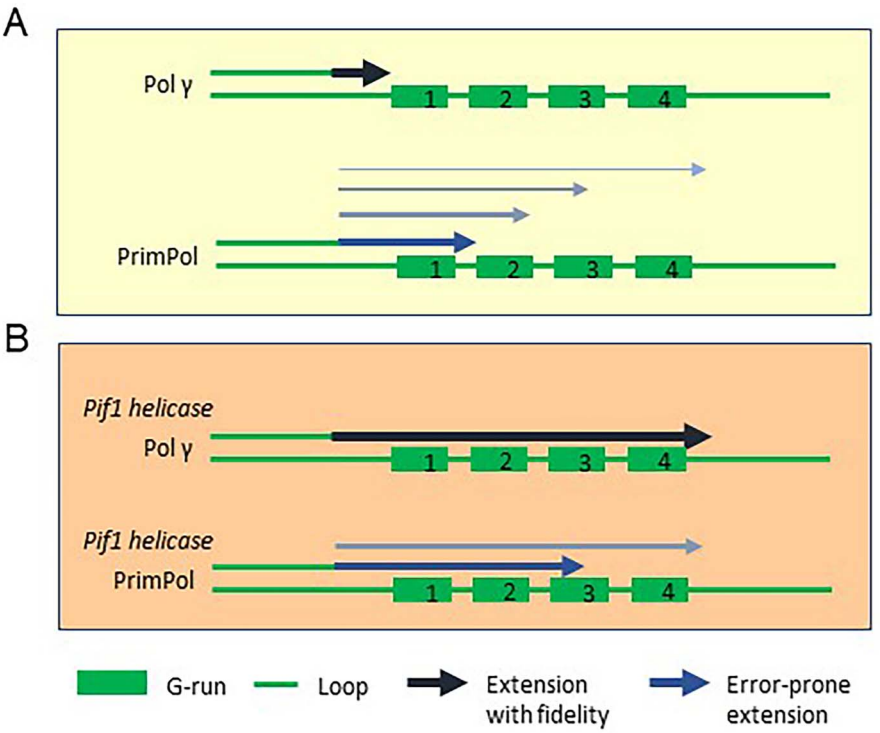


Figure 8. Model for stalling of DNA synthesis with differential fidelity by Pol γ versus PrimPol at G4 and enhancement by Pif1 helicase. (**A**) G4 structure stalls Pol γ within the first G-run. PrimPol catalyzes nucleotide incorporation using the template loop and adjacent G-runs, albeit in an error-prone manner. (**B**) Pif1 helicase enhances Pol γ and to a lesser extent PrimPol DNA synthesis past the G4; however, PrimPol bypass is still mutagenic.

utilization of the primer by the DNA polymerase. The *in vivo* relevance of this observation is unclear given that Pol γ acts on a RNA primer produced by the mt RNA polymerase to initiate DNA synthesis. The ability of catalytically active *Saccharomyces cerevisiae* Pif1 to stimulate Pol γ synthesis and produce fully extended product on the G4 template suggests that the helicase may aid in replicating mt G4 regions in eukaryotic cells—rather than Twinkle, which poorly resolves G4 DNA substrates *in vitro* (17). However, this suggestion remains conjectural and requires further testing. Overall, our biochemical results argue strongly for G4s as mitochondrial replication stalling elements (Fig. 8). Given the susceptibility to stalling at those sites, it is likely that the mt replication machinery relies on Pif1 (Fig. 8) (or another mt G4-resolving helicase) and one or more of the error-prone translesion polymerases to bypass G4s (see Introduction).

The most recently identified mt polymerase, PrimPol, is a candidate for G4 translesion synthesis. We showed that PrimPol catalyzes DNA synthesis farther than Pol γ using the G4 template under the same reaction conditions and exhibits some ability to fully bypass stable G4s (Fig. 5). Despite the evident stalling by PrimPol at mitochondrial G4 structures (this study) as well as at other model G-quadruplexes not derived from mt DNA sequence (20), our biochemical results suggest a potential role for PrimPol in G4 template-directed DNA synthesis. As suggested by Schiavone *et al.* (20), PrimPol may be better suited to replicate past relatively less stable G4 structures or catalyze DNA synthesis downstream of the G4. However, it may be relevant that the physiological concentration of PrimPol is likely to be much lower than Pol γ. Quantitative proteomic studies estimate the number of copies of Pol γ A (catalytic subunit) and Pol γ B_2 (accessory subunit) to be in range of 12 000–20 000 and 14 000–42 000, respectively (47,48). By comparison, Western blot data suggest that PrimPol is at a much lower copy number than Pol γ (A. Doherty, unpublished observations). Thus, it remains to be seen what mechanism(s) exist that regulate polymerase usage in mitochondria *in vivo* and in what contexts (e.g. replication fork stalling (41); see below).

We found through primer extension fidelity assays that catalysis by PrimPol on the mt G4 DNA substrate is highly error-prone, consistent with reports that other translesion polymerases, including Pol eta (η) and Pol kappa (κ), as well as the replicative polymerase Pol epsilon (ε), exhibit similarly compromised fidelity when copying through a G4 motif (49,50). Our *in vitro* data suggesting that PrimPol is highly mutagenic may lead one to expect a high mutation rate *in vivo* which is not the case. Presumably other pathways help to suppress mutation. Although some evidence for mammalian mt mismatch repair exists (51,52), the proteins involved appear to be distinct from nuclear mismatch repair (53). Further studies are required to characterize how mutagenesis is suppressed in the mt genome.

Taken together, our results demonstrate that PrimPol catalyzes DNA synthesis past the first G-run on its own, suggesting the polymerase may play a role with other mt proteins to destabilize the G4 structure. This partial unfolding of an intramolecular G4 might proceed through a multistep pathway that results in intermediary structures of some lesser stability, such as the G-triplex (G3), characterized by G:G:G triad planar stacks (54–56). It is tempting to speculate that the unique structure of the on-pathway G3 intermediate may bind PrimPol (or Pol γ) with different affinity compared to G4 or a single-stranded/double-stranded DNA junction, similar to what has been evidenced for Pif1 helicase (55), but that remains to be seen. We postulate that PrimPol's conversion of the first G-run in a stable G4 to duplex DNA would allow for other polymerases and replisome

proteins to disrupt the structured G-rich template. This should be further explored in future experiments, along with PrimPol's capacity to perform *de novo* priming downstream of the G4 arrest site. This repriming role for PrimPol is regarded as one of the enzyme's main functions in the nuclear genome (20,41,57); however, it has not been studied in the context of mt G4 replication. Although PRIMPOL−/− mice are viable, their embryonic fibroblasts or human cells silenced for PrimPol expression show reduction in mtDNA synthesis (3). Therefore, we sought to address whether PrimPol is capable of catalyzing DNA synthesis using a G4 template in a manner in which gaps would not be left behind that would require subsequent DNA synthesis or recombinational repair. Certainly, PrimPol's ability to bind and reprime DNA downstream of G4s *in vitro* (20,41,57) suggests a second mode whereby the DNA roadblock is overcome in a manner dependent on gap filling repair, whose mechanistic details in mtDNA metabolism are still not understood. Further studies with a fully reconstituted mt replisome and interacting proteins, such as Pif1 as evidenced in the current work or other DNA helicases localized to mt (e.g. Suv3 (58)), are required to detail the putative mechanisms of efficient G4 bypass.

Results from our biochemical reconstitution experiments and a very recently published work (46) demonstrate that Pif1 is capable of stimulating efficient DNA synthesis past the stable G4 structure by Pol γ or Pol δ, as well as PrimPol (this study), in a manner that is dependent on Pif1's ATPase/G4 resolvase activity. This may reflect Pif1's activity to periodically patrol intramolecular G4 DNA as reported by the Ha lab (59). The repeated cycles of G4 unfolding by Pif1 is important for its mechanism of action to keep the G4 resolved based on estimates that it refolds immediately (0.2 s) under conditions that Pif1 is active (59). Unlike Pif1, according to studies by Wu and Spies (60), FANCJ partially stabilizes G4 DNA structures and acts to unfold and refold G4 in repeated cycles. Another potential contributing factor for why Pif1 but not FANCJ enables efficient G4 bypass by Pol γ or PrimPol has to do with the distinct DNA substrate interactions of the two G4-resolving helicases. Pif1 has specificity for a double-stranded/single-stranded junction and patrols the nearby single-stranded to unfold G4 (59), whereas FANCJ targets the G4-associated single-stranded gaps (60). Therefore, Pif1's DNA binding specificity as well as its unique patrolling activity to keep G4 resolved may serve optimally to replicate the mitochondrial genome which has its own unique requirements and machinery compared to nuclear DNA replication. The human homologue of Pif1 shares sequence and functional similarity, but it remains possible that it would have different properties from those seen with the standard yeast nuclear version previously described (61,62) and used in the current study. Due to technical issues in terms of solubility and proteolysis, we were unable to purify recombinant human mt Pif1 protein; the reports on human Pif1 are based on truncated protein or full-length nuclear form (63,64).

A previous study demonstrated that inactivation of Pif1 causes a progressive mitochondrial myopathy in mice, resulting in reduced respiratory chain efficiency (31). Pif1 inactivation also leads to defective repair of oxidative mtDNA damage induced by exposure of mouse embryonic fibroblasts to H_2O_2; the elevated mtDNA damage can be alleviated in part by genetic complementation with a gene encoding the mitochondrial isoform of mouse Pif1 (31). These findings suggest a role of mammalian Pif1 in mtDNA maintenance. Presumably Pif1's ability to efficiently resolve G-quadruplexes abundant in the mt genome contributes to DNA synthesis past the G4 structures, but might not fully compensate for mtDNA mutations caused by PrimPol or perhaps other TLS polymerases.

The sequencing data from products of the *in vitro* primer extension assays demonstrate that PrimPol is much more error-prone than Pol γ during replication using the G4 template. However, given that Pol γ is able to efficiently replicate the G4 template with reasonably high fidelity when Pif1 is present, it may be questioned if the sites of mitochondrial genetic variation detected *in vivo* are really caused by PrimPol error-prone DNA synthesis. Further studies that assess the role of PrimPol in mtDNA replication *in vivo* are warranted. Recently, elegant studies by Torregrosa-Munumer *et al.* using 2D agarose gel electrophoresis demonstrate a role of mouse PrimPol to reprime stalled mtDNA replication forks (41). They go on to show that PrimPol is necessary for mtDNA replication after UV exposure and that PrimPol overexpression enhances the initiation of lagging strand DNA synthesis during mtDNA replication; however, the fidelity of DNA synthesis was not examined. It remains to be seen the relative importance of PrimPol versus Pol γ in replicating G4 structures that stall mtDNA replication in a physiological setting. It is plausible that the susceptibility of bases in the G4 DNA structures to damage such as that imposed by oxidative stress is a prominent source of mtDNA variation (see below).

Our findings that mt variants are enriched in stable G4 regions, combined with the observation that the major replicative mt polymerase undergoes potent arrest at these sites, support a model in which replisome stalling at G4s directly contributes to an increase in nucleotide misincorporation mutations within G4 sequences throughout the mt genome. Further studies with additional cohort data will be useful to determine a specific G4 mutagenic signature. G4s are known to be prevalent in telomeres and promoter regions of the nuclear genome and are believed to play important structural and functional roles in telomere capping and regulation of gene expression (65–68). It would therefore be interesting to confirm whether certain G4 sequences undergo selective evolutionary pressure to preserve the biological function of G4s by minimizing mutations that could potentially disrupt the Hoogsteen hydrogen bonding and tetrad stacking. An alternative explanation is that nucleotides in the vicinity of G-quadruplexes are more prone to DNA damage or refractory to repair, affecting the propensity for base-pair substitutions in the subsequent round of replication. Further studies are necessary to ascertain if the source of variation can be attributed to error-prone template-directed DNA synthesis, elevated sensitivity of guanine stacks or adjacent nucleotides to endogenous or exogenously induced stress (e.g. oxidative damage) and/or altered repair of specific DNA lesions located near G-quadruplexes. For example, the excision activities of various mammalian DNA glycosylases on G-quadruplex DNA structures with certain lesions (e.g. thymine glycol, 8-oxoguanine) varied significantly and in some cases (e.g. thymine glycol) depended strongly on the sequence context (69). However, to our knowledge, a comparative assessment of base lesion repair in the intervening sequences versus G-runs of G-quadruplex structures and comparison to duplex or single-stranded DNA has not been thoroughly performed. Clearly, further studies are required to determine to what extent the persistence of unique DNA structures in specific regions of the genome plays a role in nucleotide variation by compromised replication or repair.

In summary, our computational approach showed an enrichment of mt variants in the stable G4s. Consequent biochemical studies suggest that replication machinery stalling at G4 structures and reliance on error-prone DNA synthesis may contribute

to mutagenesis. Further studies are required to elucidate the precise biochemical mechanism by which variants are introduced and to better understand how G4 structures are faithfully replicated *in vivo*.

Materials and Methods

Whole-genome sequencing data

The SardiNIA whole-genome sequencing cohort includes 2077 individuals from the 6921 participants in the 'SardiNIA' study of the genetics of quantitative traits in the Sardinian founder population (70,71). DNA was extracted from whole blood samples, and paired-end sequence reads were generated with Illumina Genome Analyzer IIx and Illumina HiSeq2000 instruments to an average depth of 4.2X (72). The 25th percentile, median and the 75th percentile of the individual sample mtDNA average sequencing depth in the SardiNIA cohort are 156, 197 and 247, respectively. The InCHIANTI whole-genome sequencing cohort contains 680 InCHIANTI Project participants from Tuscany, Italy (73). DNA from buffy coat samples was sequenced using an Illumina HiSeq 2000 with a minimum read depth of 6X (74). The 25th percentile, median and the 75th percentile of the individual sample mtDNA average sequencing depth in the InCHIANTI cohort are 363, 432 and 511, respectively. Both experiments used the rCRS mtDNA reference sequence (NCBI Reference Sequence Number NC_012920.1) for alignment. In total, over 41 TB of whole-genome sequencing data were analyzed for this manuscript.

Ethics Statement: All SardiNIA participants gave written informed consent, with protocols approved by institutional review board of the National Institute on Aging (04-AG-N317). The InCHIANTI study baseline (1998–2000) was supported as a 'targeted project' (ICS110.1/RF97.71) by the Italian Ministry of Health and in part by the US National Institute on Aging (Contracts: 263 MD 9164 and 263 MD 821336). The study protocol was approved by the Italian National Institute of Research and Care of Aging Institutional Review and Medstar Research Institute (Baltimore, MD).

Homoplasmic and heteroplasmic variant sites

We used mitoCaller (72) to identify homoplasmic and heteroplasmic SNVs from whole-genome sequences with a likelihood-based calculation incorporating sequencing error rates at each base. Homoplasmies were conventionally defined as variants affecting all mtDNA copies within a cell compared to the reference sequence. Heteroplasmies were defined as the mixture of two or more alleles.

G-quadruplex (G4)-forming sequences

Bedrat *et al.* (25) created the G4Hunter algorithm to identify sequences predicted to form G4 structures and used this algorithm to probe the mt genome. They biophysically tested the sequences identified by G4Hunter to assign each one to one of the three stability categories *in vitro*: stable G4 showed conclusive evidence for G4 formation from all tests used; unstable G4 showed evidence both for and against G4 formation based on different tests; and verified non-G4 ('not-G4' in (25)) showed conclusive evidence of no G4 structures, despite being predicted by G4Hunter to do so. The G4Hunter algorithm produces results consistent with the literature; for example, 83% of sequences identified by Dong *et al.* (16) overlap with G4Hunter's set of predicted G4 sequences.

For information presented in Table 1, we extended the original G4Hunter sequences by 5 bp upstream and downstream to identify variants proximal to as well as within G4 sequences. Merging and overlapping these new sequences by G4 category yielded 57 stable G4 sequences (all heavy strand, 2593 bases covered), 51 unstable G4 sequences (all heavy strand, 2078 bases covered) and 62 verified non-G4 sequences (56 heavy strand, 6 light strand, 2456 bases covered).

For Tables 2 and 3, we divided the original sequences from the G4Hunter paper into the two types of regions present in G4 structures: G-runs (any run of two or more consecutive G-residues) and loops (all other bases, including single G-residues). We created a negative control by identifying the complement of all bases from the original sequences identified by G4Hunter (stable + unstable + verified non-G4, 4666 bases covered due to overlap between G4 categories and removing all light strand sequences, see Supplemental Information S1).

Statistical analyses of tables

We used a binomial test (Table 1; Supplementary Material, Tables S1, S3 and S4) to assess whether the observed number of variants in a G4-forming category exceeded the expectation under null (i.e. enrichment). We used a chi-square test (Tables 2 and 3; Supplementary Material, Tables S2 and S5–S10) to test for independence of two categorical variables in a contingency table.

Recombinant proteins

Recombinant Pol γ A (catalytic subunit) and Pol γ B_2 (accessory subunit), Twinkle helicase, TFAM, mtSSB and PrimPol were purified as previously described (21,75–77). Recombinant *S. cerevisiae* Pif1-WT and Pif1-K264A proteins (residues 40–859, nuclear form) were purified as described (61). Recombinant FANCJ was purified as described (78).

Preparation of DNA substrates

The oligonucleotide sequences chosen for substrates (Supplementary Material, Table S11) were derived from biologically relevant genes identified as sites of mt variation based on the SardiNIA genome sequence data. NADH dehydrogenase subunit 1 (mitoG4-29) and subunit 2 (mitoG4-55) variants were chosen based on their frequency in the SardiNIA population. MitoG4-29 and mitoG4-55 correspond to mtDNA sequences 3558–3595 and 5426–5465, respectively. 75-mer (mitoG4-29) or 64-mer (mitoG4-55) linear oligonucleotides harboring a G-quadruplex site were ordered from Lofstrand Labs Ltd and used as templates. Matching control oligonucleotides of the same lengths were also designed without a G4 site (non-G4-29 and non-G4-55). 10 pmol of 25-mer or 65-mer primer (to facilitate Twinkle loading at a 5′ overhang) was labeled at its 5′ end with 30 μCi of [γ-^{32}P]ATP and T4 polynucleotide kinase. Unincorporated ATP was removed using a G25 spin column (GE Healthcare Life Sciences), and the radiolabeled oligonucleotide was hybridized to the complementary non-G4 or G4 template strand (25 pmol) in the presence of 100 mM KCl or LiCl by incubating at 95°C for 5 min followed by slow cooling to room temperature (RT). For DNA substrates used in helicase assays and G4 stability assays, 10 pmol of non-G4 or G4 template was labeled at the 5′ end as above and incubated at 95°C for 5 min followed by slow cooling to RT in the presence of 100 mM KCl.

Primer extension assays

DNA substrates as described above were used in primer extension assays. The Pol γ primer extension reaction mixture (10 µl) contained 10 mM Tris-HCl pH 8.0, 50 mM KCl, 8 mM $MgCl_2$, 2 mM DTT, 100 µg/ml BSA, 5 nM labeled DNA substrate (which contributed an additional 10 mM KCl), 15 nM Pol γ A (catalytic subunit), 30 nM Pol γ B_2 (accessory subunit) and either 100 µM or 250 µM dNTPs. Primer extension reactions with PrimPol (200 nM, as previously used (3)) contained 10 mM Tris-HCl pH 6.8; 10 mM $MgCl_2$ or indicated concentration of $MnCl_2$, 1 mM DTT; and either 100 µM or 250 µM dNTP. KCl (10 mM) contributed by the DNA substrate was the only monovalent cation present in the reaction mixture to reduce inhibition of PrimPol enzymatic activity. Reactions with additional replication proteins contained 5–20 nM Twinkle (with 4 mM ATP present), 300 nM mtSSB and/or TFAM (monomer concentration indicated). Where specified, misincorporation assay reactions contained 5 mM $MgCl_2$ or $MnCl_2$ and 25 µM of one of the four dNTPs. Typical reaction mixtures were incubated for 1–30 min at 37°C and stopped by the addition of 5 µl of formamide loading buffer (20 mM formamide, 0.1% xylene cyanol, 0.1% bromophenol blue). For primer extension with FANCJ or Pif1 present, reactions were conducted as described in the figure legends and performed at 30°C to match the optimal temperature for helicase activity by Pif1 or FANCJ. DNA synthesis by Pol γ or PrimPol was moderately reduced at 30°C compared to 37°C. Reaction products were resolved on 16% polyacrylamide gels containing 8 M urea for 1 h and 45 min. Gels were imaged on a Typhoon system following exposure to detect primer extension by autoradiography.

Primer extension fidelity was analyzed by adding a single-strand oligonucleotide linker to the 3'-OH of both the extension product and the template oligos. Extension products (10 µl) were ligated in a 20 µl reaction containing 1× CircLigase II buffer, 2.5 mM $MnCl_2$, 1 M betaine, 0.3125 nM linker oligo (5′ phos-AGANNNNNNAGATCGGAAGAGCACACGTCTGAACTCCAGTCAC-biotin 3′) and 2.5 units CircLigase II (Lucigen) by heating to 60°C for 1 h. Ligation reactions were inactivated by heating to 80°C for 10 min. To differentiate the extension products from template, ligation reactions (2 µl) were PCR amplified over 30 cycles in a 50 µl reaction using Herculase II high-fidelity DNA polymerase (Agilent), extension primer oligonucleotide (5′ CGCCAGGGTTTTCCCAGTCACGACC 3′) and an internal linker oligonucleotide (5′ GTGACTGGAGTTCAGACGTGTGCTCTTCC-GATCT 3′). PCR products were ligated into Strataclone vector (Agilent), and the ligation mixture was transformed into Strataclone competent bacterial cells (Agilent); transformant DNA was Sanger sequenced (Macrogen USA). Unique extension products were scored using the 'NNNNNN' oligo tag and aligned to template sequence. Mutation frequency was calculated by dividing the number of mutations (includes insertions, deletions and substitutions) by the total number of nucleotides extended. Statistical analysis was performed using Fisher exact test.

Electrophoretic mobility shift assays

Gel shift assays were performed using reaction conditions described above and the indicated concentrations of mt replication proteins or TFAM. EMSA reactions were run for 30 min at 37°C and were stopped by the addition of 6X native dye (74% glycerol, 0.01% xylene cyanol 0.01% bromophenol blue). EMSAs were analyzed by resolution of radiolabeled DNA and DNA-protein complexes on 5% polyacrylamide non-denaturing gels and visualized using autoradiography.

Helicase assays

Pif1 or FANCJ helicase was incubated with non-G4 or G4 5′-^{32}P-labeled template strand in standard Pol γ primer extension reaction conditions at 30°C for 10 min. In addition, a trap oligonucleotide (10 nM final concentration) complementary to the G4 sequence (or the same region in the non-G4 template) was added during reaction initiation to prevent G4 formation after resolution by the indicated helicase. 10 µl of loading dye (0.04% bromophenol blue, 0.04% xylene cyanol, 25% glycerol, 50 mM KCl) was added to the reactions, and products were resolved on 8% 0.5X TBE non-denaturing PAGE gels with 10 mM KCl at 150 V for 2.5 h.

G4 stability assays

Non-G4 or G4 5′-^{32}P-labeled template strands were incubated under standard PrimPol primer extension reaction conditions in the presence of 800 mM LiCl and without additional KCl (the only contribution of KCl was from the substrate preparation resulting in a final concentration of 10 mM KCl) at 37°C for 30 min. In addition, varying concentrations of trap oligonucleotide complementary to the G4 sequence (or the same region in the non-G4 template) were added during initiation of incubation period to prevent G4 formation after unfolding. 10 µl of loading dye (0.04% bromophenol blue, 0.04% xylene cyanol, 25% glycerol) was added to the reactions after the 30 min incubation, and products were resolved on 8% 0.5X TBE non-denaturing PAGE gels (no KCl) at 150 V for 2.5 h.

Data Access

Bioinformatic resource

We provide a public session on the UCSC Genome Browser that allows users to visualize, access and explore our data. The session includes the locations of the three different G4 regions and the locations of all homoplasmies and heteroplasmies from the SardiNIA and InCHIANTI cohorts. Users can click 'Public Sessions' under the 'My Data' tab on the UCSC Genome Browser main page and enter 'butlertj3' in the search bar (username used to generate the public session) and select the session named 'mtDNA_G4_homo_hetero'.

Supplementary Material, Fig. S2 provides a snapshot of the information available in the public session.

Supplementary Material

Supplementary Material is available at *HMG* online.

Acknowledgements

We wish to thank Dr Kazuo Shin-Ya (National Institute of Advanced Industrial Science and Technology) for kindly providing telomestatin. We thank Jacob Crouch (NIA-NIH) for the assistance with some biochemical assays and Andre Rush (NIA-NIH) for the preliminary computational work.

Funding

Intramural Research Program of the National Institutes of Health, National Institute on Aging; Biotechnology and Biological Sciences Research Council (BB/H019723/1, BB/M008800/1 to A.J.D.'s laboratory); University of Sussex PhD studentship (to

T.A.G.); National Institutes of Health grant R35 GM122601 to KDR.

Conflict of Interest statement

The authors have no conflict of interest.

References

1. Falkenberg, M. (2018) Mitochondrial DNA replication in mammalian cells: overview of the pathway. *Essays Biochem.*, **62**, 287–296.
2. Bailey, L.J., Bianchi, J. and Doherty, A.J. (2019) PrimPol is required for the maintenance of efficient nuclear and mitochondrial DNA replication in human cells. *Nucleic Acids Res.*, **47**, 4026–4038.
3. Garcia-Gomez, S., Reyes, A., Martinez-Jimenez, M.I., Chocron, E.S., Mouron, S., Terrados, G., Powell, C., Salido, E., Mendez, J., Holt, I.J. *et al.* (2013) PrimPol, an archaic primase/polymerase operating in human cells. *Mol. Cell*, **52**, 541–553.
4. Singh, B., Li, X., Owens, K.M., Vanniarajan, A., Liang, P. and Singh, K.K. (2015) Human REV3 DNA polymerase zeta localizes to mitochondria and protects the mitochondrial genome. *PLoS One*, **10**, e0140409.
5. Wisnovsky, S., Jean, S.R. and Kelley, S.O. (2016) Mitochondrial DNA repair and replication proteins revealed by targeted chemical probes. *Nat. Chem. Biol.*, **12**, 567–573.
6. Sykora, P., Kanno, S., Akbari, M., Kulikowicz, T., Baptiste, B.A., Leandro, G.S., Lu, H., Tian, J., May, A., Becker, K.A. *et al.* (2017) DNA polymerase beta participates in mitochondrial DNA repair. *Mol. Cell. Biol.* **37**, pii: e00237–e00217.
7. Pohjoismaki, J.L., Wanrooij, S., Hyvarinen, A.K., Goffart, S., Holt, I.J., Spelbrink, J.N. and Jacobs, H.T. (2006) Alterations to the expression level of mitochondrial transcription factor A, TFAM, modify the mode of mitochondrial DNA replication in cultured human cells. *Nucleic Acids Res.*, **34**, 5815–5828.
8. Hansel-Hertsch, R., Di Antonio, M. and Balasubramanian, S. (2017) DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. *Nat Rev. Mol. Cell. Biol.*, **18**, 279–284.
9. Kaguni, L.S. and Clayton, D.A. (1982) Template-directed pausing in in vitro DNA synthesis by DNA polymerase a from Drosophila melanogaster embryos. *Proc. Natl. Acad. Sci. U. S. A.*, **79**, 983–987.
10. Kamath-Loeb, A.S., Loeb, L.A., Johansson, E., Burgers, P.M. and Fry, M. (2001) Interactions between the Werner syndrome helicase and DNA polymerase delta specifically facilitate copying of tetraplex and hairpin structures of the d(CGG)n trinucleotide repeat sequence. *J. Biol. Chem.*, **276**, 16439–16446.
11. Lormand, J.D., Buncher, N., Murphy, C.T., Kaur, P., Lee, M.Y., Burgers, P., Wang, H., Kunkel, T.A. and Opresko, P.L. (2013) DNA polymerase delta stalls on telomeric lagging strand templates independently from G-quadruplex formation. *Nucleic Acids Res.*, **41**, 10323–10333.
12. Lemmens, B., van Schendel, R. and Tijsterman, M. (2015) Mutagenic consequences of a single G-quadruplex demonstrate mitotic inheritance of DNA replication fork barriers. *Nat. Commun.*, **6**, 8909.
13. Lopes, J., Piazza, A., Bermejo, R., Kriegsman, B., Colosio, A., Teulade-Fichou, M.P., Foiani, M. and Nicolas, A. (2011) G-quadruplex-induced instability during leading-strand replication. *EMBO J.*, **30**, 4033–4046.
14. Wanrooij, P.H., Uhler, J.P., Shi, Y., Westerlund, F., Falkenberg, M. and Gustafsson, C.M. (2012) A hybrid G-quadruplex structure formed between RNA and DNA explains the extraordinary stability of the mitochondrial R-loop. *Nucleic Acids Res.*, **40**, 10334–10344.
15. Wanrooij, P.H., Uhler, J.P., Simonsson, T., Falkenberg, M. and Gustafsson, C.M. (2010) G-quadruplex structures in RNA stimulate mitochondrial transcription termination and primer formation. *Proc. Natl. Acad. Sci. U. S. A.*, **107**, 16072–16077.
16. Dong, D.W., Pereira, F., Barrett, S.P., Kolesar, J.E., Cao, K., Damas, J., Yatsunyk, L.A., Johnson, F.B. and Kaufman, B.A. (2014) Association of G-quadruplex forming sequences with human mtDNA deletion breakpoints. *BMC Genomics*, **15**, 677.
17. Bharti, S.K., Sommers, J.A., Zhou, J., Kaplan, D.L., Spelbrink, J.N., Mergny, J.L. and Brosh, R.M., Jr. (2014) DNA sequences proximal to human mitochondrial DNA deletion breakpoints prevalent in human disease form G-quadruplexes, a class of DNA structures inefficiently unwound by the mitochondrial replicative twinkle helicase. *J. Biol. Chem.*, **289**, 29975–29993.
18. Eddy, S., Ketkar, A., Zafar, M.K., Maddukuri, L., Choi, J.Y. and Eoff, R.L. (2014) Human Rev1 polymerase disrupts G-quadruplex DNA. *Nucleic Acids Res.*, **42**, 3272–3285.
19. Haracska, L., Prakash, S. and Prakash, L. (2002) Yeast Rev1 protein is a G template-specific DNA polymerase. *J. Biol. Chem.*, **277**, 15546–15551.
20. Schiavone, D., Jozwiakowski, S.K., Romanello, M., Guilbaud, G., Guilliam, T.A., Bailey, L.J., Sale, J.E. and Doherty, A.J. (2016) PrimPol is required for replicative tolerance of G quadruplexes in vertebrate cells. *Mol. Cell*, **61**, 161–169.
21. Bianchi, J., Rudd, S.G., Jozwiakowski, S.K., Bailey, L.J., Soura, V., Taylor, E., Stevanovic, I., Green, A.J., Stracker, T.H., Lindsay, H.D. *et al.* (2013) PrimPol bypasses UV photoproducts during eukaryotic chromosomal DNA replication. *Mol. Cell*, **52**, 566–573.
22. Zafar, M.K., Ketkar, A., Lodeiro, M.F., Cameron, C.E. and Eoff, R.L. (2014) Kinetic analysis of human PrimPol DNA polymerase activity reveals a generally error-prone enzyme capable of accurately bypassing 7,8-dihydro-8-oxo-2′-deoxyguanosine. *Biochemistry*, **53**, 6584–6594.
23. Mouron, S., Rodriguez-Acebes, S., Martinez-Jimenez, M.I., Garcia-Gomez, S., Chocron, S., Blanco, L. and Mendez, J. (2013) Repriming of DNA synthesis at stalled replication forks by human PrimPol. *Nat. Struct. Mol. Biol.*, **20**, 1383–1389.
24. Lyonnais, S., Tarres-Sole, A., Rubio-Cosials, A., Cuppari, A., Brito, R., Jaumot, J., Gargallo, R., Vilaseca, M., Silva, C., Granzhan, A. *et al.* (2017) The human mitochondrial transcription factor A is a versatile G-quadruplex binding protein. *Sci. Rep.*, **7**, 43992.
25. Bedrat, A., Lacroix, L. and Mergny, J.L. (2016) Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.*, **44**, 1746–1759.
26. Cheng, M., Cheng, Y., Hao, J., Jia, G., Zhou, J., Mergny, J.L. and Li, C. (2018) Loop permutation affects the topology and stability of G-quadruplexes. *Nucleic Acids Res.*, **46**, 9264–9275.
27. Guedin, A., Gros, J., Alberti, P. and Mergny, J.L. (2010) How long is too long? Effects of loop size on G-quadruplex stability. *Nucleic Acids Res.*, **38**, 7858–7868.
28. Piazza, A., Adrian, M., Samazan, F., Heddi, B., Hamon, F., Serero, A., Lopes, J., Teulade-Fichou, M.P., Phan, A.T. and Nicolas, A. (2015) Short loop length and high thermal stability

determine genomic instability induced by G-quadruplex-forming minisatellites. *EMBO J.*, **34**, 1718–1734.
29. Sattin, G., Artese, A., Nadai, M., Costa, G., Parrotta, L., Alcaro, S., Palumbo, M. and Richter, S.N. (2013) Conformation and stability of intramolecular telomeric G-quadruplexes: sequence effects in the loops. *PLoS One*, **8**, e84113.
30. Ribeyre, C., Lopes, J., Boule, J.B., Piazza, A., Guedin, A., Zakian, V.A., Mergny, J.L. and Nicolas, A. (2009) The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences in vivo. *PLoS Genet.*, **5**, e1000475.
31. Bannwarth, S., Berg-Alonso, L., Auge, G., Fragaki, K., Kolesar, J.E., Lespinasse, F., Lacas-Gervais, S., Burel-Vandenbos, F., Villa, E., Belmonte, F. *et al.* (2016) Inactivation of Pif1 helicase causes a mitochondrial myopathy in mice. *Mitochondrion*, **30**, 126–137.
32. Futami, K., Shimamoto, A. and Furuichi, Y. (2007) Mitochondrial and nuclear localization of human Pif1 helicase. *Biol. Pharm. Bull.*, **30**, 1685–1692.
33. Kazak, L., Reyes, A., Duncan, A.L., Rorbach, J., Wood, S.R., Brea-Calvo, G., Gammage, P.A., Robinson, A.J., Minczuk, M. and Holt, I.J. (2013) Alternative translation initiation augments the human mitochondrial proteome. *Nucleic Acids Res.*, **41**, 2354–2369.
34. Lahaye, A., Stahl, H., Thines-Sempoux, D. and Foury, F. (1991) PIF1: a DNA helicase in yeast mitochondria. *EMBO J.*, **10**, 997–1007.
35. Byrd, A.K. and Raney, K.D. (2015) A parallel quadruplex DNA is bound tightly but unfolded slowly by pif1 helicase. *J. Biol. Chem.*, **290**, 6482–6494.
36. Bharti, S.K., Sommers, J.A., George, F., Kuper, J., Hamon, F., Shin-ya, K., Teulade-Fichou, M.P., Kisker, C. and Brosh, R.M., Jr. (2013) Specialization among iron-sulfur cluster helicases to resolve G-quadruplex DNA structures that threaten genomic stability. *J. Biol. Chem.*, **288**, 28217–28229.
37. Wu, Y., Shin-ya, K. and Brosh, R.M., Jr. (2008) FANCJ helicase defective in Fanconia anemia and breast cancer unwinds G-quadruplex DNA to defend genomic stability. *Mol. Cell. Biol.*, **28**, 4116–4128.
38. Keen, B.A., Jozwiakowski, S.K., Bailey, L.J., Bianchi, J. and Doherty, A.J. (2014) Molecular dissection of the domain architecture and catalytic activities of human PrimPol. *Nucleic Acids Res.*, **42**, 5830–5845.
39. Sen, D. and Gilbert, W. (1992) Guanine quartet structures. *Methods Enzymol.*, **211**, 191–199.
40. Lim, S.E., Longley, M.J. and Copeland, W.C. (1999) The mitochondrial p55 accessory subunit of human DNA polymerase gamma enhances DNA binding, promotes processive DNA synthesis, and confers N-ethylmaleimide resistance. *J. Biol. Chem.*, **274**, 38197–38203.
41. Torregrosa-Munumer, R., Forslund, J.M.E., Goffart, S., Pfeiffer, A., Stojkovic, G., Carvalho, G., Al-Furoukh, N., Blanco, L., Wanrooij, S. and Pohjoismaki, J.L.O. (2017) PrimPol is required for replication reinitiation after mtDNA damage. *Proc. Natl. Acad. Sci. U. S. A.*, **114**, 11398–11403.
42. Stojkovic, G., Makarova, A.V., Wanrooij, P.H., Forslund, J., Burgers, P.M. and Wanrooij, S. (2016) Oxidative DNA damage stalls the human mitochondrial replisome. *Sci. Rep.*, **6**, 28942.
43. Guilliam, T.A., Jozwiakowski, S.K., Ehlinger, A., Barnes, R.P., Rudd, S.G., Bailey, L.J., Skehel, J.M., Eckert, K.A., Chazin, W.J. and Doherty, A.J. (2015) Human PrimPol is a highly error-prone polymerase regulated by single-stranded DNA binding proteins. *Nucleic Acids Res.*, **43**, 1056–1068.
44. Kunkel, T.A. and Mosbaugh, D.W. (1989) Exonucleolytic proofreading by a mammalian DNA polymerase. *Biochemistry*, **28**, 988–995.
45. Song, S., Pursell, Z.F., Copeland, W.C., Longley, M.J., Kunkel, T.A. and Mathews, C.K. (2005) DNA precursor asymmetries in mammalian tissue mitochondria and possible contribution to mutagenesis through reduced replication fidelity. *Proc. Natl. Acad. Sci. U. S. A.*, **102**, 4990–4995.
46. Sparks, M.A., Singh, S.P., Burgers, P.M. and Galletto, R. (2019) Complementary roles of Pif1 helicase and single stranded DNA binding proteins in stimulating DNA replication through G-quadruplexes. *Nucleic Acids Res.*, **47**, 8595–8605.
47. Hein, M.Y., Hubner, N.C., Poser, I., Cox, J., Nagaraj, N., Toyoda, Y., Gak, I.A., Weisswange, I., Mansfeld, J., Buchholz, F. *et al.* (2015) A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell*, **163**, 712–723.
48. Itzhak, D.N., Tyanova, S., Cox, J. and Borner, G.H. (2016) Global, quantitative and dynamic mapping of protein subcellular localization. *Elife*, **5**, pii: e16950.
49. Eddy, S., Tillman, M., Maddukuri, L., Ketkar, A., Zafar, M.K. and Eoff, R.L. (2016) Human translesion polymerase kappa exhibits enhanced activity and reduced fidelity two nucleotides from G-quadruplex DNA. *Biochemistry*, **55**, 5218–5229.
50. Eddy, S., Maddukuri, L., Ketkar, A., Zafar, M.K., Henninger, E.E., Pursell, Z.F. and Eoff, R.L. (2015) Evidence for the kinetic partitioning of polymerase activity on G-quadruplex DNA. *Biochemistry*, **54**, 3218–3230.
51. de Souza-Pinto, N.C., Mason, P.A., Hashiguchi, K., Weissman, L., Tian, J., Guay, D., Lebel, M., Stevnsner, T.V., Rasmussen, L.J. and Bohr, V.A. (2009) Novel DNA mismatch-repair activity involving YB-1 in human mitochondria. *DNA Repair*, **8**, 704–719.
52. Mason, P.A., Matheson, E.C., Hall, A.G. and Lightowlers, R.N. (2003) Mismatch repair activity in mammalian mitochondria. *Nucleic Acids Res.*, **31**, 1052–1058.
53. Kazak, L., Reyes, A. and Holt, I.J. (2012) Minimizing the damage: repair pathways keep mitochondrial DNA intact. *Nat. Rev. Mol. Cell. Biol.*, **13**, 659–671.
54. Hou, X.M., Fu, Y.B., Wu, W.Q., Wang, L., Teng, F.Y., Xie, P., Wang, P.Y. and Xi, X.G. (2017) Involvement of G-triplex and G-hairpin in the multi-pathway folding of human telomeric G-quadruplex. *Nucleic Acids Res.*, **45**, 11401–11412.
55. Hou, X.M., Wu, W.Q., Duan, X.L., Liu, N.N., Li, H.H., Fu, J., Dou, S.X., Li, M. and Xi, X.G. (2015) Molecular mechanism of G-quadruplex unwinding helicase: sequential and repetitive unfolding of G-quadruplex by Pif1 helicase. *Biochem. J.*, **466**, 189–199.
56. Amato, J., Pagano, A., Cosconati, S., Amendola, G., Fotticchia, I., Iaccarino, N., Marinello, J., De Magis, A., Capranico, G., Novellino, E. *et al.* (2017) Discovery of the first dual G-triplex/G-quadruplex stabilizing compound: a new opportunity in the targeting of G-rich DNA structures? *Biochim. Biophys. Acta*, **1861**, 1271–1280.
57. Bailey, L.J. and Doherty, A.J. (2017) Mitochondrial DNA replication: a PrimPol perspective. *Biochem. Soc. Trans.*, **45**, 513–529.
58. Minczuk, M., Piwowarski, J., Papworth, M.A., Awiszus, K., Schalinski, S., Dziembowski, A., Dmochowska, A., Bartnik, E., Tokatlidis, K., Stepien, P.P. *et al.* (2002) Localisation of the human hSuv3p helicase in the mitochondrial matrix and its preferential unwinding of dsDNA. *Nucleic Acids Res.*, **30**, 5074–5086.

59. Zhou, R., Zhang, J., Bochman, M.L., Zakian, V.A. and Ha, T. (2014) Periodic DNA patrolling underlies diverse functions of Pif1 on R-loops and G-rich DNA. *elife*, **3**, e02190.
60. Wu, C.G. and Spies, M. (2016) G-quadruplex recognition and remodeling by the FANCJ helicase. *Nucleic Acids Res.*, **44**, 8742–8753.
61. Ramanagoudr-Bhojappa, R., Blair, L.P., Tackett, A.J. and Raney, K.D. (2013) Physical and functional interaction between yeast Pif1 helicase and Rim1 single-stranded DNA binding protein. *Nucleic Acids Res.*, **41**, 1029–1046.
62. Ramanagoudr-Bhojappa, R., Byrd, A.K., Dahl, C. and Raney, K.D. (2014) Yeast Pif1 accelerates annealing of complementary DNA strands. *Biochemistry*, **53**, 7659–7669.
63. Dehghani-Tafti, S., Levdikov, V., Antson, A.A., Bax, B. and Sanders, C.M. (2019) Structural and functional analysis of the nucleotide and DNA binding activities of the human PIF1 helicase. *Nucleic Acids Res.*, **47**, 3208–3222.
64. Sanders, C.M. (2010) Human Pif1 helicase is a G-quadruplex DNA-binding protein with G-quadruplex DNA-unwinding activity. *Biochem. J.*, **430**, 119–128.
65. Rhodes, D. and Lipps, H.J. (2015) G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.*, **43**, 8627–8637.
66. Huppert, J.L. and Balasubramanian, S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.*, **35**, 406–413.
67. Moye, A.L., Porter, K.C., Cohen, S.B., Phan, T., Zyner, K.G., Sasaki, N., Lovrecz, G.O., Beck, J.L. and Bryan, T.M. (2015) Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. *Nat. Commun.*, **6**, 7643.
68. Smith, J.S., Chen, Q., Yatsunyk, L.A., Nicoludis, J.M., Garcia, M.S., Kranaster, R., Balasubramanian, S., Monchaud, D., Teulade-Fichou, M.P., Abramowitz, L. *et al.* (2011) Rudimentary G-quadruplex-based telomere capping in *Saccharomyces cerevisiae*. *Nat. Struct. Mol. Biol.*, **18**, 478–485.
69. Zhou, J., Liu, M., Fleming, A.M., Burrows, C.J. and Wallace, S.S. (2013) Neil3 and NEIL1 DNA glycosylases remove oxidative damages from quadruplex DNA and exhibit preferences for lesions in the telomeric sequence context. *J. Biol. Chem.*, **288**, 27263–27272.
70. Pilia, G., Chen, W.M., Scuteri, A., Orru, M., Albai, G., Dei, M., Lai, S., Usala, G., Lai, M., Loi, P. *et al.* (2006) Heritability of cardiovascular and personality traits in 6,148 Sardinians. *PLoS Genet.*, **2**, e132.
71. Sidore, C., Busonero, F., Maschio, A., Porcu, E., Naitza, S., Zoledziewska, M., Mulas, A., Pistis, G., Steri, M., Danjou, F. *et al.* (2015) Genome sequencing elucidates Sardinian genetic architecture and augments association analyses for lipid and blood inflammatory markers. *Nat. Genet.*, **47**, 1272–1281.
72. Ding, J., Sidore, C., Butler, T.J., Wing, M.K., Qian, Y., Meirelles, O., Busonero, F., Tsoi, L.C., Maschio, A., Angius, A. *et al.* (2015) Assessing mitochondrial DNA variation and copy number in lymphocytes of ~2,000 Sardinians using tailored sequencing analysis tools. *PLoS Genet.*, **11**, e1005306.
73. Ferrucci, L., Bandinelli, S., Benvenuti, E., Di Iorio, A., Macchi, C., Harris, T.B. and Guralnik, J.M. (2000) Subsystems contributing to the decline in ability to walk: bridging the gap between epidemiology and geriatric practice in the InCHIANTI study. *J. Am. Geriatr. Soc.*, **48**, 1618–1625.
74. Moore, A.Z., Ding, J., Tuke, M.A., Wood, A.R., Bandinelli, S., Frayling, T.M. and Ferrucci, L. (2018) Influence of cell distribution and diabetes status on the association between mitochondrial DNA copy number and aging phenotypes in the InCHIANTI study. *Aging Cell*, **17**, 1–3, e12683.
75. Khan, I., Crouch, J.D., Bharti, S.K., Sommers, J.A., Carney, S.M., Yakubovskaya, E., Garcia-Diaz, M., Trakselis, M.A. and Brosh, R.M., Jr. (2016) Biochemical characterization of the human mitochondrial replicative twinkle helicase: substrate specificity, DNA branch migration, and ability to overcome blockades to DNA unwinding. *J. Biol. Chem.*, **291**, 14324–14339.
76. Yakubovskaya, E., Guja, K.E., Eng, E.T., Choi, W.S., Mejia, E., Beglov, D., Lukin, M., Kozakov, D. and Garcia-Diaz, M. (2014) Organization of the human mitochondrial transcription initiation complex. *Nucleic Acids Res.*, **42**, 4100–4112.
77. Yakubovskaya, E., Chen, Z., Carrodeguas, J.A., Kisker, C. and Bogenhagen, D.F. (2006) Functional human mitochondrial DNA polymerase gamma forms a heterotrimer. *J. Biol. Chem.*, **281**, 374–382.
78. Wu, Y., Sommers, J.A., Loiland, J.A., Kitao, H., Kuper, J., Kisker, C. and Brosh, R.M., Jr. (2012) The Q motif of Fanconi anemia group J protein (FANCJ) DNA helicase regulates its dimerization, DNA binding, and DNA repair function. *J. Biol. Chem.*, **287**, 21699–21716.