Mammalian Polymerase Theta Promotes Alternative-NHEJ and Suppresses Recombination

Pedro A. Mateos-Gomez, Fade Gong, Nidhi Nair, Kyle M Miller, Eros Lazzerini-Denchi and Agnel Sfeir*

Supplementary information:

A) Sequence analysis of telomere fusions using illumina technology:

Alt-NHEJ junction sequences

TTAGGG	TTAA	CCCTAA	4
TTAGGG	A	CCCTAA	3
TTAGGG	TTAGGAA	CCCTAA	3
TTAGGG	TTCTAA	CCCTAA	3
TTAGGG	TTA	CCCTAA	2
TTAGGG		CCCTAA	2
TTAGGG	CCTAA	CCCTAA	2
TTAGGG	TTAG	CCCTAA	2
TTAGGG	TTAGCCTAA	CCCTAA	2
TTAGGG	TTAGGCTAA	CCCTAA	2
TTAGGG	TAA	CCCTAA	1
TTAGGG	TTAGG	CCCTAA	1
TTAGGG	AAC	CCCTAA	1
TTAGGG	CTAA	CCCTAA	2
TTAGGG	TA	CCCTAA	1
TTAGGG	TTAAA	CCCTAA	1
TTAGGG	TTAGGA	CCCTAA	1
TTAGGG	TTAGGTAA	CCCTAA	1
TTAGGG	TTATAA	CCCTAA	1
TTAGGG	TTCCCTAA	CCCTAA	1
TTAGGG	TTCCTAA	CCCTAA	1
TTAGGG	TTAAA	CCCTAA	1
TTAGGG	*AACTTGAGTCTTTAA*	CCCTAA	1
TTAGGG	*TTGGGTTAGAGAATTA*	CCCTAA	1
TTAGGG	*GAAAGGCCCTAACCCTAT*	CCCTAA	1
TTAGGG	*ATAGGGTTAGGACTAGCGTAAG*	CCCTAA	1
TTAGGG	*GCCTTCCTTTTCCATCCGGCGCCCAG*	CCCTAA	1
TTAGGG	*TTAACTCCCCTCCCCATGCCCTAACAA*	CCCTAA	1
TTAGGG	*TTAGGATAAGGGTAGTGGAACAGGTAGG*	CCCTAA	1
TTAGGG	*TTAGTGTTCTAACCAGAGCAATTAGGCAACAAA*	CCCTAA	1
		N=	46

C-NHEJ junction sequences

TTAGGG	TAA	CCCTAA	15
TTAGGG	TTA	CCCTAA	12
TTAGGG		CCCTAA	7
TTAGGG	AA	CCCTAA	6
TTAGGG	CTAA	CCCTAA	3
TTAGGG	TTAA	CCCTAA	4
TTAGGG	TTAGCCTAA	CCCTAA	3
TTAGGG	TTATAA	CCCTAA	3
TTAGGG	TTAGG	CCCTAA	4
TTAGGG	A	CCCTAA	2
TTAGGG	T	CCCTAA	2
TTAGGG	TTAG	CCCTAA	2
TTAGGG	TTAGGCCTAA	CCCTAA	5
TTAGGG	TTAGGTAA	CCCTAA	3
TTAGGG	TTCTAA	CCCTAA	3
TTAGGG	TCTAA	CCCTAA	2
TTAGGG	TTAGAA	CCCTAA	2
TTAGGG	TTAGA	CCCTAA	2
TTAGGG	TTAGGA	CCCTAA	2
TTAGGG	TTAGGCTAA	CCCTAA	1
TTAGGG	TTATCCTAA	CCCTAA	2
TTAGGG	TTAGGGCTAA	CCCTAA	2
TTAGGG	TTAGGGTTAGGA	CCCTAA	1
TTAGGG	CCTAA	CCCTAA	3
		N=	91

Sequences of fusion junctions obtained at telomeres that have been processed by alt-NHEJ and C-NHEJ. N represents the total number of fusion events: N=91 for C-NHEJ and N=46 for Alt-NHEJ. Underlined sequences in italics represent non-TTAGGG nucleotide insertions found in the context of alt-NHEJ driven fusions. No fusion events were recovered from non-Cre treated *TRF1*$^{F/F}$*TRF2*$^{F/F}$*Ku80*$^{-/-}$ cells (data not shown).

In order to calculate the expected efficiency of detecting telomere fusions based on next generation sequencing, one has to take into account the average telomere lengths (~75Kb in mice) as well as the length of the short reads obtained by illumina platforms (150 nucleotides in our experimental setting). Therefore, if at the time of analysis all telomeres were fused we would expect to identify one read containing a fusion event for every 1000 telomere reads. Our results indicate that in TRF2 null cells, the efficiency of retrieving fusion events is approximately 10

times lower. This could be attributed to the fact that not all telomeres at this time point following Cre-induction are engaged in fusions, and could also be linked to the strict requirement to identify telomere fusions by short reads. On one hand, not all telomeres are engaged in telomere fusions when analyzed on metaphase spreads. Additionally, it is likely that the actual percentage of fused telomeres in a population of cells is lower than what is observed in metaphase spreads, especially given that cells with uncapped telomeres that have not been fused are likely to interfere with cell cycle progression. With respect to the strict sequence requirement when filtering reads, we would like to point out that our criteria to detect fusion sequences imposes that the junction is in the middle portion of the short read, which diminishes the probability of detecting reads corresponding to telomere fusions. Lastly, we anticipate that a significant fraction of the insertion-type fusions generated by PolQ in shelterin-free Ku80 deficient cells to be lost in our analysis.

B) Figure 4e – Model: Resection of a DSB by MRN/CtIP generates short single stranded overhangs, which can undergo spontaneous annealing. Both intermediate structures highlighted in grey boxes are suitable substrates for Polθ. Template-independent incorporation of random nucleotides at overhangs generates more homology, which would facilitate annealing of opposite DNA ends. On the other hand, template-dependent Polθ activity stabilizes paired intermediates, including ones that have undergone spontaneous annealing. The latter activity will not manifest in random insertions, but is equally important to drive alt-NHEJ repair and inhibit HDR. Cells lacking *PolQ* (i) display elevated levels of HDR. On the other hand, cells lacking *PolQ* and *BRCA* (ii) accumulate excess unrepaired lesions/chromosomal aberrancies, which would compromise survival.