



Specific Patterns of Mutations Induced by Mutated DNA Polymerase Epsilon in Endometrial Cancer Genomes

Daria Ostroverkhova^{1*}, Kathrin Tyryshkin^{1*}, Jessy Jia Song², Igor Rogozin³, Konstantin Shaitan⁴, Polina Shcherbakova⁵ and Anna Panchenko¹

¹Queen's University, Canada; ²University of Waterloo, Canada; ³National Center for Biotechnology Information, USA; ⁴Lomonosov Moscow State University, Russia; ⁵University of Nebraska Medical Center, USA

INTRODUCTION

- DNA Polymerase Epsilon (Pol ε), encoded by the *POLE* gene, plays an essential role in maintaining DNA replication fidelity;
- POLE* mutations affecting functionally critical residues in Pol ε lead to an exceptionally high mutation burden in endometrial cancer;
- Here we investigated which features are associated with the presence/absence of *POLE* mutations beyond the trinucleotide limit;
- We applied a novel computational framework to the whole-exome sequencing data of endometrial tumors to analyse motifs for G>T mutations in various positions in the DNA sequence contexts of different lengths.

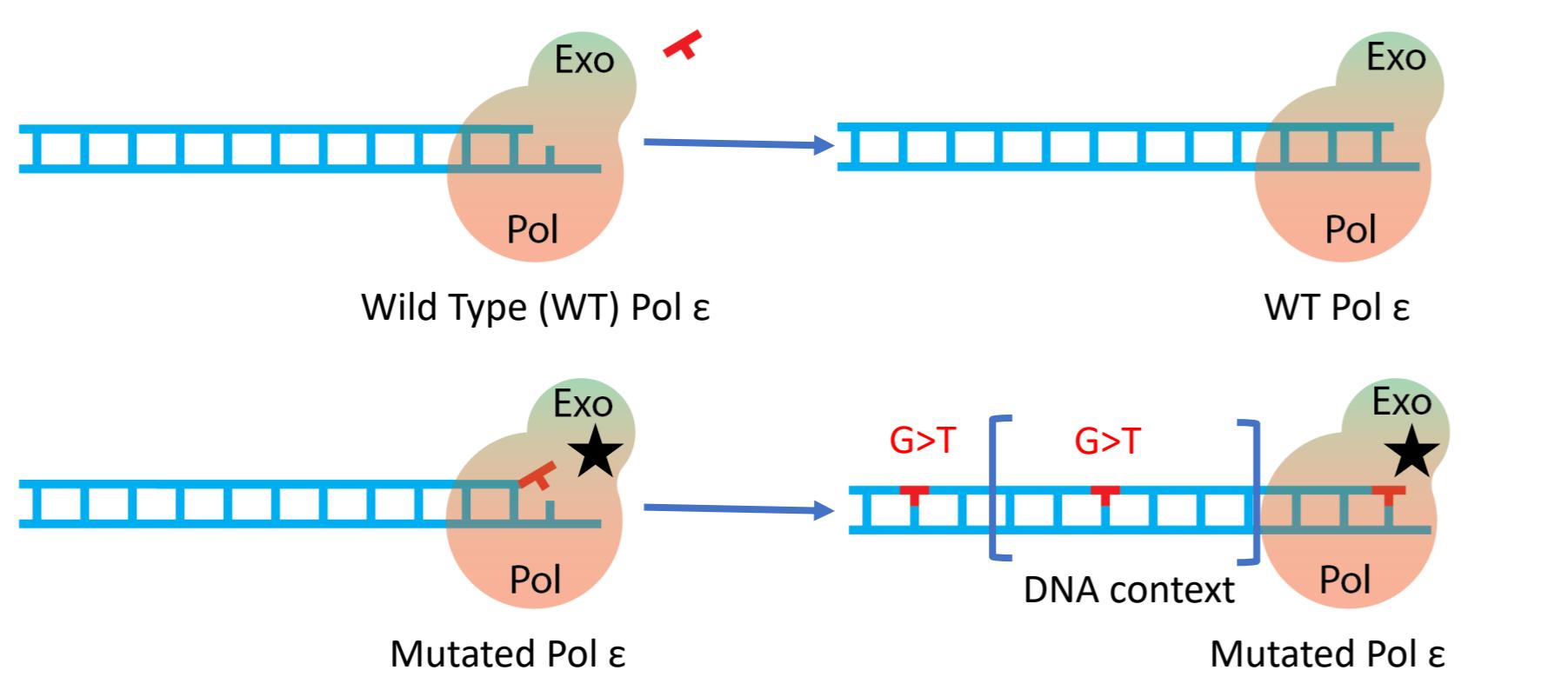
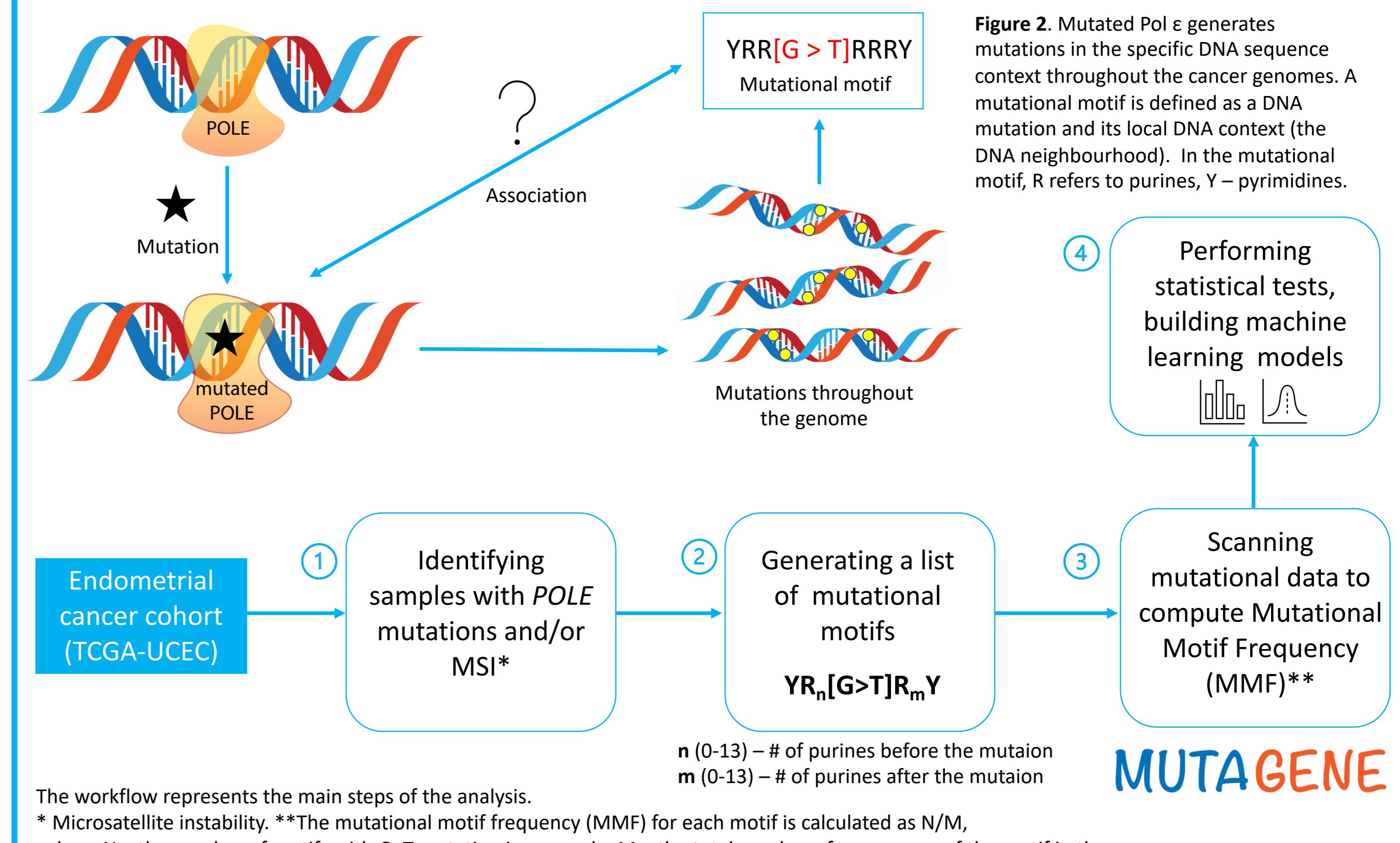


Figure 1. Mutations in Pol ε impair its functioning and lead to an increased number of G > T transversions in the cancer genome. ★ mutation

METHODS



The workflow represents the main steps of the analysis.

* Microsatellite instability. **The mutational motif frequency (MMF) for each motif is calculated as N/M, where N – the number of motifs with G>T mutation in a sample, M – the total number of occurrences of the motif in the exome.

MUTAGENE

RESULTS

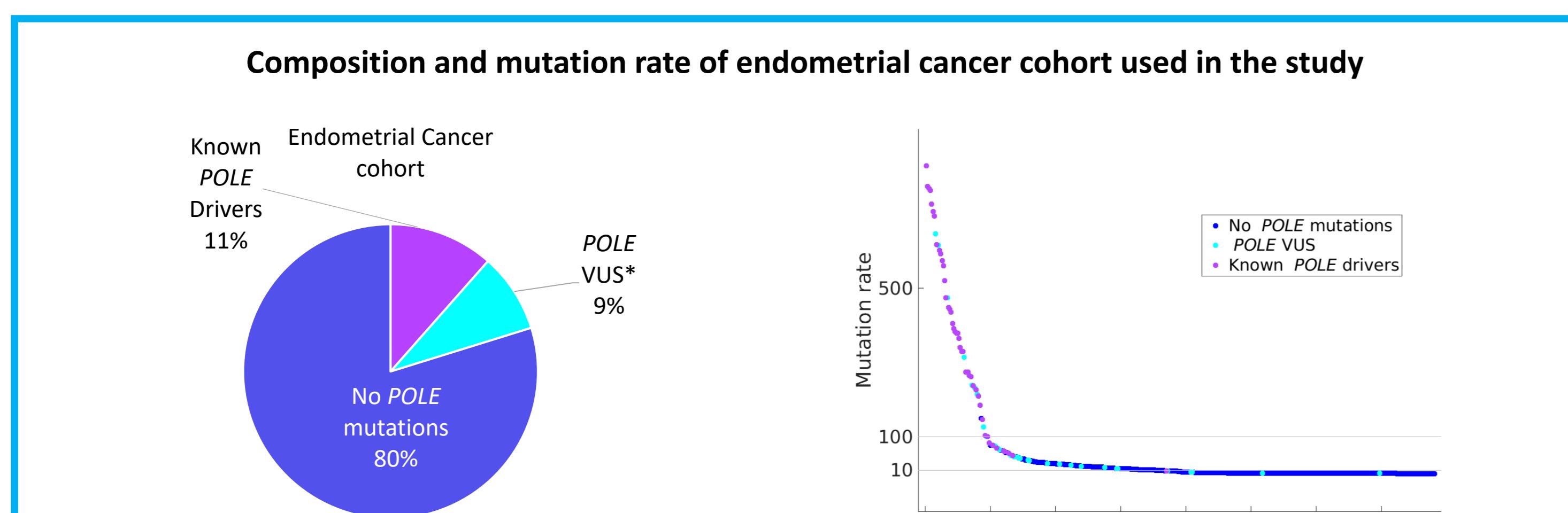


Figure 3. The number of samples with known *POLE* driver mutations (N = 45), samples with *POLE* variant of unknown significance (*VUS) (N=34), and samples with no *POLE* mutations (N=312).

I. *POLE* driver status is associated with increased frequency of mutations in polypurine tracts with a strong preference for certain positions within the tracts



Figure 5. Mutational Motif Frequency (MMF) of mutational motifs in samples with *POLE* drivers, *POLE* VUS, and no mutations in *POLE*. R - purines, Y - pyrimidines.

Figure 6. Comparing the distribution of MMF among the group of samples with *POLE* driver mutations. Each box plot represents the MMF distribution for motifs where the number of purines after the mutation is varied. The highest MMF is observed for mutational motifs where the mutation is followed by two or three purines.

II. The polypurine tract signature is associated with the presence of *POLE* driver mutations and not MMR deficiency

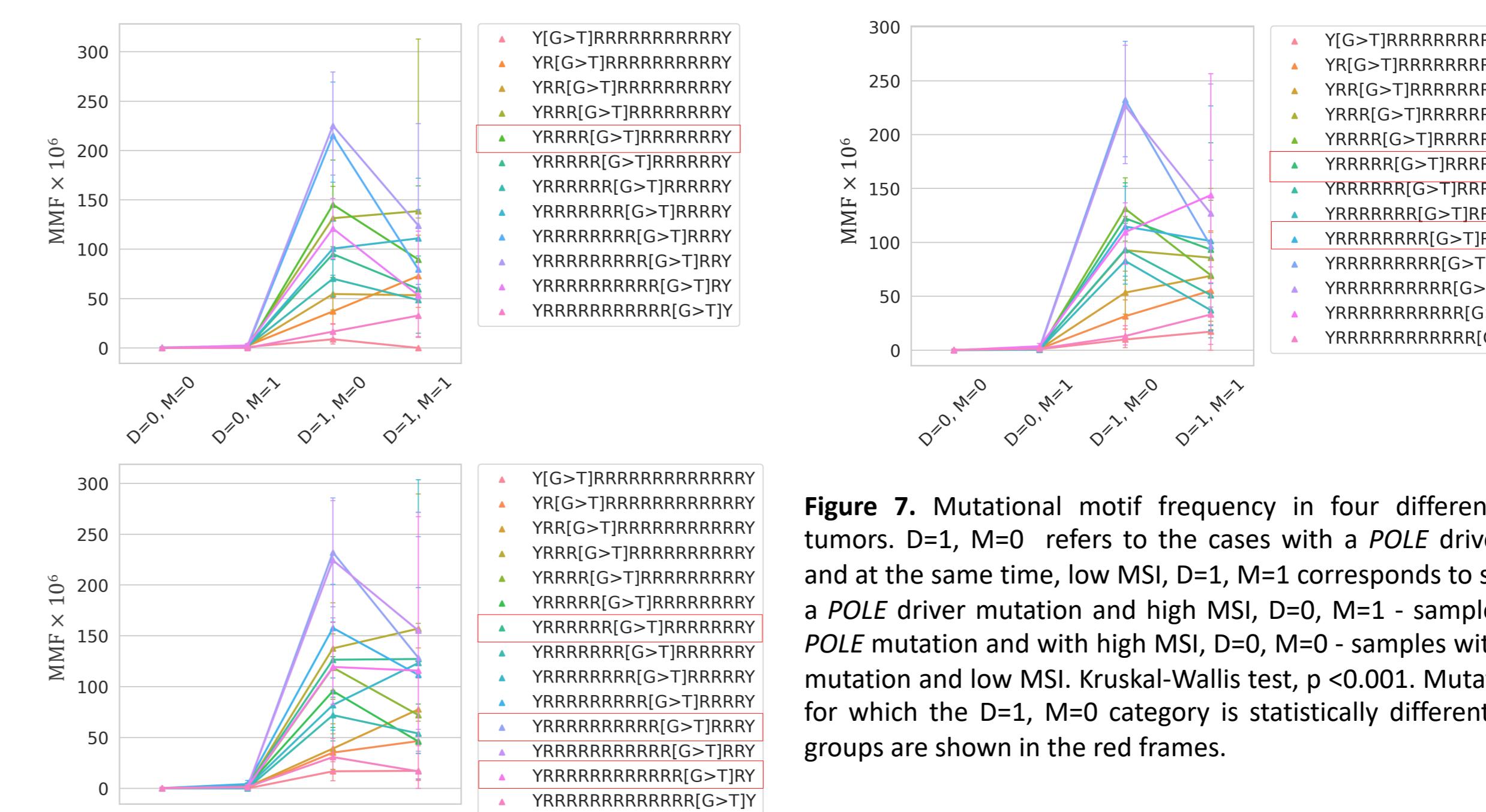


Figure 7. Mutational motif frequency in four different groups of tumors. D=1, M=0 refers to the cases with a *POLE* driver mutation, and at the same time, low MSI, D=1, M=1 corresponds to samples with a *POLE* driver mutation and high MSI, D=0, M=0 - samples without a *POLE* mutation and with low MSI, D=0, M=1 - samples with a *POLE* mutation and with high MSI. Kruskal-Wallis test, p <0.001. Mutational motifs for which the D=1, M=0 category is statistically different from other groups are shown in the red frames.

III. Machine learning approach predicts samples with novel *POLE* driver mutations

Table 1. Comparing of machine learning models' performance.

Predictors	Evaluation Metrics	
	AUC-ROC	Matthew's correlation
Mutational motifs $YR_n[G>T]R_mY$	0.98	0.97
Mutation Burden	0.97	0.91
G>T (without context)	0.81	0.69

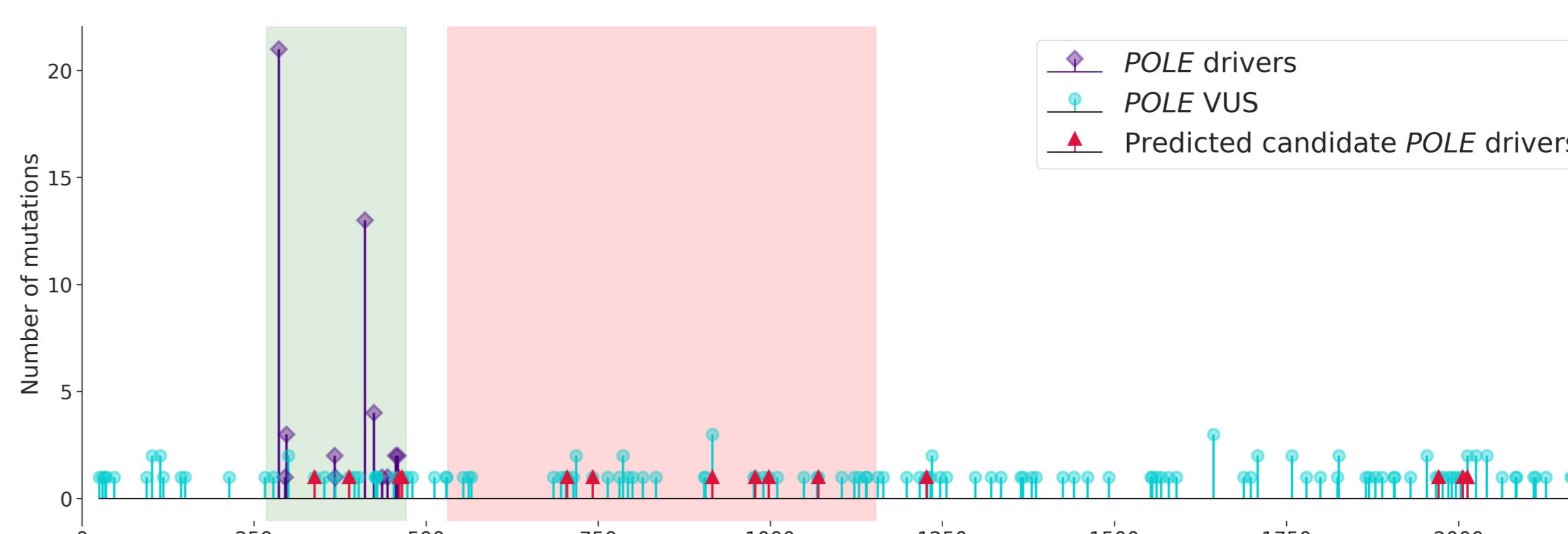


Figure 8. A schematic representation of *POLE*. Each lollipop shows a *POLE* mutation reported in endometrial cancer. Red lollipops display candidate *POLE* driver mutations predicted by our approach.

CONCLUSIONS

- Presence of driver mutations in *POLE* is associated with the preferential occurrence of G>T mutations in polypurine tracts throughout the endometrial cancer genomes;
- There is the strong association between the presence of *POLE* driver mutations and MMR proficiency;
- Samples with potential/candidate *POLE* driver mutations were identified by applying machine learning approach.