Frequent POLE-Driven Hypermutation in Ovarian Endometrioid Cancer Revealed by Mutational Signatures in RNA Sequencing

Jaime Davila (davila3@stolaf.edu)
St. Olaf College
Pritha Chanana
Fred Hutchinson Cancer Research Center
Vivekananda Sarangi
Mayo Clinic
Zach Fogarty
Mayo Clinic
John Weroha
Mayo Clinic
Ruifeng Guo
Mayo Clinic
Goode Ellen
Mayo Clinic
Yajue Huang
Mayo Clinic
Wang Chen
Mayo Clinic

Research Article

Keywords: hypermutation, mutational signatures, ovarian endometrioid cancer, POLE, RNA-seq

Posted Date: January 18th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-145368/v1

Abstract

Background: DNA polymerase epsilon (POLE) is encoded by the POLE gene, and *POLE*-driven tumors are characterized by high mutational rates. *POLE*-driven tumors are relatively common in endometrial and colorectal cancer, and their presence is increasingly recognized in ovarian cancer (OC) of endometrioid type. *POLE*-driven cases possess an abundance of TCT>TAT and TCG>TTG somatic mutations characterized by mutational signature 10 from the Catalog of Somatic Mutations in Cancer (COSMIC). By quantifying the contribution of COSMIC mutational signature 10 in RNA sequencing (RNA-seq) we set out to identify *POLE*-driven tumors in a set of unselected Mayo Clinic OC.

Methods: Mutational profiles were calculated using expressed single-nucleotide variants (eSNV) in the Mayo Clinic OC tumors (n=195), The Cancer Genome Atlas (TCGA) OC tumors (n=419), and the Genotype-Tissue Expression (GTEx) normal ovarian tissues (n=84). Non-negative Matrix Factorization (NMF) of the mutational profiles inferred the contribution per sample of four distinct mutational signatures, one of which corresponds to COSMIC mutational signature 10.

Results: In the Mayo Clinic OC cohort we identified six tumors with a predicted contribution from COSMIC mutational signature 10 of over five mutations per megabase. These six cases harbored known POLE hotspot mutations (*P286R*, *S297F*, *V411L*, and *A456P*) and were of endometrioid histotype (P=5e-04). These six tumors were hypermutated with a higher tumor mutation load (mean, 54.02 mutations per megabase) compared to non-POLE endometrioid OC cases (mean, 7.69 mutations per megabase; P=5e-04), and had an early onset (average age of patients at onset, 48.33 years) when compared to non-POLE endometrioid OC cohort (average age at onset, 60.13 years; P=.008). Samples from TCGA and GTEx had a low COSMIC signature 10 contribution (median 0.16 mutations per megabase; maximum 1.78 mutations per megabase) and carried no POLE hotspot mutations.

Conclusions: From the largest cohort of RNA-seq from endometrioid OC to date (n=53), we identified six hypermutated samples likely driven by POLE (frequency, 11%). Our result suggests the clinical need to screen for POLE driver mutations in endometrioid OC, which can guide enrollment in immunotherapy clinical trials.

Background

DNA polymerase epsilon (POLE) is a key member of the DNA polymerase family and is responsible for error correction during replication (1). Somatic mutations in the exonuclease domain of *POLE* have been identified in 7–12% of uterine endometrial cancers, in 1–2% of colorectal cancers, and at low frequencies in stomach cancer, glioblastoma, breast cancer, and others (2–5). Recurrent hotspot mutations include *P286R*, *S297F*, and *V411L*, which are characterized by ultramutation (5, 6). Patients with *POLE*-driven endometrial cancer have a favorable prognosis, a higher neoantigen load, an increased number of tumor-infiltrating lymphocytes, and they may benefit from immunotherapy (6–8).

Somatic *POLE* mutations are uncommon in serous ovarian cancer (OC) (9, 10) but the presence of *POLE* mutations is increasingly recognized in ovarian endometrioid cancer (OEC) (10–14). Patients with *POLE*-driven OEC have earlier disease onset, and an increased number of *CD8 +* intraepithelial tumor-infiltrating lymphocytes (11, 14). The prevalence of *POLE*-driven tumors in different OEC studies ranges from 3–13% (10–12, 14–16), and is likely driven by sample size, cohort selection criteria, and type of mutation detection assay. Interestingly, in a Japanese cohort of concurrent ovarian and endometrial cancer, the frequency of *POLE* mutations was high (five of 8 cases; 62%) detected by Sanger sequencing (13); however, the reported *POLE* mutations (*Q292E*, *E396V*, *D287N*, and *N293D*) do not correspond to known hotspot mutations. This result highlights the need to ascertain the effects of particular *POLE* mutations and whether they are accompanied by hypermutation.

POLE-driven tumors are associated with a distinct mutational signature found in whole-genome sequencing from tumor-normal pairs which is characterized by a high number of TCT > TAT and TCG > TTG mutations (4, 17). Such distinct mutational profile is known as COSMIC mutational signature 10 (17). Exome sequencing detected an abundance of COSMIC mutational signature 10 in OEC cases with known *POLE* hotspot mutations (16). Previously, we developed a novel method for inferring and quantifying distinct mutational signatures using RNA sequencing (RNA-seq) data (18, 19). We applied this method in tumor-only fresh-frozen RNA sequencing (RNA-seq) samples in endometrial cancer and colorectal cancer tumors from The Cancer Genome Atlas (TCGA) and identified *POLE* cases with high specificity and sensitivity (19). Additionally, our method can be used to calculate the number of somatic mutations per megabase also known as tumor mutational burden (TMB).

Using previously published RNA-seq from a Mayo Clinic OC cohort (n = 195), which included the largest set of OEC transcriptomes to date (n = 53), we sought to leverage the mutational signatures approach to identify *POLE*-driven cases and to characterize their clinical characteristics (20, 21).

Methods

Mayo Clinic and Public Data Selection and RNA-seq

Selected participants were patients who were at least 20 years old and were ascertained at Mayo Clinic from 1992 through 2009 within one year after receiving a pathologically confirmed diagnosis of primary invasive epithelial OC, fallopian tube cancer, or primary peritoneal cancer. (Table 1). Tumors were snap frozen immediately after surgery and stored at − 80 ºC. A gynecologic pathologist confirmed the clinical diagnoses and verified the tumor histology and grade and the presence of 70% tumor content before RNA extraction from fresh frozen tissue. As described previously (20, 21) transcriptomic sequencing was performed in four batches with TruSeq Library Preparation kits (Stranded Total RNA Library Preparation Kit or RNA Library Preparation Kit v2; Illumina, Inc) and sequenced on the Illumina HiSeq 2000 sequencer with 100–base pair paired-end reads. All patients gave informed consent; all protocols were approved by the Mayo Clinic Institutional Review Board.

Table 1
Characteristics of the 195 Patients in the Mayo Clinic Ovarian Cancer Cohort

Feature	Value
Age at diagnosis, mean (SD), y	62 (12)
Histology, No. (%)	
Serous	114 (58)
Endometrioid	53 (27)
Clear cell	14 (7)
Undifferentiated	8 (4)
Mucinous	3 (2)
Other	3 (2)
Grade, No. (%)	
1	20 (10)
2	24 (12)
3	146 (75)
Unknown	5 (3)
Stage, No. (%)	
I	45 (23)
II	15 (8)
III	100 (51)
IV	33 (17)
Unknown	2 (1)

We also used OC RNA-seq from TCGA (n = 419) and from normal ovarian tissue from the Genotype-Tissue Expression (GTEx) project (Broad Institute) (n = 84) (9, 22).

Bioinformatics Methods

RNA-seq of Mayo Clinic, TCGA, and GTEx data sequencing reads were processed through the Mayo Clinic MAP-RSeq v.2.1.5 computational workflow, and variants were calculated with RVboost 0.1 (18, 23). We considered expressed single nucleotide variants from RVboost with a Q score greater than 10%, read depth greater than 10, a minor allelic frequency less than 2% in the 1000 Genomes Project (24), and not present in recurrent expressed single-nucleotide variants identified in RNA-seq from adjacent normal

tissue. The RNA-seq capture region was defined as positions with 20× coverage as calculated by the Genome Analysis Toolkit (GATK; Broad Institute). Samples with read depth over 20× at less than five million positions were excluded from this analysis. Tumor mutation burden (TMB) was calculated as the number of considered expressed variants per capture region $\times 10^6$.

We used the mutational signatures v2 from the Catalogue of Somatic Mutations in Cancer (COSMIC) (25) where the mutational profiles are represented as the proportion of each substitution type (C > A, C > G, C > T, T > A, T > C, and T > G) and its dinucleotide context (the nucleotide before and after each mutated base) (17). For the detection of mutational signatures and their contribution to each sample, we used R version 3.4.2 (R Foundation) with the MutationalPatterns v1.4.3 package (26). To measure the similarity between two mutational signatures we used the cosine similarity as implemented in the function *cos_sim* in the MutationalPatterns v1.4.3 package. The cosine similarity takes values between 0 and 1, with a value close to 1 if there is great similarity between signatures and close to 0 if the two signatures are dissimilar.

Statistical Analysis

We used R version 3.4.2 with the tidyverse 1.2.1 package to perform statistical analyses and generate graphs. Two-sided Mann-Whitney tests (Wilcoxon rank sum tests) were used for comparisons of TMB and age across patients with and without *POLE* mutations. Histology findings from patients with and without *POLE* mutations were compared with the Fisher exact test.

Results

The Mayo Clinic OC cohort with existing RNA-seq consisted of 195 patients whose clinical characteristics are described in Table 1 (20, 21). This cohort, contains an abundance of nonserous histologies (81 of 195) and includes 53 with OEC, is the largest OEC RNAseq collection to date.

Mutational profiles from eSNV were calculated in the Mayo Clinic OC tumors (n = 195), TCGA OC tumors (n = 419), and the GTEx normal ovarian tissues (n = 84). Using Non-negative Matrix Factorization (NMF) (26) we were able to approximate each sample's mutational profile as a combination of four distinct mutational signatures (Fig. 1A and Supplemental Fig. 1). The cosine similarity measures the proximity between two mutational profiles and takes values between 0 and 1, with values closer to 1 corresponding to highly similar mutational profiles. Using the cosine similarity, we established a high degree of resemblance between the approximated mutational profiles from NMF and the original mutational profile (median cosine similarity, 0.92; interquartile range, 0.89–0.93) (Supplemental Fig. 2).

Of the four mutational signatures identified, three are found within the COSMIC mutational signatures catalog v2 (Fig. 1B) (17). With a cosine similarity of 0.88, the first signature corresponded to COSMIC signature 10, which is associated with *POLE* defects (2). With a cosine similarity of 0.85, a second signature coincides with COSMIC signature 6, which is associated with microsatellite instability and defects of MMR (2). With a cosine similarity of 0.72, a third signature is equivalent to COSMIC signature

5, which is related to the patient's age (27). Finally with a cosine similarity of 0.68, a fourth signature corresponded to a mutational signature found in germline variants (28).

Normal samples from GTEx and serous OC samples from TCGA had low levels of the *POLE* signature (median 0.16 mutations per megabase; maximum 1.78 mutations per megabase) and can be used as negative *POLE* controls (Fig. 2A). Six samples (of 195) from our Mayo OC cohort had a *POLE* signature contribution of more than five mutations per megabase and were clearly outliers when compared to the negative *POLE* controls (Fig. 2A). The individual mutational profiles for each of those six samples are shown in Fig. 2B.

Within these six samples, we found expressed mutations corresponding to the *POLE* hotspot COSMIC mutations *P286R*, *S297F*, *V411L*, and *A456P* (Table 2 and on Supplemental Fig. 3) (29). No other samples in the Mayo OC, the TCGA OC, or the normal ovarian GTEx cohort harbored expressed mutations in any of the *POLE* hotspots.

Table 2
Clinical Characteristics of six *POLE* Samples from the Mayo Clinic Ovarian Cancer Cohort

Sample	*POLE* mutation	TMB	Age, y	Histology	Stage	Grade	Vital status
P1	*A456P*	76.13	46	Endometrioid	III	2	Alive
P2	*S297F*	71.75	49	Endometrioid	I	3	Alive
P3	*V411L*	115.01	49	Endometrioid	III	3	Deceased
P4	*P286R*	28.91	46	Endometrioid	I	2	Alive
P5	*P286R*	21.76	49	Endometrioid	I	1	Alive
P6	*V411L*	10.59	51	Endometrioid	I	1	Alive

Abbreviations: *POLE*, DNA polymerase epsilon; TMB, tumor mutation burden.

Of particular interest, Mayo OEC Sample P3 had a high *POLE* contribution of 64.2 mutations per megabase as well as a high MMR signature contribution of 50.8 mutations per megabase. On further inspection, this sample harbored a pathogenic *R167C* mutation in *MLH1*. Samples with defects in both *POLE* and MMR have been previously described in OEC (14).

All six samples were of endometrioid histotype (P = 5e-04) and constituted 11% of OEC cases. Average age at onset was earlier (48.33 years) than in the non-*POLE* OEC cohort (60.13 years; P = .008). Four of these six samples corresponded to stage 1, and two corresponded to stage 3.

By excluding the contribution from the germline signature found in the NMF analysis, we calculated the tumor mutational burden (TMB). Samples from the three cohorts had a median of nearly five mutations per megabase (Supplemental Fig. 4). The six OEC tumors with *POLE* signature from Mayo Clinic had a mean mutational burden of 54.02 mutations per megabase and was greater than the non-*POLE* OEC cases (mean, 7.69 mutations per megabase; P = 3e-04) (Table 2).

Discussion

Although RNA-seq was developed for and used in the context of gene expression quantification, it is a well-suited technique for mutational signature analysis. It allows for the quantification of tumor mutational burden and the identification of causative *POLE* mutations in a single assay as demonstrated in our study.

POLE-driven tumors have a favorable prognosis, an increased number of tumor-infiltrating lymphocytes, and can benefit from immunotherapy. Despite these distinct characteristics of *POLE*-driven tumors, *POLE* mutation status is not routinely evaluated in a clinical settting for OEC. By using RNA-seq on the largest OEC cohort to our knowledge (n = 53), we found six of 53 OEC samples (11%) that were *POLE*-driven. Those six samples had high mutational burden, highly specific mutational profile corresponding to the well-characterized *POLE* COSMIC signature 10 profile, and *POLE* hotspot mutations (*P286R*, *S297F*, *V411L*, and *A456P*). Two of the 6 patients in the group of *POLE* cases had advanced-stage OEC with relatively high recurrence risks; therefore, OEC patients with *POLE*-driven tumors can be eligible for immunotherapy trials if their cancers progress or recur.

The tumor mutational burden of our *POLE* cases spans from ten mutations per megabase to more than 115 mutations per megabase. Patients with higher TMB have worse clinical characteristics and outcomes than among non-POLE OECs; the sample with the lowest TMB was from a patient with a low-grade, stage 1 cancer, and the sample with the highest TMB was from a deceased patient with a high-grade, stage 3 cancer. However, the limited number of *POLE* cases in this study precluded any statistical analysis, and further independent studies with larger sample sizes are necessary to validate and confirm such a trend.

Conclusions

Using RNA-seq mutational signatures from the largest OEC cohort to date (n = 53), we found that 6 hypermutated samples (11%) had evidence of *POLE*-driven tumors.

Abbreviations

COSMIC, Catalogue of Somatic Mutations in Cancer

GTEx, Genotype-Tissue Expression

MMR, mismatch repair

MSK-IMPACT, Memorial Sloan Kettering integrated mutation profiling of actionable cancer targets

OC, ovarian cancer

OEC, ovarian endometrioid cancer

POLE, DNA polymerase epsilon

RNA-seq, RNA sequencing

TCGA, The Cancer Genome Atlas

TMB, tumor mutation burden

Declarations

Ethics Approval and Consent to Participate

All patients gave informed consent, and all protocols were approved by the Mayo Clinic Institutional Review Board. All methods were carried out in accordance with relevant guidelines and regulations.

Consent for Publication

Consent for publication is not applicable.

Availability of Data and Materials

The datasets and code generated during and/or analyzed during the current study are available in https://github.com/jdavilal/mutational_sig_oec

Competing Interests

The authors declare that they have no competing interests.

Funding

This study was supported by R21-CA222867, R01-CA122443, R01-CA248288 from the National Institutes of Health

Authors' Contributions

J.I.D. and C.W. performed and supervised bioinformatics analyses, participated in study design and interpretation of results, and manuscript drafting. P.C., V.S., and Z.C.F. participated in generation of results and performed bioinformatics analysis and manuscript drafting. E.L.G., S. J.W., and Y.H. participated in study design and manuscript drafting. All authors read and approved the final manuscript.

Acknowledgments

We wish you to acknowledge English Language editing and review services by Emelina Bly at the department of Scientific Publications at Mayo Clinic, Rochester, MN

References

1. Pursell ZF, Isoz I, Lundström E-B, Johansson E, Kunkel TA. Yeast DNA polymerase epsilon participates in leading-strand DNA replication. Science. 2007;317(5834):127-30.
2. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012;487(7407):330-7.
3. Kandoth C, Schultz N, Cherniack AD, Akbani R, Liu Y, Shen H, et al. Integrated genomic characterization of endometrial carcinoma. Nature. 2013;497(7447):67-73.
4. Shinbrot E, Henninger EE, Weinhold N, Covington KR, Göksenin AY, Schultz N, et al. Exonuclease mutations in DNA polymerase epsilon reveal replication strand specific mutation patterns and human origins of replication. Genome Res. 2014;24(11):1740-50.
5. Rayner E, van Gool IC, Palles C, Kearsey SE, Bosse T, Tomlinson I, et al. A panoply of errors: polymerase proofreading domain mutations in cancer. Nat Rev Cancer. 2016;16(2):71-81.
6. Church DN, Stelloo E, Nout RA, Valtcheva N, Depreeuw J, ter Haar N, et al. Prognostic significance of POLE proofreading mutations in endometrial cancer. J Natl Cancer Inst. 2015;107(1):402.
7. Howitt BE, Shukla SA, Sholl LM, Ritterhouse LL, Watkins JC, Rodig S, et al. Association of Polymerase e-Mutated and Microsatellite-Instable Endometrial Cancers With Neoantigen Load, Number of Tumor-Infiltrating Lymphocytes, and Expression of PD-1 and PD-L1. JAMA Oncol. 2015;1(9):1319-23.
8. Mehnert JM, Panda A, Zhong H, Hirshfield K, Damare S, Lane K, et al. Immune activation and response to pembrolizumab in POLE-mutant endometrial cancer. J Clin Invest. 2016;126(6):2334-40.
9. Integrated genomic analyses of ovarian carcinoma. Nature. 2011;474(7353):609-15.
10. Zou Y, Liu FY, Liu H, Wang F, Li W, Huang MZ, et al. Frequent POLE1 p.S297F mutation in Chinese patients with ovarian endometrioid carcinoma. Mutat Res. 2014;761:49-52.
11. Hoang LN, McConechy MK, Köbel M, Anglesio M, Senz J, Maassen M, et al. Polymerase Epsilon Exonuclease Domain Mutations in Ovarian Endometrioid Carcinoma. Int J Gynecol Cancer. 2015;25(7):1187-93.
12. Cybulska P, Paula ADC, Tseng J, Leitao MM, Jr., Bashashati A, Huntsman DG, et al. Molecular profiling and molecular classification of endometrioid ovarian carcinomas. Gynecol Oncol. 2019;154(3):516-23.
13. Ishikawa M, Nakayama K, Nakamura K, Ono R, Yamashita H, Ishibashi T, et al. High frequency of POLE mutations in synchronous endometrial and ovarian carcinoma. Hum Pathol. 2019;85:92-100.

14. Leskela S, Romero I, Rosa-Rosa JM, Caniego-Casas T, Cristobal E, Pérez-Mies B, et al. Molecular Heterogeneity of Endometrioid Ovarian Carcinoma: An Analysis of 166 Cases Using the Endometrial Cancer Subrogate Molecular Classification. Am J Surg Pathol. 2020;44(7):982-90.
15. Parra-Herran C, Lerner-Ellis J, Xu B, Khalouei S, Bassiouny D, Cesari M, et al. Molecular-based classification algorithm for endometrial carcinoma categorizes ovarian endometrioid carcinoma into prognostically significant groups. Mod Pathol. 2017;30(12):1748-59.
16. Teer JK, Yoder S, Gjyshi A, Nicosia SV, Zhang C, Monteiro ANA. Mutational heterogeneity in non-serous ovarian cancers. Sci Rep. 2017;7(1):9728.
17. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013;500(7463):415-21.
18. Wang C, Davila JI, Baheti S, Bhagwate AV, Wang X, Kocher JP, et al. RVboost: RNA-seq variants prioritization using a boosting method. Bioinformatics. 2014;30(23):3414-6.
19. Jessen E, Liu Y, Davila J, Kocher J-P, Wang C. Determining mutational burden and signature using RNA-seq from tumor-only samples2020.
20. Earp MA, Raghavan R, Li Q, Dai J, Winham SJ, Cunningham JM, et al. Characterization of fusion genes in common and rare epithelial ovarian cancer histologic subtypes. Oncotarget. 2017;8(29):46891-9.
21. Fridley BL, Dai J, Raghavan R, Li Q, Winham SJ, Hou X, et al. Transcriptomic Characterization of Endometrioid, Clear Cell, and High-Grade Serous Epithelial Ovarian Carcinoma. Cancer Epidemiology Biomarkers & Prevention. 2018;27(9):1101.
22. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013;45(6):580-5.
23. Kalari KR, Nair AA, Bhavsar JD, O'Brien DR, Davila JI, Bockol MA, et al. MAP-RSeq: Mayo Analysis Pipeline for RNA sequencing. BMC Bioinformatics. 2014;15:224.
24. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, et al. A global reference for human genetic variation. Nature. 2015;526(7571):68-74.
25. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res. 2019;47(D1):D941-d7.
26. Blokzijl F, Janssen R, van Boxtel R, Cuppen E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. Genome Med. 2018;10(1):33.
27. Alexandrov LB, Jones PH, Wedge DC, Sale JE, Campbell PJ, Nik-Zainal S, et al. Clock-like mutational processes in human somatic cells. Nature genetics. 2015;47(12):1402-7.
28. Milholland B, Dong X, Zhang L, Hao X, Suh Y, Vijg J. Differences between germline and somatic mutation rates in humans and mice. Nat Commun. 2017;8:15183.
29. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res. 2018;47(D1):D941-D7.

Figures

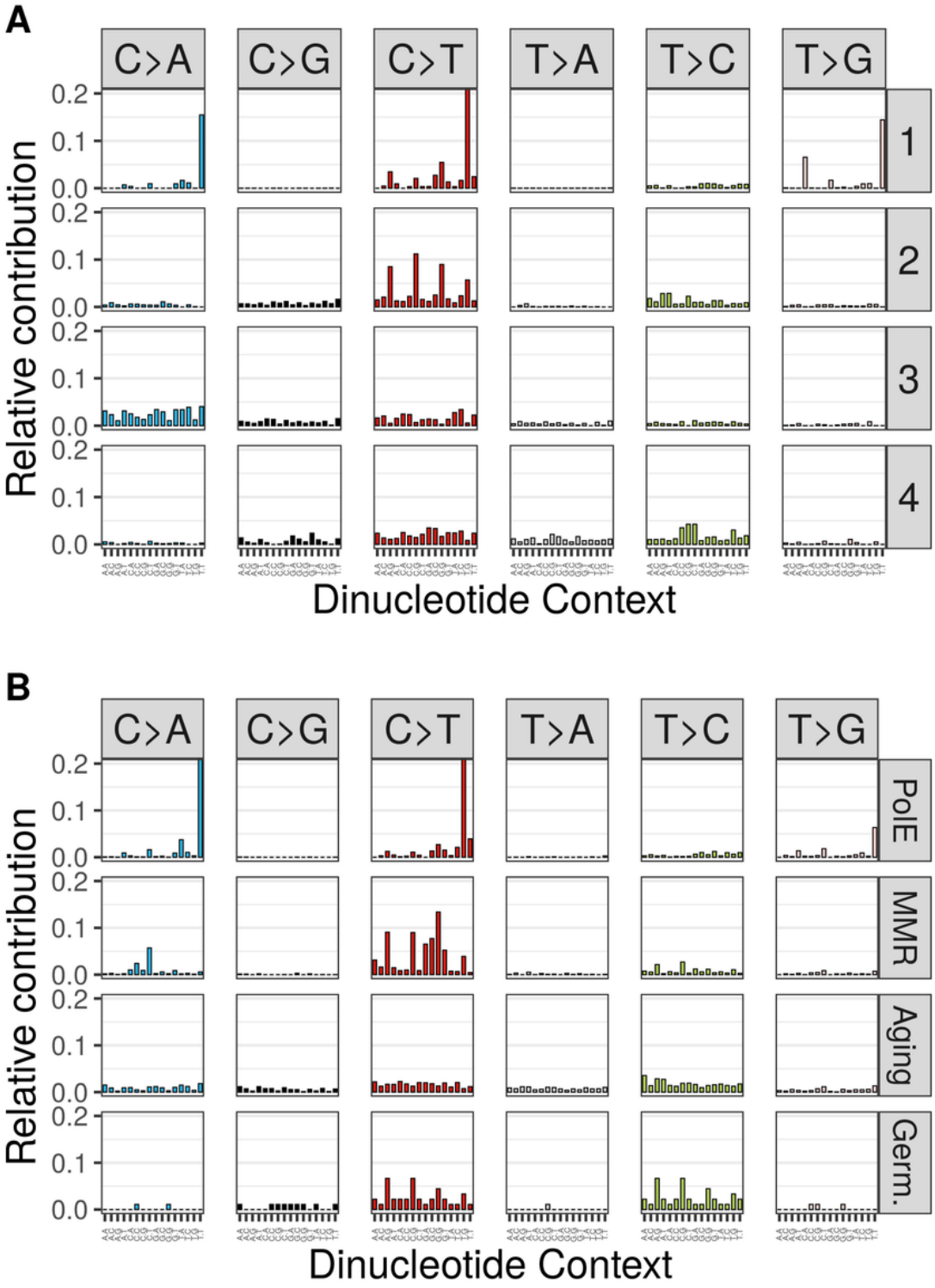

Figure 1

Mutational Signatures from RNA Sequencing of Tissues with and without Ovarian Cancer (OC). A, The four mutational signatures according to dinucleotide context in the Mayo Clinic and TCGA OC tumors combined. B, The reference Catalogue of Somatic Mutations in Cancer (COSMIC) signatures with high similarity to the signatures in the analyzed samples. The four signatures are DNA polymerase epsilon mutations (POLE), mismatch repair defects (MMR), patient age, and germline signature (Germ).

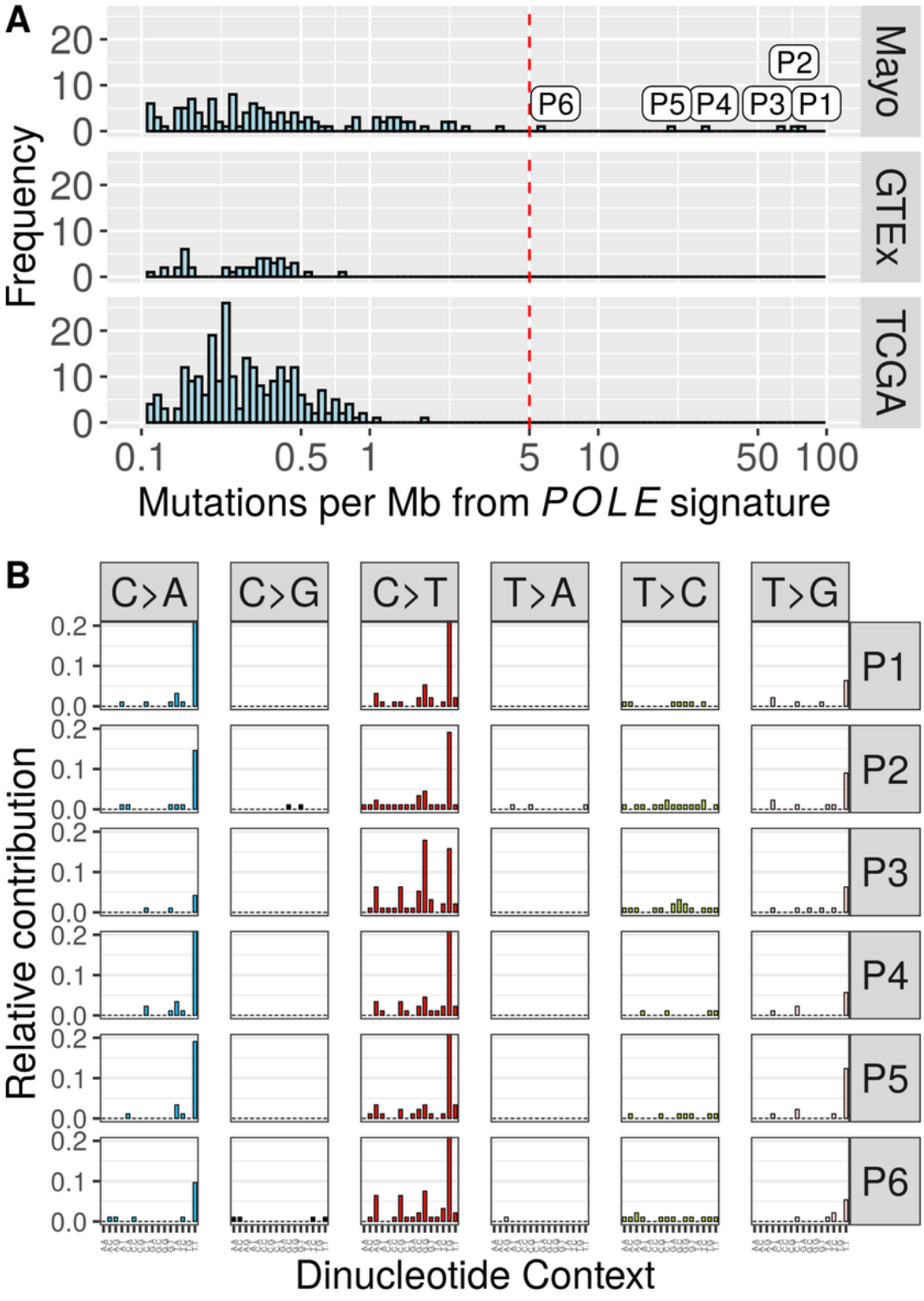

Figure 2

Identification of DNA Polymerase Epsilon (POLE)-Driven Ovarian Cancer Samples. A, Distribution of the tumor mutation burden (TMB) attributed to the POLE signature across the cohorts from Mayo Clinic (Mayo), Genotype-Tissue Expression (GTEx), and The Cancer Genome Atlas (TCGA), which allows the identification of six samples (P1-P6) with high POLE contribution and more than five mutations per megabase. The red dashed line indicates a threshold used to separate between POLE positive and

negative cases. B, Mutational profiles of six POLE-driven Mayo Clinic OEC tumors according to the frequency of dinucleotide context for the six samples; their distinct C>A and C>T peaks are associated with the POLE signature.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- poleeocbmcsubmision111supplemental.docx