Mod Pathol 38 (2025) 100843

MODERN PATHOLOGY

USCAP UNITED STATES AND CANADIAN ACADEMY OF PATHOLOGY
Advancing Pathology Together

Journal homepage: https://modernpathology.org/

Research Article

# Comparative Analysis of Somatic and Germline Polymerase Proofreading Deficiencies in Cancer: Molecular and Clinical Implications

Julen Viana-Errasti[a,b,c,d], Raúl Marín[a,b], Sandra García-Mulero[b,g,h,i], Tirso Pons[e], Mariona Terradas[a,b,c,f], Gabriel Capellá[a,b,c], Victor Moreno[b,g,h,i], Pilar Mur[a,b,c,j,*], Laura Valle[a,b,c,*]

[a] *Hereditary Cancer Program, Catalan Institute of Oncology, IDIBELL, Hospitalet de Llobregat, Barcelona, Spain;* [b] *Program in Molecular Mechanisms and Experimental Therapy in Oncology (Oncobell), IDIBELL, Hospitalet de Llobregat, Barcelona, Spain;* [c] *Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Madrid, Spain;* [d] *Programa de Doctorat en Biomedicina, Universitat de Barcelona (UB), Barcelona, Spain;* [e] *Department of Immunology and Oncology, National Center for Biotechnology (CNB-CSIC), Spanish National Research Council, Madrid, Spain;* [f] *Unitat de Biologia Cel·lular i Genètica Mèdica, Facultat de Medicina, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain;* [g] *Oncology Data Analytics Program (ODAP), Catalan Institute of Oncology, Hospitalet de Llobregat, Barcelona, Spain;* [h] *Department of Clinical Sciences, Faculty of Medicine and Health Sciences, Universitat de Barcelona Institute of Complex Systems (UBICS), University of Barcelona, Hospitalet de Llobregat, Barcelona, Spain;* [i] *Consorcio de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain;* [j] *Catalan Cancer Plan, Department of Health of Catalonia, Barcelona, Spain*

## ARTICLE INFO

*Article history:*
Received 28 April 2025
Revised 16 June 2025
Accepted 8 July 2025
Available online 16 July 2025

*Keywords:*
hereditary cancer
mutational signatures
POLD1
POLE
polymerase proofreading-associated polyposis
tumor mutational burden

## ABSTRACT

Polymerases ε and δ maintain genome integrity through exonuclease proofreading. Germline and somatic pathogenic variants (PVs) in the exonuclease domain (ED) of *POLE* and *POLD1* impair proofreading, causing hypermutated tumors. Despite shared mutational features that make these tumors highly immunogenic, molecular and clinical distinctions between *POLE* and *POLD1* mutations and between somatic and germline variants remain incompletely understood. We compared the molecular and clinical characteristics of *POLE* and *POLD1* ED PVs (n = 31), assessing their location, pathogenicity, clinical phenotypes, mismatch repair (MMR) status, tumor mutational burden, and signatures. We analyzed 360 proofreading-deficient tumors (source: The Cancer Genome Atlas [TCGA] and Catalogue Of Somatic Mutations In Cancer [COSMIC]) and 70 families (249 individuals) with polymerase proofreading-associated polyposis. All germline and somatic PVs had high AlphaMissense scores (0.87-1) and clustered within or near Exo motifs. Recurrent, non-founder germline PVs, *POLE* L424V and *POLD1* S478N, showed low/modest REVEL scores. Somatic variants occurred mainly in endometrial cancers (75% of proofreading-deficient TCGA cancers), whereas colorectal cancer predominated in polymerase proofreading-associated polyposis (56% of carriers). Cancer risks and tumor spectra differed between *POLE* and *POLD1* PV carriers. Aggressive hereditary phenotypes were linked to either specific *POLE* PVs (eg, S297F, V411L, P436R, M444K, A456P, and S461T) or the co-occurrence of germline ED PVs with germline MMR gene PVs. Distinct hypermutator profiles were confirmed for polymerase ε and polymerase δ proofreading deficiencies via unique mutational signatures (Polymerase ε: SBS10a/b, SBS28; Polymerase δ: SBS10c/d). Tumors with combined proofreading and MMR deficiencies had significantly higher tumor mutational burden and a shift in the associated mutational spectra. Unlike *POLE*, *POLD1* ED PVs exhibited haplosufficiency, typically requiring a somatic second hit (eg, loss of heterozygosity) or MMR deficiency to drive hypermutation. In conclusion, differences between *POLE* and *POLD1* and between somatic and germline mutations influence clinical presentation, mutagenic potential, and reliance on cooperating defects in tumorigenesis. These insights advance the understanding of

* Corresponding authors.
E-mail addresses: lvalle@idibell.cat (L. Valle), pmur@catsalut.cat (P. Mur).


https://doi.org/10.1016/j.modpat.2025.100843

proofreading-deficient cancers, with implications for diagnostics, genetic counseling, and precision oncology.



## Introduction

Replicative polymerases ε (Pol ε) and δ (Pol δ) have an exonuclease domain (ED) responsible for proofreading, ensuring high replication fidelity.[1,2] *POLE* and *POLD1* genes encode the largest subunit of the corresponding polymerases, which contains the catalytic polymerase and ED. The proofreading function, provided by the exonuclease, is essential to maintain genomic stability, and its disruption leads to hypermutation, contributing to tumorigenesis.

Somatic *POLE* ED pathogenic variants (PVs) are well-known cancer drivers (https://www.intogen.org/search?gene=POLE),[3] occurring in 8% to 10% of endometrial cancers and 1% to 3% of colorectal cancers (CRCs).[4-19] These PVs are also found in glioblastoma, ovarian, pancreatic, gastric, and other cancers, although at lower frequencies. Somatic *POLD1* ED PVs are rarely identified in tumors. Germline PVs (gPVs) in *POLE* and *POLD1* EDs cause a rare cancer predisposition syndrome—polymerase proofreading-associated polyposis (PPAP)—characterized by an increased risk of gastrointestinal polyposis, CRC, endometrial cancer, and other malignancies.[20,21]

Both somatic and germline proofreading deficiencies result in hypermutated tumors, typically exceeding 10 mutations per megabase (Mb) and frequently surpassing 100 mutations/Mb, classifying them as ultramutated. These tumors exhibit distinctive mutational spectra, notably enriched for TCT>TAT, TCG>TTG, and TTT>TGT substitutions. Copy number alterations are rare in this context.[14,6,22-25]

Mutational signatures SBS10a, SBS10b, and SBS28 are associated with Pol ε proofreading deficiency, whereas SBS10c and SBS10d are linked to Pol δ proofreading deficiency.[21,23,26-28] Importantly, tumors with proofreading deficiency exhibit high immunogenicity due to their elevated tumor mutational burden (TMB), which has both prognostic and therapeutic implications. These tumors are often responsive to immune checkpoint inhibitors.[29-31]

In this study, we compare the characteristics of known somatic and germline pathogenic ED variants and evaluate the molecular and clinical features of the associated tumors.

## Materials and Methods

### *Germline and Somatic Pathogenic Variants*

Germline variants classified as pathogenic, likely pathogenic, or variant of unknown significance (VUS) with strong evidence of pathogenicity ("hot VUS")—hereinafter referred to as gPVs—were included in the study. Variant classification was performed following the specifications developed by our group for the classification of *POLE* and *POLD1* ED variants based on the guidelines provided by the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP)[32] (Supplementary Table S1; last update of evidence: May 2024).[21] The term "hot VUS" was applied to those variants that did not reach the "likely pathogenic" classification due to the lack of only 1 supporting piece of evidence.

Somatic driver PVs were considered when the variant was identified in at least 1 tumor that harbored the corresponding gene-specific mutational signature(s), in the absence of another variant (P, LP, or VUS) in the ED of the same polymerase gene. For Pol δ proofreading- and mismatch repair-deficient (dMMR) tumors, contributions >50% for SBS20 were considered, due to the lower specificity of this signature.[21]

### *Variant Nomenclature*

Variant nomenclature follows the Human Genome Variation Society (HGVS) recommendations (v.21.0.4),[34] with nucleotide 1 corresponding to the A of the ATG translation initiation codon. All variants were annotated according to RefSeq IDs LRG_789; NM_006231.4 (*POLE*) and LRG_785; NM_001256849.1 (≈NM_002691.4) (*POLD1*). POLE ED includes amino acids 268 to 471, and POLD1 ED, amino acids 304 to533 (based on NCBI: "region_name DNA_polB_epsilon_exo and DNA_polB_delta_exo").

### *In Silico Analyses*

Pathogenicity predictions were obtained using REVEL[35,33] and AlphaMissense.[36] The scores were obtained from Ensembl Variant Effect Predictor (VEP).[37]

Direct contact of an amino acid with the DNA (positioned for proofreading) was defined when any atom of the amino acid is accessible to the cavity where the DNA binds and is at less than 6 Å from the DNA, and indirect contact when it is accessible to the cavity by at ≥ 6 Å from the DNA. No contact was considered when the atoms of an amino acid are not accessible to the DNA-binding cavity. Methodologic details and predictions for each residue of the ED of Pol ε and Pol δ are available elsewhere.[21]

### *Families With Polymerase Proofreading-Associated Polyposis*

A total of 70 families carrying any of the *POLE* or *POLD1* gPVs listed in Table 1 were identified in the literature (Supplementary Table S1). These families, with a total of 249 heterozygous carrier individuals, were considered for the phenotypic characterization of the hereditary cancer syndrome.

### *Polymerase Proofreading Tumors From The Cancer Genome Atlas (TCGA) And The Catalogue Of Somatic Mutations In Cancer (COSMIC)*

A total of 78 polymerase proofreading-deficient cancers from The Cancer Genome Atlas (TCGA) were included in the study: 74 harbored PVs in *POLE* ED, and 4 in *POLD1*. The cancer types represented among the proofreading-deficient tumors included 55 endometrial adenocarcinomas, 11 CRCs, 3 gliomas, 2 breast cancers, 2 cervix cancers, 1 bladder cancer, 3 extracolonic gastrointestinal cancers, and 1 prostate cancer (Supplementary Table S2). Histopathologic information was obtained from TCGA portal, and

**Table 1**
List of germline variants classified as pathogenic, or likely pathogenic according to gene-specific recommendations of the ACMG/AMP guidelines[21] and of somatic variants classified as drivers based on the presence of the corresponding proofreading-deficient mutational signatures (see Materials and Methods)

| ED variant | Variant classification[a] | AM | REVEL | Exo Motif[b] | DNA-binding cleft (distance to DNA)[c] | No. of tumors with somatic variant (MSS/MSI)[d] | No. of families with germline variant (affected/ unaffected heterozygotes)[e] |
|---|---|---|---|---|---|---|---|
| *Predominantly germline variants* | | | | | | | |
| *POLE*:c.830A>G; p.E277G | LP | 1 | 0.835 | Exo I -catalytic site | Yes (<6Å) | - | 1 (7/0) |
| *POLE*:c.833C>A; p.T278K | LP | 1 | 0.666 | Exo I | Yes (<6Å) | - | 1 (4/0) |
| *POLE*:c.881T>G; p.M294R | LP | 0.99 | 0.815 | Flanking Exo I | Yes (<6Å) | - | 3 (7/2) |
| *POLE*:c.1089C>G; p.N363K | P | 0.99 | 0.735 | Exo II | Yes (<6Å) | - | 4 (41/0) |
| *POLE*:c.1102G>A; p.D368N | Hot VUS | 0.99 | 0.529 | Exo II | Yes (<6Å) | - | 1 (6/1) |
| *POLE*:c.1270C>G; p.L424V | P/SomD | 0.96 | 0.654 | Exo IV | Yes (<6Å) | 6 (4/0) | 30 (71/4) |
| *POLE*:c.1373A>T; p.Y458F | LP | 0.87 | 0.492 | Exo III | Yes (<6Å) | - | 1 (25/2) |
| *POLE*:c.1381T>A; p.S461T | LP | 0.91 | 0.587 | Exo III | Yes (>6Å) | - | 1 (1/0) |
| *POLD1*:c.946G>C; p.D316H | LP | 1 | 0.743 | Exo I -catalytic site | Yes (<6Å) | - | 1 (2/0) |
| *POLD1*:c.1204G>A; p.D402N | LP | 1 | 0.484 | Exo II | Yes (<6Å) | - | 2 (2/0) |
| *POLD1*:c.1421T>C; p.L474P | P | 1 | 0.913 | Exo IV | Yes (<6Å) | - | 5 (14/8) |
| *POLD1*:c.1433G>A; p.S478N | P/SomD | 0.99 | 0.377 | Exo IV | Yes (>6Å) | 2 (1/1) | 6 (16/1) |
| *Predominantly somatic variants* | | | | | | | |
| *POLE*:c.857C>G; p.P286R | P/SomD | 1 | 0.837 | Flanking Exo I | Yes (<6Å) | 161 (46/4) | 1 (1/0) |
| *POLE*:c.890C>A; p.S297Y | SomD | 1 | 0.792 | Flanking Exo I | No | 1 (0/1) | - |
| *POLE*:c.890C>T; p.S297F | P/SomD | 1 | 0.799 | Flanking Exo I | No | 21 (5/1) | 1 (1/0) |
| *POLE*:c.1087A>G; p.N363D | SomD | 0.98 | 0.536 | Exo II | Yes (<6Å) | 1 (0/1) | - |
| *POLE*:c.1100T>C; p.F367S | SomD | 1 | 0.784 | Exo II | Yes (<6Å) | 3 (2/1) | - |
| *POLE*:c.1102G>T; p.D368Y | SomD | 0.99 | 0.74 | Exo II | Yes (<6Å) | 2 (1/1) | - |
| *POLE*:c.1231G>T; p.V411L | P/SomD | 0.95 | 0.457 | Flanking Exo IV | Yes (>6Å) | 107 (29/8) | 2 (2/0) |
| *POLE*:c.1270C>A; p.L424I | SomD | 0.94 | 0.593 | Exo IV | Yes (<6Å) | 4 (2/2) | - |
| *POLE*:c.1307C>G; p.P436R | LP/SomD | 1 | 0.592 | Exo V | Yes (>6Å) | 6 (2/1) | 1 (1/0) |
| *POLE*:c.1366G>C; p.A456P | LP/SomD | 1 | 0.62 | Exo III | No | 22 (7/1) | 1 (1/0) |
| *POLE*:c.1376C>T; p.S459F | SomD | 1 | 0.418 | Exo III | Yes (>6Å) | 11 (8/1) | - |
| *POLE*:c.1382C>T; p.S461L | SomD | 0.99 | 0.582 | Exo III | Yes (>6Å) | 1 (0/1) | - |
| *POLE*:c.1394C>T; p.A465V | SomD | 0.98 | 0.599 | Exo III | Yes (>6Å) | 5 (1/3) | - |
| *Germline and somatic variants* | | | | | | | |
| *POLE*:c.824A>T; p.D275V | Hot VUS | 1 | 0.817 | Exo I -catalytic site | Yes (>6Å) | 1 (1/0) | 1 (2/0) |
| *POLE*:c.857C>T; p.P286L | LP/SomD | 0.99 | 0.812 | Flanking Exo I | Yes (<6Å) | 2 (1/1) | 1 (15/1) |
| *POLE*:c.1331T>A; p.M444K | LP/SomD | 0.99 | 0.621 | Flanking Exo V | Yes (<6Å) | 2 (2/0) | 1 (1/0) |
| *POLD1*:c.947A>G; p.D316G | LP/SomD | 1 | 0.773 | Exo I -catalytic site | Yes (<6Å) | 1 (0/1) | 1 (2/0) |
| *POLD1*:c.946G>A; p.D316N | P/SomD | 0.99 | 0.587 | Exo I -catalytic site | Yes (<6Å) | 2 (0/2) | 2 (5/1) |
| *POLD1*:c.952G>A; p.E318K | SomD | 1 | 0.6 | Exo I -catalytic site | Yes (<6Å) | 1 (0/1) | 2 (2/0) |

Germline VUS lacking only 1 supporting evidence to be classified as (L)P were also included and called "hot VUS." Variants were categorized as predominantly germline, predominantly somatic, or both germline and somatic, according to their relative prevalence in tumors (somatic) or as a germline alteration. In silico pathogenicity prediction scores, and location of the variants with respect to the exonuclease Exo Motif and within the DNA-binding cleft of the exonuclease, were included in the table. ACMG/AMP, the American College of Medical Genetics and Genomics and the Association for Molecular Pathology; AM, AlphaMissense; COSMIC, the Catalogue Of Somatic Mutations In Cancer; LP, likely pathogenic; MSI, microsatellite instability (number of tumors with MSI); MSS, microsatellite stability (number of tumors with MSS); P, pathogenic; SomD, somatic driver; TCGA, The Cancer Genome Atlas; VUS, variant of unknown significance.

[a] See Material and Methods to know about the criteria applied for variant classification, and Supplementary Table S1 for details on the evidence used for the classification of germline variants.

[b] Regions within ±15 amino acids of the Exo motifs were considered flanking.

[c] This data indicates if the affected amino acid is located within the DNA-binding pocket of the exonuclease (yes), or if it is outside (no). See Materials and Methods section for details.

[d] TCGA and COSMIC cancers considered. MMR deficiency status (MSI/MSS) was only available for 60 of 282 COSMIC tumors, whereas all TCGA cases included this information. pMMR tumors include those categorized as MSS, low-MSI and low-MSI/MSS in the databases. dMMR tumors were considered when categorized as MSI or high-MSI.

[e] Any type of cancer and/or the presence of gastrointestinal adenomas were considered. Details of the families and phenotypes in Supplementary Table S1.

overall survival (Kaplan-Meier) curves comparing TCGA endometrial and CRCs with and without PVs in *POLE* or *POLD1* ED were obtained through cBioPortal (https://www.cbioportal.org/).[5]

Tumor sequencing data from 282 proofreading-deficient cancers (with confirmed somatic PVs and excluding cultured samples and tumors of unknown origin) were obtained from the Catalogue Of Somatic Mutations in Cancer (COSMIC; accessed May 2024): 281 harbored exonuclease PVs in *POLE* and 1 in *POLD1*. Cancer types represented included 214 endometrial carcinomas, 43 CRCs, 10 ovarian cancers, 8 brain tumors, and 7 tumors in other locations (Supplementary Table S2).

All variants in TCGA and COSMIC tumors were somatic in nature, and therefore, no overlapping with the PPAP cases occurred.

### *Tumor Mismatch Repair Deficiency Determination*

Scores of microsatellite instability (MSI) estimated by MANTIS and MSIsensor were retrieved from TCGA.

Thresholds of 0.4 for MANTIS and 3.5 for MSIsensor were applied.[38,39] Tumors were classified as dMMR if at least 1 of the 2 scores exceeded the respective threshold. When available, MSI status from the Bethesda panel was prioritized for classification.[40,41] For samples with undefined MMR status, classification was inferred based on COSMIC mutational signatures: tumors were considered dMMR if the contribution of dMMR-associated signatures exceeded 20% of the total signature profile.

For tumors obtained from COSMIC, MMR status was taken directly from the database annotations. Microsatellite stable (MSS) and MSI-low tumors were considered mismatch repair proficient (pMMR), whereas MSI-high tumors were classified as dMMR.

### *Tumor Mutational Burden and Signatures*

Exome sequencing data (BAM files) or targeted sequencing data (covering ≥100 genes) from tumors harboring *POLE* and *POLD1* ED variants were obtained from TCGA and COSMIC v.94 (accessed May 2024). Of the tumors meeting these criteria, all but 1 had exome sequencing data available. The remaining case, sourced from COSMIC, had been sequenced using a 410-gene targeted panel.

Sequencing data were processed to calculate TMB and mutational signatures as previously described.[21] In total, 78 TCGA and 12 COSMIC tumors were analyzed. Signature attribution was performed using FitMS through the Signal web application (https://signal.mutationalsignatures.com/),[42] accessed in May 2024, without selection of tissue-specific signatures. For *POLE* variants, COSMIC v3 signatures were considered, which include SBS10a, SBS10b, SBS28, and SBS14. For *POLD1* variants, the Cancer Reference Signatures were considered, which include SBS10d and SBS20. Although SBS10c is also associated with Pol δ proofreading deficiency, it is typically observed only in normal tissues—such as sperm, blood, or normal colon crypts—from individuals carrying germline *POLD1* ED variants.[26,43] Since this study focuses on tumor samples, where SBS10c is generally absent, it was excluded from the analysis. Furthermore, SBS10c is not included in the Cancer Reference Signatures used by Signal, reflecting its limited relevance in cancer.

### *Statistical Analysis*

Statistical analyses were conducted using R statistical software (v4.4.1, R Foundation for Statistical Computing) and GraphPad QuickCalcs (www.graphpad.com). Group comparisons were performed using parametric tests, including Spearman *t* tests and analysis of variance. Differences in frequencies were assessed using the proportional and Fisher exact tests. Survival curves were compared using log-rank tests, as provided in cBioPortal. All statistical tests were 2-sided, with significance set at $P < .05$.

Unsupervised hierarchical clustering of TCGA tumors with and without *POLE* ED PVs was based on the contribution percentages of Pol ε proofreading-deficient signatures (SBS10a, SBS10b, SBS14, and SBS28), dMMR signatures (SBS6, SBS15, SBS21, SBS26, and SBS44), and "other signatures," encompassing any assigned COSMIC v.3 SBS and unassigned signatures (FitMS-Signal results). Euclidean distance was used to compute sample distances, and the Ward-D2 method was applied for clustering. The ComplexHeatmap package was used for both clustering computation and heatmap visualization.

## Results

A diagram that summarizes the design of this study is included as Figure 1.

### *Nature, Frequency, and Location of Germline and Somatic Exonuclease Domain Pathogenic Variants Affecting Polymerase Proofreading*

Based on the criteria detailed in the Materials and Methods section for the inclusion of germline and somatic variants (Supplementary Tables S1 and S3), a total of 31 PVs, all of them of missense nature, were identified for this study (Table 1). The relative frequency of each PV and its location in the ED linear sequence are represented in Figure 2.

Among the 22 gPVs—classified as P, LP, or hot VUS—identified in individuals with PPAP, 12 were categorized as predominantly germline. The most recurrent variants (ie, identified in ≥3 unrelated families), listed in order of decreasing frequency, were: *POLE* L424V, *POLD1* S478N, *POLD1* L474P, *POLE* N363K, and *POLE* M294R (Table 1).

In total, 20 somatic driver variants were identified across tumors from TCGA and COSMIC data sets—16 in *POLE* and 4 in *POLD1*. Thirteen of these were classified as predominantly somatic. The most recurrent somatic hotspots (defined as occurring in ≥10 tumors), in order of decreasing frequency, were: *POLE* P286R, V411L, A456P, S297F, and S459F (Table 1).

Of the 31 PVs identified—either as gPVs or somatic drivers—13 were found in both PPAP germline and somatic contexts. Notably, the germline-to-somatic distribution of these shared variants was often highly unbalanced: Variants such as *POLE* P286R, V411L, S297F, A456P, and S459F are somatic hotspots but rarely observed in the germline, while *POLE* L424V and *POLD1* S478N are the most frequent gPVs in PPAP and are seldom detected as somatic alterations in tumors (Table 1; Fig. 2).

Polymerase proofreading activity relies on highly conserved motifs within the ED—the so-called Exo motifs (I-V)—and 2 catalytic residues located in EXO I that coordinate metal ion binding: D275 and E277 in *POLE*, and D316 and E318 in *POLD1*.[44,45] Germline and somatic PVs cluster within or near the Exo motifs (Fig. 2).

In *POLE*, ED PVs—whether germline or somatic—are broadly distributed across all 5 Exo motifs and their flanking regions. All known pathogenic *POLD1* ED variants, which affect only 5 protein residues to date, are located within Exo motifs I, II, or IV. Notably, the exonuclease catalytic sites are affected by both somatic and germline variants in both genes (Fig. 2).

Beyond their linear position in the ED, pathogenicity may also relate to a variant's location within the DNA-binding cleft of the exonuclease's 3-dimensional structure. Variants that either directly contact the DNA or are embedded within the DNA-binding region are thought to have a greater impact on proofreading function and/or may induce a stronger mutagenic effect.[46–49] The majority of the 31 variants classified gPVs or somatic drivers affect residues that form the DNA-binding cleft of the exonuclease (28/31; 90%). Of these, 71% (20/28) affect amino acids with atoms that are at less than 6 Å from the DNA, that is, in direct contact with it (Table 1).

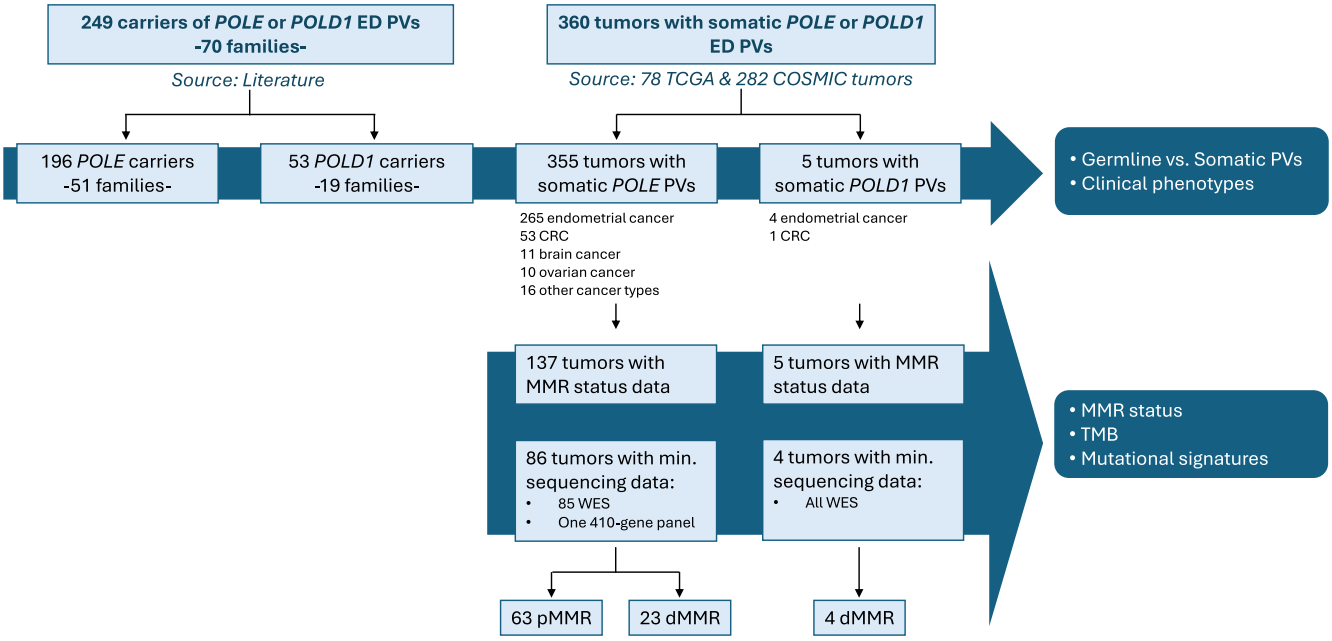


**Figure 1.**
Schematic representation of the study design.

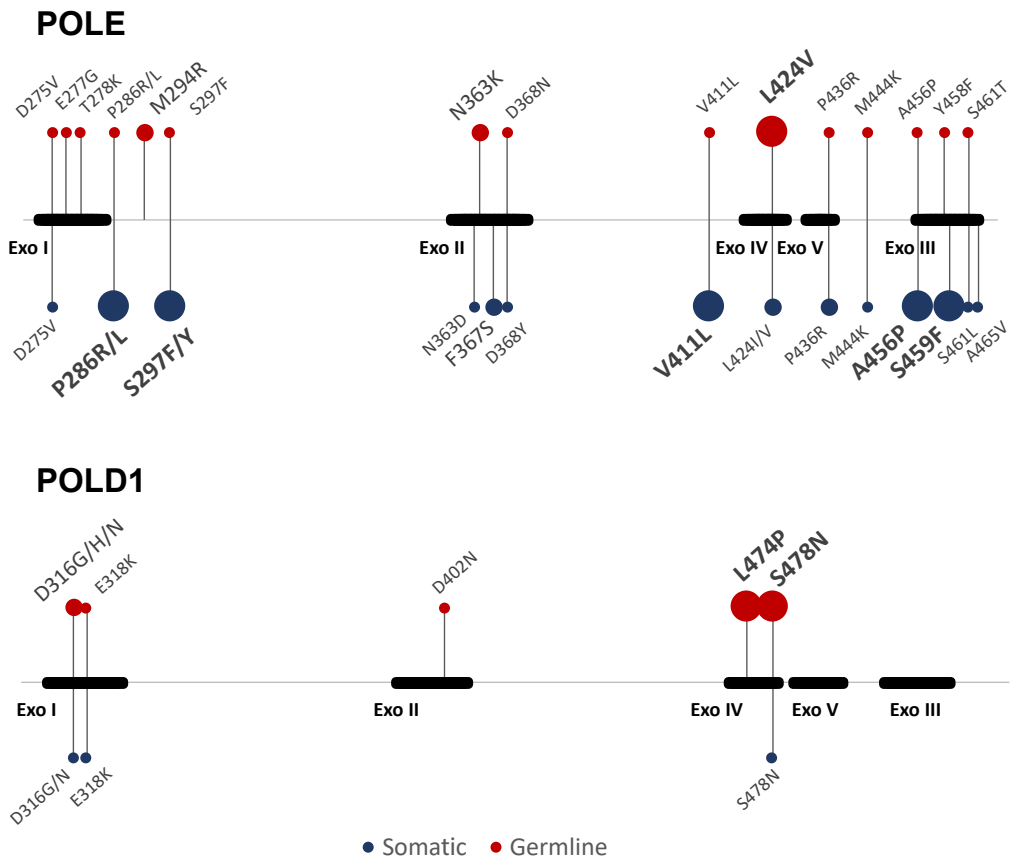


**Figure 2.**
Location in the exonuclease domain sequence of POLE and POLD1 of somatic (blue dots) and germline (red dots) variants. Exo motifs (I–V) are represented in the diagram as thick black lines (POLE: I, aa 271-285; II, aa 359-372; III, aa 453-466; IV, aa 420-428; and V, aa 430-438. POLD1: I, aa 312-326; II, aa 393-406; III, aa 506-519; IV, aa 470-478; and V, aa 485-493). The size of each colored dot represents the frequency of the corresponding variant: large dots represent the most frequent somatic mutation hotspots (≥10 tumors in Table 1) and the most prevalent germline variants (≥5 families in Table 1); medium-size dots, less recurrent variants (somatic: 3-9 tumors; germline: 3-4 families); and small dots, variants identified in ≤ 2 tumors or ≤ 2 families. aa, amino acid.

**Table 2**
Clinical phenotypes of individuals carrying a gPV in the ED of *POLE* or *POLD1*

| Clinical phenotypes | *POLE* or *POLD1* ED PV heterozygotes<br>Affected vs total no. of heterozygotes<br>Median age at diagnosis (range) | *POLE* ED PV heterozygotes<br>Affected vs total no. of heterozygotes<br>Median age at diagnosis (range) | *POLD1* ED PV heterozygotes<br>Affected vs total no. of heterozygotes<br>Median age at diagnosis (range) | *POLE* vs *POLD1* PV ED heterozygotes<br>(2-tailed Fisher exact test) |
|---|---|---|---|---|
| Colorectal cancer<br>Median age (range) | 139/249 (56%)<br>Age (y): 42 (13-88) | 115/196 (59%)<br>Age (y): 43 (13-88) | 24/53 (45%)<br>Age (y): 34.5 (17-80) | $P = .0884$ |
| Endometrial cancer[a], female<br>Median age (range) | 28/127 (22%)<br>Age (y): 50 (30-57) | 15/94 (16%)<br>Age (y): 50 (30-56) | 13/33 (39%)<br>Age (y): 51 (31-57) | $P = .0077$ |
| Breast cancer[a], female<br>Median age (range) | 13/127 (10%)<br>Age (y): 52 (38-65) | 8/94 (9%)<br>Age (y): 48.5 (38-65) | 5/33 (15%)<br>Age (y): 57 (52-65) | $P = .3202$ |
| Ovarian cancer[a], female<br>Median age (range) | 9/127 (7%)<br>Age (y): 40 (33-57) | 9/94 (10%)<br>Age (y): 40 (33-57) | 0/33 (0%) | $P = .1104$ |
| Brain cancer<br>Median age (range) | 23/249 (9%)<br>Age (y): 30 (4-66) | 22/196 (11%)<br>Age (y): 30 (4-66) | 2/53 (4%)<br>Age (y): 26[d] | $P = .1210$ |
| Extracolonic GI cancers<br>Median age (range) | 20/249 (8%)<br>Age (y): 48 (34-78) | 19/196 (10%)<br>Age (y): 48 (34-78) | 1/53 (2%)<br>Age (y): 36 | $P = .0851$ |
| Other cancers<br>Median age (range) | 14/249 (6%)<br>Age (y): 48 (17-78) | 11/196 (6%)<br>Age (y): 47 (31-78) | 3/53 (6%)<br>Age (y): 52 (17-58) | $P = 1.0000$ |
| Multiple primary cancers | 83/249 (33%) | 63/196 (32%) | 20/53 (38%) | $P = .5116$ |
| GI polyposis[b]<br>Median age (range) | 69/249 (28%)<br>Age (y): 40 (13-71) | 57/196 (29%)<br>Age (y): 38 (13-71) | 12/53 (23%)<br>Age (y): 41 (19-58) | $P = .3918$ |
| CALMs (reported) | 13/249 (5%) | 13/196 (7%) | 0/53 (0%) | $P = .0765$ |
| No cancer | 63/249 (25%) | 45/196 (23%) | 18/53 (34%) | $P = .1112$ |
| No cancer, no benign tumor phenotype, and no CALMs<br>Median age (range) | 20/249 (8%)<br>Age (y): 34 (18-44)[c] | 10/196 (5%)<br>Age (y): 34 (2-44) | 10/53 (19%)<br>Age (y): 31 (18-44) | $P = .0028$ |

Cases with variants classified as P, LP, hot VUS or somatic drivers (Table 1) reported in the literature were included (last search: May 2024). Polyp information and other phenotypic details are shown in Supplementary Table S4.
CALMs, Café-au-lait macules; GI, gastrointestinal; PV, pathogenic variant (includes pathogenic, likely pathogenic and "hot VUS").
[a] Only female heterozygotes were considered. No male heterozygotes were diagnosed with breast cancer.
[b] Individuals reported as having >10 polyps (type not determined), >10 adenomas, polyposis, multiple polyps or multiple adenomas. Sixty-nine heterozygotes fulfilled this criterion, 13 had <10 but ≥5 polyps, 64 had 1 to 4 polyps, and 112 had 0 polyps at the time of ascertainment or no data regarding polyps.
[c] Available age information in 4 of 20 individuals.
[d] Available information on age at brain cancer diagnosis for 1 of 2 patients.

*POLE* and *POLD1* gPVs and somatic drivers exhibit consistently high AlphaMissense scores (mean, 0.98; median, 0.99; range, 0.87-1), with 93.5% (29/31) scoring ≥0.94 (Table 1). In contrast, REVEL scores are more variable and generally lower, with a mean score of 0.658 (range, 0.377-0.913) across all 31 variants. There was no significant difference in REVEL scores between predominantly germline variants (mean, 0.653; range, 0.377-0.913) and somatic drivers (mean, 0.642; range, 0.418-0.837). Additional details on REVEL predictions are provided in Supplementary Results section.

### *Clinical Phenotype I: Tumor Spectrum of Polymerase Proofreading-Associated Polyposis (Hereditary Cancer Syndrome)*

Table 2 summarizes the clinical phenotypes observed in PPAP, based on data from 249 heterozygous carriers across 70 families reported in the literature (clinical features of each individual are detailed in Supplementary Table S4). As previously described by our group,[21] the primary PPAP-associated tumor types are colorectal, endometrial, ovarian, breast, brain, and upper gastrointestinal cancers, as well as polyposis (>10 adenomas). All these tumor phenotypes show a prevalence >10% among cancer-affected heterozygotes and are frequently associated with multiple primary cancers.

Although current sample sizes remain limited—particularly for *POLD1*—emerging data suggest distinct cancer risks and tumor spectra for *POLE* vs *POLD1* ED gPV carriers. Reported phenotypes indicate that endometrial cancer is more commonly diagnosed in heterozygous carriers of *POLD1* PVs, whereas ovarian cancer has so far been observed exclusively in *POLE* heterozygotes. Notably, up to 20% of *POLD1* ED heterozygotes and only 5% of *POLE* heterozygotes have been reported without tumors or café-au-lait macules, suggesting variable penetrance between the 2 genes (Table 2; Fig. 3A).

### *Clinical Phenotype II: Constitutional Mismatch Repair Deficiency-Like Presentations (Aggressive Hereditary Forms)*

Although PPAP-associated tumors—both benign (eg, polyps) and malignant—are typically diagnosed in adulthood, with mean ages at cancer diagnosis ranging 30-60 years (Table 2), a subset of cases presents with unusually aggressive phenotypes. These include polyposis and cancers—such as CRC and brain tumors—manifesting in childhood, adolescence, or early adulthood (Table 3).[50–56] The mean age at CRC diagnosis in reported cases is 16 years, with brain tumors, including medulloblastomas, developing as early as 4-5 years. Café-au-lait macules and nonmalignant tumors (eg, pilomatricomas) are often reported in these cases, whose phenotype mimics constitutional mismatch repair deficiency (CMMRD).

Notably, 3 of the 6 *POLE* variants linked to aggressive PPAP phenotypes—V411L, S297F, and A456P—are among the most recurrent somatic drivers. The M444K variant has been identified in 2 tumors as a somatic driver and in 1 CMMRD-like case as a germline variant. S461T had not been previously reported as

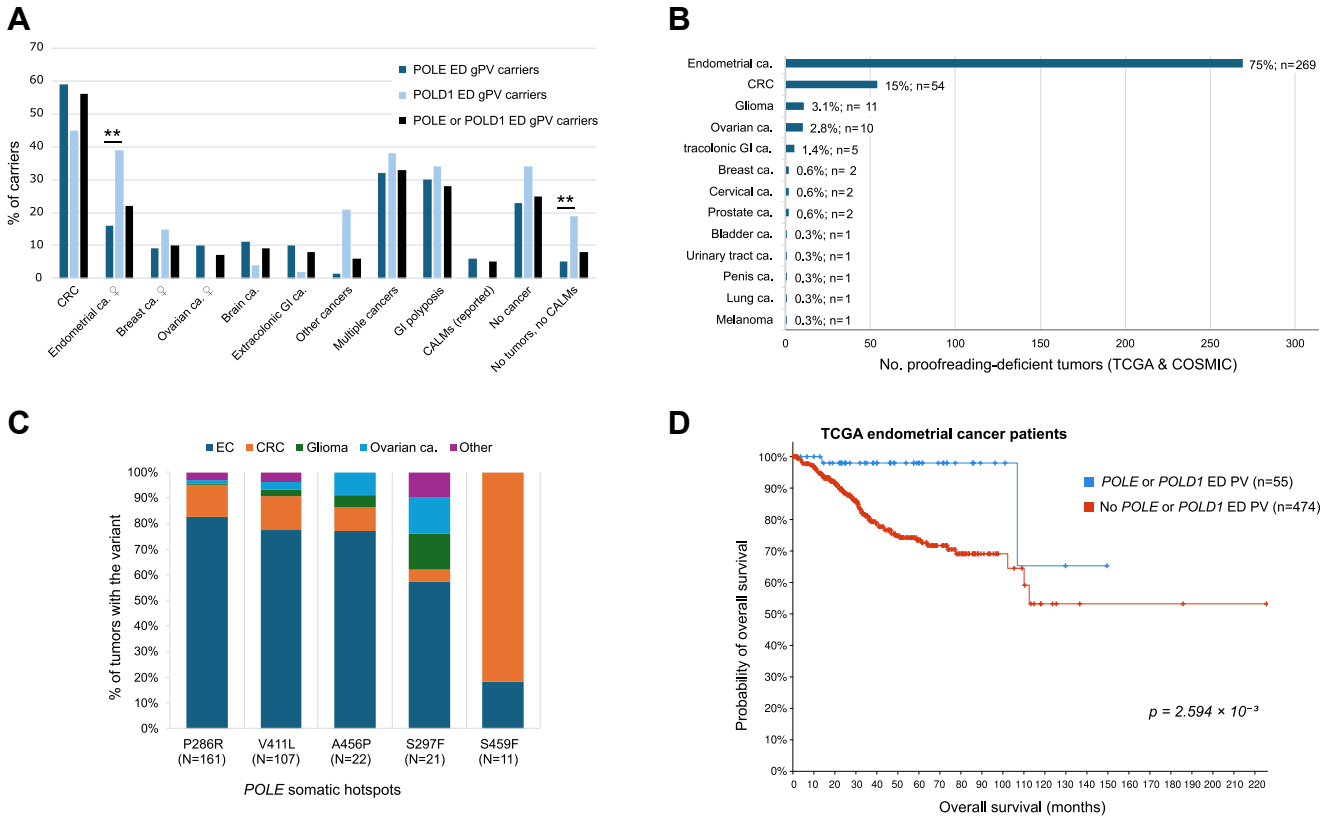


**Figure 3.**

(A) Prevalence of different tumor types and non-tumoral phenotypes among heterozygous carriers of *POLE* and *POLD1* ED pathogenic variants, based on the information detailed in Table 2. Phenotypic prevalences were compared between *POLE* and *POLD1* ED PV heterozygotes using a 2-tailed Fisher's exact test (significance levels: $*P < .05$, $**P < .01$, $***P < .001$). (B) TCGA and COSMIC tumor types harboring *POLE* or *POLD1* somatic driver variants and relative frequency of the different tumor types. (C) Distribution of tumor types according to the presence of the most recurrent somatic *POLE* mutation hotspots. (D) Kaplan-Meier survival curves showing overall survival for TCGA endometrial cancer patients with (n = 55) or without (n = 474) somatic pathogenic variants in the ED of *POLE* or *POLD1*. Log-rank test: $P = 2.594 \times 10^{-3}$ (Source: cBioPortal). ca., cancer; CALMs, café-au-lait macules; CRC, colorectal cancer; EC, endometrial cancer; ED, exonuclease domain; GI, gastrointestinal; gPV, germline pathogenic variant; PV, pathogenic variant.

either a somatic or germline variant except in the case described here. To date, no aggressive PPAP phenotypes have been linked to germline *POLD1* PVs following a monogenic model.

CMMRD-like phenotypes may also arise through digenic inheritance, where a *POLE* or *POLD1* ED germline variant co-occurs with a germline variant in an MMR gene. This has been documented in 2 unrelated families (Table 3).[50–56] One involved 2 siblings carrying both *POLD1* c.946G>A p.(D316N) and *PMS2* c.2007-786_2174+493del1447.[56] Another involved a 4-year-old medulloblastoma patient who inherited *POLE* c.830A>G p.(E277G) and carried a de novo *PMS2* c.2148dupC variant.[57] A potential third case, reported by the European Consortium "Care for CMMRD" (C4CMMRD) in 2019, involved a patient diagnosed with duodenal cancer at the age of 14 years who inherited a maternal *PMS2* germline variant and a paternal *POLE* ED VUS.[58]

### *Clinical Phenotype III: Proofreading-Deficient Cancers (Somatic Pathogenic Variants)*

Among the 360 tumors from TCGA and COSMIC datasets harboring somatic drivers in the ED of *POLE* and *POLD1*, 355 (98.6%) harbored PVs in *POLE* and 5 (1.4%) in *POLD1*. *POLE* ED-mutated tumor types included endometrial cancer (n = 269; 74.72%), CRC (n = 54; 15%), glioma (n = 11; 3.06%), and ovarian cancer (n = 10; 2.78%), with a small number of other tumor types represented (Fig. 3B).

Analysis of the tumor spectra associated with the most prevalent somatic drivers—*POLE* P286R, V411L, A456P, S297F, and S459F—revealed potential genotype-phenotype correlations, suggesting that certain variants may be preferentially associated with specific tumor types (Fig. 3C; Fisher exact test: 0.0005).

Histopathologic data provided further characterization of the proofreading-deficient subtype. Among 234 *POLE* ED-mutated endometrial cancers with available histologic classification, 192 (92.3%) were of endometrioid subtype, including 3 with mixed histologic features. Of the 79 endometrial cancers with histologic grading information, 56 (71%) were high-grade tumors (grade 3), consistent with prior reports.[59] All 54 (100%) CRCs were adenocarcinomas. Among the 10 central nervous system tumors with available histology, 8 (80%) were glioblastomas multiforme and 2 (20%) were astrocytomas. Of the 10 ovarian cancers, 9 (90%) exhibited endometrioid histology.

Consistent with previous analyses,[29] comparison of the overall survival of TCGA patients diagnosed with endometrial cancer with and without *POLE* or *POLD1* ED PVs confirmed the favorable prognosis associated with the proofreading-deficient subtype (source: TCGA; Kaplan-Meier curves and survival analysis generated via cBioPortal [https://www.cbioportal.org/]; Fig. 3D). The differences observed among CRC subtypes did not reach

**Table 3**
Characteristics of reported heterozygous carriers of *POLE* and *POLD1* ED pathogenic variants who developed aggressive phenotypic traits resembling CMMRD phenotypes

| Germline ED variant[a] | De novo | Digenic inheritance with an MMR pathogenic variant | Phenotype (age [y] at diagnosis) | Publication |
|---|---|---|---|---|
| *POLE*:c.890C>T; p.S297F [somatic hotspot] | de novo | No | CRC (20 y; pMMR); 50-60 adenomas; epithelioma/ pilomatricoma (28 y); glioblastoma (29 y; 2 somatic *MSH6* pathogenic variants; TMB 320 mut/Mb); multiple CALMs and skin papules and cysts. | Sehested et al, 2022[50] |
| *POLE*:c.1231G>T; p.V411L [somatic hotspot] | [b]Probably de novo | No | CRC (14 y; pMMR); >10 polyps (14 y); 6 CALMs; pilomatricoma. | Wimmer et al, 2017[51] |
| *POLE*:c.1307C>G; p.P436R | n.a. | No | CRC (17 y); anaplastic astrocytoma (17 y). | Galati et al, 2020[52] Shuen et al, 2019[53] |
| *POLE*:c.1331T>A; p.M444K | n.a. | No | Tectal plate glioma (11 y); CRC (13 y; pMMR; TMB 169 mut/ Mb); >100 adenomas (13 y); several CALMs; intramuscular venous malformation in right shoulder (3 y). | Sehested et al, 2022[50] |
| *POLE*:c.1366G>C; p.A456P [somatic hotspot] | [b]Probably de novo | No | Anaplastic medulloblastoma (5 y; pMMR; TMB >150 mut/ Mb), osteochondroma (5 y) and 3 pilomatricomas (5 y), renal cyst (5 y) >100 hyperpigmented macules and papules, including some with CALM features. | Lindsay et al, 2019[54] |
| *POLE*:c.1381T>A; p.S461T | de novo | No | Medulloblastoma (4 y; MSH6 loss; TMB 266 mut/Mb); multiple hyperpigmented skin spots, >6 reminiscent of CALMs. | Sehested et al, 2022[50] |
| *POLE*:c.830A>G; p.E277G | No | *PMS2*:c.2148dupC (de novo) | Medulloblastoma (4 y; PMS2 loss; dMMR; TMB 144-276 mut/Mb; SBS10a, SBS14 and dMMR signatures); CALMs. | Michaeli et al, 2022[55] |
| *POLD1*:c.946G>A; p.D316N | No | *PMS2*:c.2007-786_2174+493del1447 | Sibling I: CRC (17 y; dMMR); multiple adenomas (17 y); 2x CRC (27 y); urothelial carcinoma (34 y); nephrogenic adenoma of urinary bladder (36 y).<br>Sibling II (deceased at 21 y): 2x CRC (19 y); ≥7 adenomas (19 y). | Schamschula et al 2022[56] |

CALMs, café-au-lait macules; CRC, colorectal cancer; dMMR, mismatch repair deficiency; n.a., not available information; pMMR, mismatch repair proficiency; TMB, tumor mutation burden; y, years.
[a] There are 2 more individuals listed in Table 1 that carry in the germline a predominantly somatic variant (*POLE* V411L and P286R). However, there are no phenotypic details (except for having been diagnosed with CRC) in the publication where they were reported,[57] which is the reason why they were not included in the table.
[b] No family history of early-onset cancers or gastrointestinal polyposis, or other PPAP-associated features; mother wildtype (paternal branch not tested).

statistical significance, likely due to the small number of TCGA proofreading-deficient CRC patients (n = 11; Supplementary Fig. S1).

### *Tumor Molecular Features: Mismatch Repair Deficiency Status and Mutational Burden*

To ensure high-quality data for comparative analysis, sequencing data from PPAP tumors—primarily derived from formalin-fixed paraffin-embedded samples—were excluded. Due to the rarity of somatic *POLD1* ED PVs and the unique characteristics of Pol δ proofreading deficiency, these were analyzed separately.

A total of 86 tumors from TCGA and COSMIC met the sequencing quality criteria (see Materials and Methods) and harbored 1 of 24 *POLE* ED variants listed in Table 1 (Supplementary Table S3).

Sixteen *POLE* PVs were represented: 1 predominantly germline (L424V), 13 predominantly somatic (P286R, S297Y, S297F,

**Table 4**
MMR status and TMB of the 86 TCGA and COSMIC tumors with any of the 24 *POLE* ED PVs included in the study

| | Total | pMMR cancers | dMMR cancers |
|---|---|---|---|
| No. of tumors | 86 | 63 (73%) | 23 (27%) |
| TMB (median; range) | 160 mut/Mb (2.6-938) | 129 mut/Mb (2.6-938) | 228 mut/Mb (109-532) |
| Hypermutated (10-99 mut/Mb) | 23/86 (27%) | 23/63 (37%) | 0/23 (0%) |
| Ultramutated (≥100 mut/Mb) | 60/86 (70%) | 38/63 (60%) | 23/23 (100%) |
| MUTATIONAL SIGNATURES | SBS10a: 23% (0%-49%) | SBS10a: 30% (0%-49%) | SBS10a: 0% (0%-28%) |
| Contribution of SBS10a, 10b, 28, and 14 (median; range) | SBS10b: 34.5% (0%-65%) | SBS10b: 39% (0%-65%) | SBS10b: 10% (0%-52%) |
| | SBS28: 0% (0%-19%) | SBS28: 0% (0%-19%) | SBS28: 0% (0%-0%) |
| | SBS14: 0% (0%-38%) | SBS14: 0% (0%-29%) | SBS14: 11% (0%-38%) |
| | TOTAL: 69% (0%-87%) | TOTAL: 76% (17%-87%) | TOTAL: 29% (0%-68%) |
| No. of tumors with proofreading-deficient SBS signatures | SBS10a: 65/86 (76%) | SBS10a: 59/63 (94%) | SBS10a: 6/23 (26%) |
| | SBS10b: 77/86 (90%) | SBS10b: 62/63 (98%) | SBS10b: 15/23 (65%) |
| | SBS28: 18/86 (21%) | SBS28: 18/63 (29%) | SBS28: 0/23 (0%) |
| | SBS14: 27/86 (31%) | SBS14: 9/63 (14%) | SBS14: 17/23 (74%) |
| | Any of the 4: 85/86 (99%) | Any of the 4: 63/63 (100%) | Any of the 4: 22/23 (96%) |

Each tumor's characteristics are detailed in Supplementary Table S3.
dMMR, mismatch repair deficiency; mut/Mb, mutations per megabase; pMMR, mismatch repair proficiency; PV, pathogenic variant; SBS, single base substitution; TMB, tumor mutation burden.

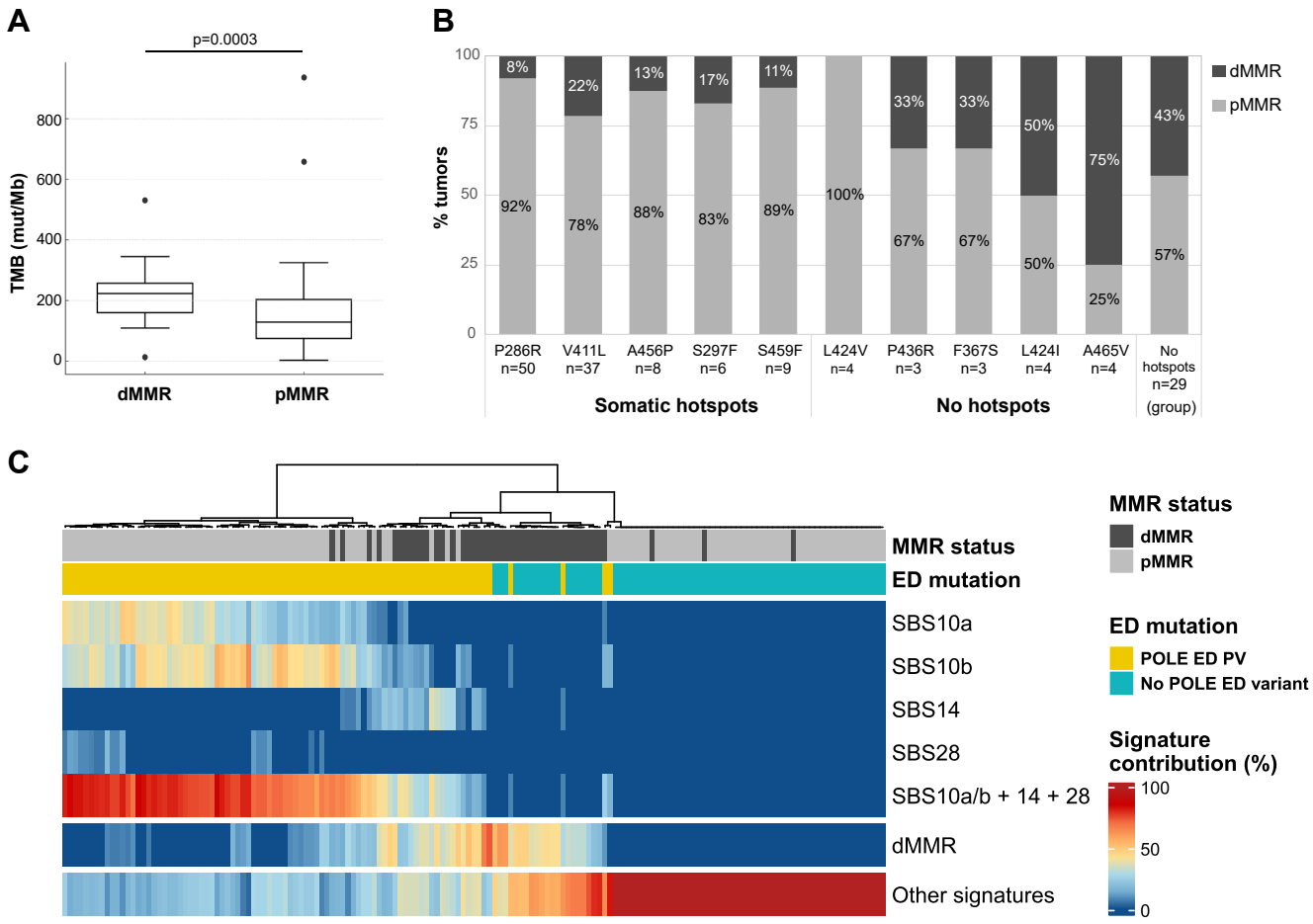


**Figure 4.**
(A) TMB (mut/Mb) of 63 pMMR and 23 dMMR Pol ε proofreading-deficient tumors. Lower and upper whiskers of the boxplots represent percentile 10 and 90, respectively. Bottom and top edges of the box represent the first and the third quartile, respectively, and the central line, the median. (B) Percentage of pMMR and dMMR tumors for each *POLE* ED somatic driver variant. Variants with MMR status information from at least 3 tumors are included in the graph. Variants are classified as somatic hotspots following the criteria described in this article (present in ≥10 tumors). The "no hotspots" group includes all tumors with MMR status information and a *POLE* somatic variant that is not considered a somatic hotspot. (C) Heatmap representing the unsupervised hierarchical clustering analysis of 86 tumors with *POLE* ED PVs and 71 tumors without *POLE* ED variants, using the Pol ε proofreading deficiency-associated mutational signatures SBS10a, SBS10b, SBS14, and SBS28, MMR deficiency signatures grouped together (SBS6, SBS15, SBS21, SBS26, and SBS44), and "other signatures" that includes any other COSMIC v3 SBS assigned and "unassigned" mutational signatures. ED, exonuclease domain; dMMR, MMR deficient; MMR, DNA mismatch repair; pMMR, MMR proficient; PV, pathogenic variant.

N363D, F367S, D368Y, V411L, L424I, P436R, A456P, S459F, S461L, and A465V), and 2 present in both germline and somatic contexts (P286L and M444K).

Among the 86 *POLE*-mutated tumors, 61 were pMMR, 21 were dMMR, and 4 had unknown MMR status. Samples with undefined MMR status were categorized based on dMMR-associated mutational signatures (>20% contribution), resulting in 63 pMMR and 23 dMMR proofreading-deficient tumors (Table 4). Eighty-two tumors harbored *POLE* ED PVs categorized as predominantly somatic, whereas 2 tumors (both pMMR) had somatic *POLE* L424V, categorized as predominantly germline. Two tumors carried *POLE* ED variants (P286L and M444K) found in both germline and somatic cases.

As already mentioned, 73% (63/86) of Pol ε proofreading-deficient tumors were pMMR, whereas 27% (23/86) were dMMR. When stratified by tumor type, 26% (14/53) of endometrial cancers, 16% (3/19) of CRCs, and 43% (6/14) of other cancer types exhibited dMMR. However, these differences did not reach statistical significance (Fisher exact test results: endometrial vs CRC, $P = .531$; endometrial vs other, $P = .325$; CRC vs other, $P = .122$).

Consistent with prior (non-TCGA) studies,[27] dMMR cancers exhibited significantly higher TMB than their pMMR counterparts (median TMB: 228 mut/Mb vs 129 mut/Mb; $P = .0003$, Mann-Whitney U test; Table 4; Fig. 4A).

MMR status analysis stratified by variant revealed that most tumors with recurrent somatic driver variants (hotspots) were pMMR (78%-92%), supporting the hypothesis that these variants alone are sufficient to drive hypermutation. However, preliminary data suggest that certain somatic *POLE* ED variants—such as A465V and L424I—may require co-occurring MMR deficiency to drive hypermutation (Fig. 4B). Notably, 2 of 4 tumors with somatic L424I were dMMR (with no other POLE ED driver), whereas the other 2 were pMMR but harbored a second somatic *POLE* ED driver (F367S or P286R). For *POLE* A465V, 3 tumors were dMMR (with no additional *POLE* ED drivers), whereas only 1 was pMMR.

*Tumor Molecular Features: Mutational Signatures*

COSMIC mutational signatures linked to Pol ε proofreading deficiency were detected in 99% (85/86) of tumors at contributions >5% (range, 8%-87%). SBS10b and SBS10a were the most prevalent, present in 90% (77/86) of tumors. Stratified by MMR status, 98% (62/63) of pMMR and 65% (15/23) of dMMR tumors had SBS10a and/or SBS10b. SBS14 was found in 14% (9/63) of pMMR and 74% (17/23) of dMMR tumors, whereas SBS28—despite its specificity for Pol ε proofreading deficiency—was detected in only 21% (18/86) of tumors, all pMMR (Table 4; Supplementary Table S3).

To assess the specificity of *POLE*-associated signatures, we performed unsupervised hierarchical clustering incorporating SBS10a, SBS10b, SBS14, SBS28, MMR-associated signatures (SBS6, SBS15, SBS21, SBS26, and SBS44), and other signatures as a group (Fig. 4C). The analysis included 86 *POLE* ED-mutated tumors (63 pMMR and 23 dMMR) and 71 randomly selected TCGA tumors without *POLE/POLD1* ED variants (gastric, colorectal, and endometrial cancers; 49 pMMR, 22 dMMR) (Supplementary Table S3). As previously observed,[21] SBS10a, SBS10b, SBS28, and SBS14 were highly specific (100%) for Pol ε proofreading deficiency. The unsupervised clustering effectively separated proofreading-deficient from dMMR tumors, although 4 *POLE*-mutant tumors overlapped with dMMR *POLE*-wildtype cases, likely due to a high contribution of dMMR and/or other signatures. However, these 4 tumors still exhibited *POLE*-associated signatures, distinguishing them from *POLE*-wildtype cases.

A direct comparison of tumors with predominantly somatic vs germline variants was not feasible due to limited cases: only 2 tumors harbored a predominantly germline variant (L424V) as a somatic event. Notably, these tumors exhibited some of the lowest TMBs in the series (2.6 and 4.4 mut/Mb) but retained clear Pol ε proofreading deficiency-associated signatures: SBS10a (0% and 23%) and SBS10b (17% and 26%).

*Recessive Effect of Polymerase δ in Hereditary Cases Versus Co-Occurrence of a POLD1 Exonuclease Domain Somatic Driver and Mismatch Repair Deficiency in Sporadic Tumors*

Compared with *POLE*, somatic *POLD1* ED PVs are rare. In TCGA and the COSMIC data sets, only 6 tumors (5 endometrial cancers and 1 CRC) harbored a somatic *POLD1* ED driver. One dMMR endometrial cancer had a co-occurring *POLE* V411L and displayed *POLE*-associated mutational signatures. The pMMR tumor, a CRC listed in COSMIC as low-MSI, lacked sequencing data for mutational signature analysis. The remaining 4 tumors (all dMMR) showed >80% contribution of SBS20, linked to combined MMR and Pol δ proofreading deficiencies (Supplementary Table S2). To validate this, we analyzed the Memorial Sloan Kettering Clinicogenomic Harmonized Oncologic Real-world Dataset (MSK-CHORD) via cBioPortal (https://www.cbioportal.org/),[5,60,61] which included 6 tumors with *POLD1* somatic drivers (3 D402N, 2 S478N, and 1 D316N), 5 of which were dMMR, supporting TCGA and the COSMIC findings.

In contrast, most tumors from *POLD1* ED gPV heterozygotes are pMMR (13/16, Supplementary Table S5). We previously demonstrated that *POLD1* gPVs require a somatic second hit (copy-neutral loss of heterozygosity [LOH]) to induce hypermutagenesis and carcinogenesis.[43] This mechanism has been observed, to date, in all analyzed tumors from *POLD1* ED gPV heterozygotes,[43,62] including 1 adenoma, 1 endometrial cancer, and 2 CRCs (Supplementary Table S5). These tumors consistently exhibited the Pol δ proofreading deficiency-associated signature SBS10d.[43]

## Discussion

This study reveals key distinctions between somatic and germline ED variants affecting Pol ε and Pol δ polymerase proofreading. Predominantly somatic variants seem to be more aggressive and mutagenic, with gPVs in these sites often leading to severe, early-onset phenotypes (CMMRD-like). Variants with extreme mutagenic potential, if they occur in the germline, may be incompatible with life, and if/when viable, these exhibit high penetrance, increasing cancer risk at an early age. On the other hand, differences in tumor phenotypes and penetrance might occur between *POLE* and *POLD1* ED gPVs.

Recurrent gPVs, such as *POLE* L424V and *POLD1* S478N, show lower REVEL pathogenicity scores and may be associated with lower TMB. The lower mutagenic potential of these variants suggests negative selection does not occur, which aligns with their higher frequency in germline cases. An exception is *POLD1* L474P, which has a high REVEL pathogenicity score despite being the second most recurrent *POLD1* gPV. However, this is a regional founder variant,[63] rather than a germline mutational hotspot. Comparing hereditary and sporadic tumors is challenging due to variability in tissue conservation, tumor types, stages, and MMR status. Future studies using controlled in vitro systems could provide clearer insights into the relative mutagenic strength of somatic and germline PVs.

The cooperation between polymerase proofreading deficiency and MMR deficiency in tumorigenesis is well supported, based on available tumor data and the existence of CMMRD-like cases caused by the co-occurrence of *POLE* or *POLD1* and MMR gene gPVs. Tumors harboring both deficiencies exhibit significantly higher TMBs and distinct mutational spectra, with SBS14 enrichment in *POLE*-mutated tumors and SBS20 in *POLD1*-mutated tumors. Some *POLE* variants, such as L424I and A465V, probably with less mutagenic potential, may require concurrent MMR deficiency or other DNA repair defects to induce hypermutation and drive tumor formation.

Although Pol ε and Pol δ are both replicative polymerases with exonuclease proofreading capabilities, *POLE* and *POLD1* differ in their tumorigenic mechanisms. *POLE* ED PVs are typically haploinsufficient, meaning a single allele alteration is sufficient to cause hypermutability. In contrast, *POLD1* ED PVs are haplosufficient, necessitating either a second hit—typically LOH—or MMR deficiency to drive mutagenesis.[26,43,64] *POLD1* gPVs predominantly follow a second hit model (LOH), whereas somatic *POLD1* PVs frequently co-occur with MMR deficiency.

Pol δ's extrinsic proofreading function provides a possible explanation for these differences. Although Pol ε primarily proofreads its own replication errors, Pol δ can also proofread errors from other replicative polymerases.[43] This extrinsic proofreading, combined with efficient MMR on the lagging strand, where Pol δ exerts its replicative activity,[65] likely mitigates the mutagenic impact of heterozygous *POLD1* variants. Consequently, *POLD1* ED heterozygous variants alone do not significantly elevate error rates, whereas MMR deficiency is necessary to prevent error correction on the lagging strand.[56]

It remains unclear why somatic *POLE* PVs are primarily found in endometrial cancers (accounting for 75% proofreading-

deficient tumors in TCGA), whereas *POLE/POLD1* gPVs are predominantly associated with increased risk of CRC (with 56% of reported gPV carriers affected with this tumor type; Fig. 3A, B). One possible explanation is ascertainment bias—germline testing of *POLE* and *POLD1* is often guided by the presence of CRC and/or polyposis, potentially underestimating endometrial cancer risk in heterozygous carriers. Additionally, the variable extent of endometrial cancer detection in these individuals may further skew the observed distribution toward CRC. On the other hand, somatic *POLE* testing is now widely implemented across many clinical centers as a biomarker for immunotherapy eligibility in endometrial cancer, which should, in theory, facilitate the identification of hereditary cancer cases enriched in endometrial cancer. However, this has not been the case, suggesting the presence of biological, cellular, or environmental factors that could be influencing this tissue specificity. Understanding the mechanisms underlying this tissue preference could offer insights into tumorigenesis in *POLE/POLD1*-associated cancers and aid in the development of prevention strategies.

Our analysis has identified, for the first time, statistically significant differences in tumor type distributions among different somatic *POLE* ED hotspot variants (Fig. 3C). Nevertheless, these findings remain preliminary due to limited sample sizes, with only 11 and 21 tumors representing the variants showing the most evident differences. Until larger tumor cohorts are analyzed to validate these observations, conclusions should be drawn with caution.

Since PPAP was first described in 2013,[20] accumulating data suggest that disease penetrance and phenotypic expression vary depending on the gene (*POLE* vs *POLD1*) and the specific variant involved (Tables 2 and 3). Highly mutagenic variants result in more severe phenotypes, characterized by early-onset tumors and a broader spectrum of malignancies and nonmalignant manifestations. The presence of co-occurring germline variants in *POLE* or *POLD1* and an MMR gene should be considered, particularly in cases resembling CMMRD.

The main limitation of this study is the lack of access to high-quality sequencing data from tumors developed in PPAP patients. In most cases, sequencing has been performed on DNA extracted from formalin-fixed paraffin-embedded samples, which generally yields lower-quality data that are not directly comparable with TCGA exome data generated from fresh-frozen tumor specimens. Furthermore, most of these data sets are not publicly available for independent reanalysis. Access to high-quality sequencing data from PPAP-associated tumors would allow for more accurate and comprehensive comparisons between hereditary and sporadic polymerase proofreading-deficient cancers. In addition, validating some of our findings—derived from a limited number of tumors—will require sequencing data from a larger cohort of polymerase proofreading-deficient tumors. Finally, assessing the mutagenic potential of ED variants in a robust and reproducible manner would necessitate controlled in vitro experiments, tracking mutation accumulation over time.

This study advances our understanding of cancer risks associated with *POLE* and *POLD1* PVs, providing insights that may improve clinical management, genetic testing interpretation, and research directions. Larger, diverse clinical data sets will be crucial for validating these findings and uncovering additional genotype-phenotype correlations. Understanding the interplay between polymerase proofreading deficiencies and other DNA repair mechanisms is essential for refining diagnostic and therapeutic strategies based on specific molecular/mutational profiles. Future research integrating experimental and clinical data will enhance patient care by further elucidating the molecular features of proofreading-deficient tumors.

### *Author Contributions*

Conceptualization: L.V. and P.M.; Formal analysis: J.V.-E., R.M., S.G.-.M., V.M., P.M., and L.V.; Funding acquisition: L.V. and G.C.; Investigation: J.V.-E., P.M., and L.V.; Methodology: J.V.-E., R.M., P. M., and L.V.; Project administration: L.V.; Supervision: L.V. and P. M.; Writing-original draft: L.V., P.M., and J.V.-E.; Writing-review and editing: all authors.

### *Data Availability*

All data, including genomic, genetic, clinical, molecular, and family information, used for this study was obtained from TCGA and COSMIC data sets and from published articles. The results supporting the conclusions of this article are included within the article and its additional files.

### *Funding*

This study was funded by the Spanish Ministry of Science and Innovation (Agencia Estatal de Investigación), cofunded by FEDER funds a way to build Europe (PID2024-162582OB-I00; PID2020-112595RB-I00, AEI/10.13039/501100011033 and "Contrato predoctoral para la formación de doctores" [J.V.-E.]); Instituto de Salud Carlos III [CIBERONC CB16/12/00234; PMP22/00064, also funded by NextGenerationEU (MRR)/PRTR]; Government of Catalonia [AGAUR 2021SGR01112; CERCA Program for institutional support (IDIBELL)]. None of the study sponsors had any involvement in the study design, collection, analysis and interpretation of data, writing of the report and decision to submit the article for publication.

### *Declaration of Competing Interest*

The authors declare no conflict of interest.

### *Ethics Approval and Consent to Participate*

The study received the approval of the IDIBELL Ethics Committee (Reference number: PR252/21). Consent to participate: not applicable.

## Supplementary Material

The online version contains supplementary material available at https://doi.org/10.1016/j.modpat.2025.100843.

## References

1. Shevelev IV, Hübscher U. The 3' 5' exonucleases. *Nat Rev Mol Cell Biol*. 2002;3(5):364–376.
2. McCulloch SD, Kunkel TA. The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res*. 2008;18(1):148–161.
3. Martínez-Jiménez F, Muiños F, Sentís I, et al. A compendium of mutational cancer driver genes. *Nat Rev Cancer*. 2020;20(10):555–572.

4. Zhao S, Choi M, Overton JD, et al. Landscape of somatic single-nucleotide and copy-number mutations in uterine serous carcinoma. *Proc Natl Acad Sci USA*. 2013;110(8):2916–2921.
5. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012;2(5):401–404.
6. Shinbrot E, Henninger EE, Weinhold 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):1740–1750.
7. Shlien A, Campbell BB, de Borja R, et al. Combined hereditary and somatic mutations of replication error repair genes result in rapid onset of ultra-hypermutated cancers. *Nat Genet*. 2015;47(3):257–262.
8. Yaeger R, Chatila WK, Lipsyc MD, et al. Clinical sequencing defines the genomic landscape of metastatic colorectal cancer. *Cancer Cell*. 2018;33(1):125–136.e3.
9. Zehir A, Benayed R, Shah RH, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med*. 2017;23(6):703–713.
10. Jones S, Stransky N, McCord CL, et al. Genomic analyses of gynaecologic carcinosarcomas reveal frequent mutations in chromatin remodelling genes. *Nat Commun*. 2014;5:5006.
11. Erson-Omay EZ, Çağlayan AO, Schultz N, et al. Somatic POLE mutations cause an ultramutated giant cell high-grade glioma subtype with better prognosis. *Neuro Oncol*. 2015;17(10):1356–1364.
12. Zou Y, Liu FY, Liu H, et al. Frequent POLE1 p.S297F mutation in Chinese patients with ovarian endometrioid carcinoma. *Mutat Res*. 2014;761:49–52.
13. Kothari N, Teer JK, Abbott AM, et al. Increased incidence of FBXW7 and POLE proofreading domain mutations in young adult colorectal cancers. *Cancer*. 2016;122(18):2828–2835.
14. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*. 2012;487(7407):330–337.
15. Domingo E, Freeman-Mills L, Rayner E, et al. Somatic POLE proofreading domain mutation, immune response, and prognosis in colorectal cancer: a retrospective, pooled biomarker study. *Lancet Gastroenterol Hepatol*. 2016;1(3):207–216.
16. Campbell BB, Light N, Fabrizio D, et al. Comprehensive analysis of hypermutation in human cancer. *Cell*. 2017;171(5):1042–1056.e10.
17. Billingsley CC, Cohn DE, Mutch DG, Stephens JA, Suarez AA, Goodfellow PJ. Polymerase ϵ (POLE) mutations in endometrial cancer: clinical outcomes and implications for Lynch syndrome testing. *Cancer*. 2015;121(3):386–394.
18. Stelloo E, Bosse T, Nout RA, et al. Refining prognosis and identifying targetable pathways for high-risk endometrial cancer; a TransPORTEC initiative. *Mod Pathol*. 2015;28(6):836–844.
19. Wong A, Kuick CH, Wong WL, et al. Mutation spectrum of POLE and POLD1 mutations in South East Asian women presenting with grade 3 endometrioid endometrial carcinomas. *Gynecol Oncol*. 2016;141(1):113–120.
20. Palles C, Cazier JB, Howarth KM, et al. Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nat Genet*. 2013;45(2):136–144.
21. Mur P, Viana-Errasti J, García-Mulero S, et al. Recommendations for the classification of germline variants in the exonuclease domain of POLE and POLD1. *Genome Med*. 2023;15(1):85.
22. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. *Nature*. 2013;500(7463):415–421.
23. Alexandrov LB, Kim J, Haradhvala NJ, et al. The repertoire of mutational signatures in human cancer. *Nature*. 2020;578(7793):94–101.
24. Church DN, Briggs SEW, Palles C, et al. DNA polymerase epsilon and delta exonuclease domain mutations in endometrial cancer. *Hum Mol Genet*. 2013;22(14):2820–2828.
25. Cancer Genome Atlas Research Network, Kandoth C, Schultz N, Cherniack AD, et al. Integrated genomic characterization of endometrial carcinoma. *Nature*. 2013;497(7447):67–73.
26. Robinson PS, Coorens THH, Palles C, et al. Increased somatic mutation burdens in normal human cells due to defective DNA polymerases. *Nat Genet*. 2021;53(10):1434–1442.
27. Shah SM, Demidova EV, Ringenbach S, et al. Exploring co-occurring POLE exonuclease and non-exonuclease domain mutations and their impact on tumor mutagenicity. *Cancer Res Commun*. 2024;4(1):213–225.
28. Hodel KP, Sun MJS, Ungerleider N, et al. POLE mutation spectra are shaped by the mutant allele identity, its abundance, and mismatch repair status. *Mol Cell*. 2020;78(6):1166–1177.e6.
29. Casanova J, Duarte GS, da Costa AG, et al. Prognosis of polymerase epsilon (POLE) mutation in high-grade endometrioid endometrial cancer: systematic review and meta-analysis. *Gynecol Oncol*. 2024;182:99–107.
30. Ambrosini M, Rousseau B, Manca P, et al. Immune checkpoint inhibitors for POLE or POLD1 proofreading-deficient metastatic colorectal cancer. *Ann Oncol*. 2024;35(7):643–655.
31. Garmezy B, Gheeya J, Lin HY, et al. Clinical and molecular characterization of *POLE* mutations as predictive biomarkers of response to immune checkpoint inhibitors in advanced cancers. *JCO Precis Oncol*. 2022;6:e2100267.
32. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17:405–423.
33. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164.
34. den Dunnen JT, Dalgleish R, Maglott DR, et al. HGVS recommendations for the description of sequence variants: 2016 update. *Hum Mutat*. 2016;37(6):564–569.
35. Ioannidis NM, Rothstein JH, Pejaver V, et al, REVEL. An ensemble method for predicting the pathogenicity of rare missense variants. *Am J Hum Genet*. 2016;99(4):877–885.
36. Cheng J, Novati G, Pan J, et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science*. 2023;381(6664):eadg7492.
37. McLaren W, Gil L, Hunt SE, et al. The ensembl variant effect predictor. *Genome Biol*. 2016;17(1):122.
38. Kautto EA, Bonneville R, Miya J, et al. Performance evaluation for rapid detection of pan-cancer microsatellite instability with MANTIS. *Oncotarget*. 2017;8(5):7452–7463.
39. Niu B, Ye K, Zhang Q, et al. MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. *Bioinformatics*. 2014;30(7):1015–1016.
40. Boland CR, Thibodeau SN, Hamilton SR, et al. A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res*. 1998;58(22):5248–5257.
41. Berg KD, Glaser CL, Thompson RE, Hamilton SR, Griffin CA, Eshleman JR. Detection of microsatellite instability by fluorescence multiplex polymerase chain reaction. *J Mol Diagn*. 2000;2(1):20–28.
42. Degasperi A, Amarante TD, Czarnecki J, et al. A practical framework and online tool for mutational signature analyses show inter-tissue variation and driver dependencies. *Nat Cancer*. 2020;1(2):249–263.
43. Andrianova MA, Seplyarskiy VB, Terradas M, et al. Discovery of recessive effect of human polymerase δ proofreading deficiency through mutational analysis of POLD1-mutated normal and cancer cells. *Eur J Hum Genet*. 2024;32(7):837–845.
44. Bernad A, Blanco L, Lázaro JM, Martín G, Salas M. A conserved 3'—–5' exonuclease active site in prokaryotic and eukaryotic DNA polymerases. *Cell*. 1989;59(1):219–228.
45. Chilkova O, Jonsson BH, Johansson E. The quaternary structure of DNA polymerase epsilon from Saccharomyces cerevisiae. *J Biol Chem*. 2003;278(16):14082–14086.
46. Mur P, García-Mulero S, Del Valle J, et al. Role of POLE and POLD1 in familial cancer. *Genet Med*. 2020;22(12):2089–2100.
47. Hamzaoui N, Alarcon F, Leulliot N, et al. Genetic, structural, and functional characterization of POLE polymerase proofreading variants allows cancer risk prediction. *Genet Med*. 2020;22(9):1533–1541.
48. Barbari SR, Shcherbakova PV. Replicative DNA polymerase defects in human cancers: consequences, mechanisms, and implications for therapy. *DNA Repair (Amst)*. 2017;56:16–25.
49. Barbari SR, Kane DP, Moore EA, Shcherbakova PV. Functional analysis of cancer-associated DNA polymerase ε variants in *Saccharomyces cerevisiae*. *G3 (Bethesda)*. 2018;8(3):1019–1029.
50. Sehested A, Meade J, Scheie D, et al. Constitutional POLE variants causing a phenotype reminiscent of constitutional mismatch repair deficiency. *Hum Mutat*. 2022;43(1):85–96.
51. Wimmer K, Beilken A, Nustede R, et al. A novel germline POLE mutation causes an early onset cancer prone syndrome mimicking constitutional mismatch repair deficiency. *Fam Cancer*. 2017;16(1):67–71.
52. Galati MA, Hodel KP, Gams MS, et al. Cancers from Novel Pole-mutant mouse models provide insights into polymerase-mediated hypermutagenesis and immune checkpoint blockade. *Cancer Res*. 2020;80(24):5606–5618.
53. Shuen AY, Lanni S, Panigrahi GB, et al. Functional repair assay for the diagnosis of constitutional mismatch repair deficiency from non-neoplastic tissue. *J Clin Oncol*. 2019;37(6):461–470.
54. Lindsay H, Scollon S, Reuther J, et al. Germline POLE mutation in a child with hypermutated medulloblastoma and features of constitutional mismatch repair deficiency. *Cold Spring Harb Mol Case Stud*. 2019;5(5):a004499.
55. Michaeli O, Ladany H, Erez A, et al. Di-genic inheritance of germline POLE and PMS2 pathogenic variants causes a unique condition associated with pediatric cancer predisposition. *Clin Genet*. 2022;101(4):442–447.
56. Schamschula E, Kinzel M, Wernstedt A, et al. Teenage-onset colorectal cancers in a digenic cancer predisposition syndrome provide clues for the interaction between mismatch repair and polymerase δ - proofreading deficiency in tumorigenesis. *Biomolecules*. 2022;12(10):1350.
57. Chang SC, Lan YT, Lin PC, et al. Patterns of germline and somatic mutations in 16 genes associated with mismatch repair function or containing tandem repeat sequences. *Cancer Med*. 2020;9(2):476–486.

58. Suerink M, Wimmer K, Brugieres L, et al. Report of the fifth meeting of the European Consortium 'Care for CMMRD' (C4CMMRD), Leiden, The Netherlands, July 6th 2019. *Fam Cancer*. 2021;20(1):67–73.
59. Wu Q, Zhang N, Xie X. The clinicopathological characteristics of POLE-mutated/ultramutated endometrial carcinoma and prognostic value of POLE status: a meta-analysis based on 49 articles incorporating 12,120 patients. *BMC Cancer*. 2022;22(1):1157.
60. Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013;6(269):pl1.
61. de Bruijn I, Kundra R, Mastrogiacomo B, et al. Analysis and visualization of longitudinal genomic and clinical data from the AACR project GENIE biopharma collaborative in cBioPortal. *Cancer Res*. 2023;83(23):3861–3867.
62. Wei CH, Wang EW, Ma L, et al. POLD1 DEDD motif mutation confers hypermutation in endometrial cancer and durable response to pembrolizumab. *Cancers (Basel)*. 2023;15(23):5674.
63. Ferrer-Avargues R, Díez-Obrero V, Martín-Tomás E, et al. Characterization of a novel POLD1 missense founder mutation in a Spanish population. *J Gene Med*. 2017;19(4).
64. Haradhvala NJ, Polak P, Stojanov P, et al. Mutational strand asymmetries in cancer genomes reveal mechanisms of DNA damage and repair. *Cell*. 2016;164(3):538–549.
65. Andrianova MA, Bazykin GA, Nikolaev SI, Seplyarskiy VB. Human mismatch repair system balances mutation rates between strands by removing more mismatches from the lagging strand. *Genome Res*. 2017;27(8):1336–1343.