POLE and POLD1 pathogenic variants in proofreading domain in papillary thyroid cancer

Running title: POLE and POLD1 mutations in PTC

Keywords: papillary thyroid cancer, IHC, pathogenic variants, POLE, POLD1

Word count: 3177

Abdul K. Siraj\textsuperscript{a}\textsuperscript{*}, Rong Bu\textsuperscript{a}\textsuperscript{*}, Maham Arshad\textsuperscript{a}, Kaleem Iqbal\textsuperscript{a}, Sandeep Kumar Parvathareddy\textsuperscript{a}, Tariq Masoodi\textsuperscript{b}, Laila Omar Ghazwani\textsuperscript{a}, Saif S. Al-Sobhi\textsuperscript{b}, Fouad Al-Dayel\textsuperscript{c} and Khawla S. Al-Kuraya\textsuperscript{a}

*authors involved equally

\textsuperscript{a}Human Cancer Genomic Research, Research Centre, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia.
\textsuperscript{b}Department of Surgery, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia
\textsuperscript{c}Department of Pathology and Laboratory Medicine, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia

Corresponding author:

Khawla S. Al-Kuraya, MD, FCAP
Human Cancer Genomic Research, Research Center
King Faisal Specialist Hospital and Research Cancer
MBC#98-16, P.O. Box 3354
Riyadh 11211, Saudi Arabia
Tel; (966)-1-205-5167
Fax (966)-1-205-5170
Email: kkuraya@kfshrc.edu.sa
Abstract

Thyroid cancer is the most frequent endocrine cancer with an increasing incidence rate worldwide and is the second most common malignancy among females in Saudi Arabia. Papillary thyroid cancer (PTC) is the most common subtype. Germline pathogenic variants in the proofreading domain of the POLE and POLD1 genes predispose to several types of cancers. However, the role of pathogenic variants of these two genes in PTC remains unknown. Capture sequencing, Sanger sequencing and Immunohistochemistry were performed on 300 PTC cases from the Middle Eastern region. One germline pathogenic variant each of POLE (1/300, 0.33%) and POLD1 (1/300, 0.33%) genes was identified. Low expression of POLD1 was detected in 46.5% (133/286) of cases and was significantly associated with the follicular variant of PTC (p=0.0006), distant metastasis (p=0.0033) and stage IV tumors (p=0.0081). However, no somatic pathogenic variant was detected in POLE gene. Furthermore, low expression of POLE was noted in 61.7% (175/284) of cases with no significant clinico-pathological associations. Our study shows that pathogenic variant in the POLE and POLD1 proofreading domain is a cause of PTC and low expression of POLD1 is associated with poor prognostic markers in the Middle Eastern population. Further studies from different geographic populations are needed to determine the frequency and spectrum of proofreading domain pathogenic variants in POLE and POLD1 genes and in PTC from different ethnicities.
Introduction

Thyroid cancer is the most frequently occurring endocrine cancer with an increasing incidence rate worldwide and with a frequency among the younger age group (1,2). It is also the second most common malignancy, after breast cancer, among females in Saudi Arabia, with a significantly higher incidence rate when compared to other countries such as the USA (2,3). Amongst the different subtypes, papillary thyroid cancer (PTC) is the most common and accounts for 80%–90% of all thyroid cancers (4). PTC is well differentiated and curable with thyroidectomy (often followed by radioiodine treatment); however, the clinical behaviour of PTC varies widely and many patients suffer disease recurrence (5,6). Although differentiated PTC is usually sporadic, cases of the familial form that is thought to be more aggressive in nature have also been reported (7,8).

Faithful DNA replication and error control are imperative for the maintenance of genomic stability and repression of mutagenesis (9). The acquisition and accumulation of deleterious mutations can trigger carcinogenesis; therefore, a rigorous system to regulate DNA synthesis is essential. There are multiple mechanisms in place to ensure accurate replication of DNA; these include: highly selective integration of nucleotides in the daughter strand; exonucleolytic proofreading to remove mismatched nucleotides; and post-replication mismatch repair (MMR) (10,11). The fidelity of eukaryotic DNA replication is primarily governed by the DNA polymerases ε (Pol ε) and δ (Pol δ); the catalytic and proofreading domains of which are encoded by POLE and POLD1, respectively (10). The leading strand of the replication fork is synthesised by Pol ε whilst Okazaki fragments of the lagging strand are synthesized by Pol δ (in tandem with DNA polymerase α). Both polymerases have an error rate of $\sim10^{-4}$–$10^{-5}$ which is further reduced to $\sim10^{-9}$–$10^{-10}$ via proofreading at the 3′–5′ exonuclease domain as well as post replication repair by the MMR apparatus (9-12).
Given that several mutations are needed for tumorigenesis, and DNA replication is extremely meticulous, it can be postulated that mutant polymerases and erroneous DNA synthesis play a key role in cancer development. Indeed, studies have reported an increase in the rate of spontaneous mutations in yeast and mice harbouring mutations in the 3′–5′ exonuclease domain of Pol ε and Pol δ (13-15). Furthermore, inactivation of the proofreading domain by a point mutation (POLD1 Asp400Ala) has been shown to increase the rate of oncogenesis in a recessive manner in mice (16,17). Similarly, studies in humans using whole-genome and targeted sequencing have identified germline pathogenic variants in the exonuclease domain of POLE and POLD1 which predispose, with a high penetrance, to multiple and/or large colorectal adenomas, early-onset colorectal cancer (CRC) as well as endometrial cancer in female carriers (18-20). In addition, somatic pathogenic variants of POLE have also been identified in different types of cancers with various incidence previously, most commonly reported in endometrial cancer (up to 10%) (21-23), followed by CRC (3-4%) (24,25). However, among 560 PTC cases characterized by TCGA, only one case harboured POLE somatic pathogenic variant (26).

Whilst the effect of mutant POLE and POLD1 on susceptibility to colorectal cancer and endometrial cancer has been widely reported in the literature; the role of POLE and POLD1 in PTC remains unknown. Identification of genetic risk factors can help with screening and early detection, and have important implications with respect to genetic counselling and implementation of treatment guidelines.

In our study, we investigated the prevalence and clinical relevance of germline and somatic variants in POLE and POLD1 proofreading domain and their protein expression in PTC cases from the Middle Eastern region.
Materials and methods

Sample Selection

Archival samples from 300 PTC patients diagnosed at King Faisal Specialist Hospital and Research Centre (Riyadh, Saudi Arabia) between 1999 to 2013 were included in the study. Detailed clinico-pathological data were noted from case records and have been summarized in Table 1. All samples were obtained from patients with approval from the Institutional Review Board of the hospital. For the study, waiver of consent was obtained for archived paraffin tissue blocks from Research Advisory Council (RAC) under project RAC # 2110 031.

DNA Extraction

DNAs were isolated from formalin-fixed, paraffin-embedded (FFPE) PTC tumor and corresponding non-tumour tissues using Gentra DNA isolation kit (Gentra, Minneapolis, MN, USA) following the manufacturer’s recommendations as described previously (27,28).

PCR and Sanger Sequencing for detection of Germline and Somatic variants in Proofreading Domain of POLE and POLD1 genes

Initially direct sequencing of the entire coding/splicing region of the proofreading domains of POLE and POLD1 genes were performed on DNA samples from 50 cases both at somatic and germline level. Primer 3 software was used to design the primers for all coding exons and their flanking intronic sequences of proofreading domain of POLE and POLD1 genes (available upon request). PCR was performed in a total volume of 25μL using 20 ng of genomic DNA, 2.5μL 10X Taq buffer, 2.3mM MgCl2, 0.2mM dNTPs, 1 unit Taq polymerase (all reagents were from Qiagen Inc), 0.2μM of each primer, and water. The efficiency and quality of the amplified PCR products were confirmed by running the PCR products on a 2% agarose gel.
For Sanger sequencing, the PCR products were subsequently subjected to direct sequencing with BigDye terminator V 3.1 cycle sequencing reagents and analysed on an ABI 3730XL DNA analyser (Applied Biosystems, Foster City, CA). Reference sequence was downloaded from NCBI GenBank. Sequencing results were compared with the reference sequence by Mutation Surveyor V4.04 (Soft Genetics, LLC, State College, PA).

**Whole Exome sequencing of Germine and Somatic Mutations in Proofreading Domain of POLE and POLD1 genes**

Furthermore, Whole Exome Sequencing (WES) was performed on 250 PTC cases as described previously (29). The tumor and corresponding normal DNA samples with A260/A280 ratio between 1.8 and 2.0 were processed for library construction. The sequencing library was prepared by random fragmentation of the DNA, followed by 5′ and 3′ adapter ligation. Adapter-ligated fragments were then Polymerase chain reaction (PCR) amplified and gel purified. Clusters were generated by loading the library into a flow cell where fragments were captured on a lawn of surface-bound oligos complementary to the library adapters. Each fragment was then amplified into distinct, clonal clusters through bridge amplification. Raw data was generated utilizing HCS (HiSeq control software v3.3) and RTA (real-time analysis. v2.5.2).

The BCL (base calls) generated by Illumina Hiseq 4000 were converted into FASTQ files by bcl2fastq (v2.16). The sequence reads in fastq format from each sample were aligned to the reference human genome (GRCh37/hg19) using burrows-wheeler aligner (BWA) (30). BAM file generation, PCR duplicates and local realignment was performed using Picard-tools and genome analysis toolkit (GATK) (31).

The variant calling was performed by GATK, subsequently the variants were annotated by ANNOVAR (32), with dbSNP138, 1000 Genomes, ESP6500, Exome
Aggregation Consortium (ExAC), Clinvar and other genome databases. The variants detected by WES were further confirmed by Sanger sequencing.

**Assessment of Pathogenicity of Variants**

ACMG/AMP 2015 guideline were utilized first for interpretation of sequence variants. All the uncertain significant variants by ACMG/AMP 2015 guideline were further analysed using five *in silico* pathogenicity prediction tools: CADD, MutationAssessor, PolyPhen-2, SIFT and Mutation Taster (33). The variants predicted as damaging or possibly damaging by three or more *in silico* prediction tools were classified as potentially pathogenic variant and further investigated.

**Tissue microarray construction & Immunohistochemistry**

All samples were analysed in a tissue microarray (TMA) format. TMA construction was performed as described earlier (34). Briefly, tissue cylinders with a diameter of 0.6 mm were punched from representative tumour regions of each donor tissue block and brought into a recipient paraffin block using a modified semiautomatic robotic precision instrument (Beecher Instruments, Woodland, WI). Two cores of PTC were arrayed from each case.

Standard protocol was followed for IHC staining. For antigen retrieval, Dako (Dako Denmark A/S, Glostrup, Denmark) Target Retrieval Solution pH 9.0 (Catalog number S2367) was used, and the slides were placed in a Pascal pressure cooker for 8 minutes at 1200°C. The slides were then incubated with primary antibodies against POLE (ab-134941, Abcam, Cambridge, UK) and POLD1 (ab-186407, Abcam, Cambridge, UK) at a dilution of 1:1000 (pH 9.0). The Dako Envision Plus System kit was used as the secondary detection system with 3, 30-diaminobenzidine as chromogen. All slides were counterstained with haematoxylin, dehydrated, cleared and mounted. Negative controls included omission of the primary antibody. Normal tissues of different organ system were also included in the TMA to
serve as control. An array of normal thyroid tissue (n = 242) was also stained. Only fresh cut slides were stained simultaneously to minimize the influence of slide aging and maximize reproducibility of the experiment.

Scoring and interpretation was performed as described previously (35). Briefly, each TMA spot was assigned an intensity score from 0–3 (I0–I3) corresponding to no, weak, moderate and strong staining, and the proportion of tumour staining for that intensity was recorded as 5% increments from a range of 0–100 (P0–P3). A final H score (range, 0–300) was obtained by adding the products of scores obtained for each intensity and proportion of area stained (H score = I1XP1 + I2xP2 + I3xP3). Using X-tile version 3.6.1 (36), we defined the optimal cut-off point for POLE and POLD1 expression as H=130 and H=125, respectively. Based on H scores, PTC cases were classified into two subgroups: those below the cut-off score were defined as low expression, whereas those above the cut-off score were defined as over expression.

Staining and evaluation of mismatch repair proteins (MLH1, MSH2, MSH6 and PMS2) was performed as described previously (37).

**Statistical Analysis**

Contingency table analysis and Chi square tests were used to study the relationship between clinico-pathological variables and protein expression. Disease free survival curves were generated using the Kaplan–Meier method, with significance evaluated using the Mantel-Cox log-rank test. The limit of significance for all analyses was defined as p value of < 0.05; two-sided tests were used in these calculations. The JMP11.0 (SAS Institute, Inc., Cary, NC) software package was used for data analyses.
Results

Sample characteristics

A total of 300 unselected PTC cases were included in the study cohort. Majority of the patients were females (77.7%) and had the classical variant of PTC (48.4%). 68.3% of patients had stage I tumour and 4.3% of patients had distant metastasis (Table 1).

Identification of Germline and Somatic variants in Proofreading Domain of POLE and POLD1 genes

Initially we performed Sanger sequencing of 50 PTC samples to look for germline and somatic mutations in proofreading domain of POLE and POLD1 genes. We found one germline variant (c.1370C>T: p. Thr457Met) of the POLE gene which was interpreted as of uncertain significance by ACMG/AMP 2015 guideline. Further analysis showed that this variant was potentially pathogenic as predicted by three out of five in-silico pathogenicity prediction tools. To further analyse the frequency of POLE and POLD1 pathogenic variants in PTC, we expanded our study to include another 250 PTC cases, which were analysed by WES. In this cohort of 250 cases, one germline variant (c.1006C>T:p. Gln336Ter) of POLD1 gene was detected and interpreted as pathogenic by ACMG/AMP 2015 guideline. However, no pathogenic variants were identified at somatic level in our cohort. Altogether, two germline variants (0.66%), one pathogenic and one potentially pathogenic, were identified in the entire cohort of 300 PTC cases.

The POLD1 p. Gln336Ter heterozygous pathogenic variant was detected in a 36-year-old female who had a mixed classical and follicular PTC. The tumor was stage I with no extra-thyroidal extension. The POLE p. Thr457Met heterozygous variant was identified in an early-onset PTC patient (diagnosed at age of 18 years). The patient with POLE p. Thr457Met variant had stage I classical variant of PTC. This variant is conserved in four out of six
species and absent in the population database of ExAC. (Table 2). Both the cases were MMR proficient by immunohistochemistry.

Since germline mutations in the proofreading domains of POLE and POLD1 are associated with an increased risk of colorectal cancer, endometrial cancer and giant cell glioblastoma (38-40), we looked for the co-existence of these malignancies in our cohort. One patient had developed rectal cancer prior to PTC, whereas another patient had endometrial cancer before developing PTC. However, neither of these patients harbored pathogenic variants of POLE or POLD1.

**POLE and POLD1 expression in PTC and their association with clinicopathological features**

Next, we evaluated the expression of POLE and POLD1 by immunohistochemistry in 300 PTC cases and 242 normal thyroid tissues using tissue microarray. The tumor tissues showed a significantly lower expression of POLE and POLD1 proteins compared to normal thyroid tissue (Figure 1). POLE immunohistochemical expression was interpretable in 284 cases. Low expression of POLE was noted in 61.7% (175/284) of cases (Figure 2A & B). No significant clinico-pathological associations were noted (Table 3). The single POLE mutant case also showed low expression of POLE by IHC. POLD1 expression was interpretable in 286 cases. Low expression of POLD1 was noted in 46.5% (133/286) of cases (Figure 2C & D) and was significantly associated with the follicular variant of PTC (p=0.0006), distant metastasis (p=0.0033) and stage IV tumors (p=0.0081) (Table 4). The POLD1 mutant case showed over-expression of POLD1 protein. Both POLE and POLD1 protein expression were not associated with disease-free survival (p = 0.1590 and 0.5197, respectively; Figure 3)
Discussion

During the past decades there has been a significant increase in incidence of PTC worldwide (1-4). Hence, further research is necessary to identify the cellular and genetic changes which trigger oncogenesis in order to administer appropriate targeted therapies as well as improve the prevention and assessment of cancer risk. Knowledge of mutation distribution can also be used to plan cost-effective genetic testing strategies on a national scale.

Pathogenic variants of the proofreading domains in POLE and POLD1 genes are widely known to be associated with different cancers including endometrial cancer, colorectal cancer, lung cancer, and glioblastoma (19,23,25,40-45). However, there are very limited current literature regarding these variants in PTC, with TCGA reporting only a single POLE somatic pathogenic variant among 560 cases (46). To the best of our knowledge, we are the first to report the frequency of proofreading domain germline pathogenic variant of the POLE and POLD1 genes in PTC cases. In our cohort, one pathogenic variant in proofreading domain of POLD1 and one potentially pathogenic variant in proofreading domain of POLE were identified, accounting for 0.66% of all PTC cases. Interestingly, these two variants were not reported previously. The variant of POLD1 p. Gln336Ter caused a stop gain and was identified in an early onset patient, indicating a functional loss and predisposition to PTC. Although the POLE p.Thr457Met variant was classified as uncertain significance by ACMG, it was partially conserved and completely absent in the population database of ExAC. Furthermore, the carrier of this variant developed PTC at a very early age of 18 years. The early onset of PTC suggests that POLE p.Thr457Met might have strong cause-effect on pathogenesis of PTC. Interestingly, this variant was also identified in one CRC patient previously by us (47). Our data indicates that this variant might predispose to not only PTC but also CRC. However, further functional analysis is required to investigate the effect of this
variant on the function of the *POLE* gene. In our cohort, we did not identify any germline pathogenic variants which were common in Western populations. These data imply the unique spectrum of proofreading domain pathogenic variants of *POLE* and *POLD1* in our population. Despite the large number of cases screened, no somatic pathogenic variants were identified in our cohort, indicating that somatic proofreading domain pathogenic variants of *POLE* and *POLD1* might not play important role in development of sporadic PTC in our population.

We have previously studied the expression of *POLE* and *POLD1* in endometrial carcinomas and their clinico-pathological associations (35). However, to the best of our knowledge, this is the first study to report the clinico-pathological associations of *POLE* and *POLD1* protein expression in PTC. Low expression of *POLD1* was associated with poorer prognostic markers such as distant metastasis and stage IV tumors. We also found low expression of *POLE* and *POLD1* to be associated with poor prognostic factors such as lymph node involvement, grade 3 tumors and stage III tumors in CRC (In press). However, no significant association with disease free survival was noted with both biomarkers. Interestingly, while the case harbouring *POLE* p.Thr457Met showed a low expression of *POLE* protein, another case harbouring the pathogenic variant of *POLD1* p. Gln336Ter showed high expression of *POLD1*. This discrepancy between protein expression and mutation has been highlighted previously by us (35) and others (48) in different organ sites. It might be partly explained by the fact that only pathogenic variants in proofreading region were investigated. In addition, a previous study also demonstrated that there were more truncated variants detected in the region outside of proofreading domain in *POLE* and *POLD1* genes (49) and these truncated variants could also cause loss of *POLE* or *POLD1* expression. Further studies are warranted to validate our findings.
Conclusions

In conclusion, this study presents data on the frequency of POLE and POLD1 proofreading domain somatic and germline pathogenic variants and association of POLE and POLD1 expression with prognostic biomarkers in a cohort of patients with PTC; more specifically, in the Middle East. Our data suggests that pathogenic variants in proofreading domain of POLE and POLD1 is a cause of PTC and low expression of POLD1 is found to be associated with poor prognostic markers. Nonetheless, it may be of interest to include these two genes in the gene panel for future studies.

Declaration of Interest

There is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Funding

This research did not receive any specific grant from any funding agency in the public, commercial or not-for-profit sector.
References

1. Gonzalez-Gonzalez R, Bologna-Molina R, Carreon-Burciaga RG, Gomezpalacio-Gastelum M, Molina-Frechero N, Salazar-Rodriguez S. Papillary thyroid carcinoma: differential diagnosis and prognostic values of its different variants: review of the literature. ISRN Oncol 2011 2011 915925-915934

2. Bu R, Siraj AK, Divya SP, Kong Y, Parvathareddy SK, Al - Rasheed M, Al - Obaisi KA, Victoria IG, Al - Sobhi SS, Al - Dawish MJjoc. Telomerase reverse transcriptase mutations are independent predictor of disease - free survival in M iddle E astern papillary thyroid cancer. 2018 142 2028-2039

3. Hussain F, Iqbal S, Mehmod A, Bazrabashi S, ElHassan T, Chaudhri NJHo, therapy sc. Incidence of thyroid cancer in the Kingdom of Saudi Arabia, 2000–2010. 2013 6 58-64

4. Schneider DF, Chen H. New developments in the diagnosis and treatment of thyroid cancer. CA Cancer J Clin 2013 63 374-394

5. Mazzaferri EL, Jhiang SM. Long-term impact of initial surgical and medical therapy on papillary and follicular thyroid cancer. Am J Med 1994 97 418-428

6. Fonseca E, Soares P, Rossi S, Sobrinho-Simoes M. Prognostic factors in thyroid carcinomas. Verh Dtsch Ges Pathol 1997 81 82-96

7. Yu Y, Dong L, Li DP, Chuai SK, Wu ZG, Zheng XQ, Cheng YA, Han L, Yu JP, Gao M. Targeted DNA Sequencing Detects Mutations Related to Susceptibility among Familial Non-medullary Thyroid Cancer. Sci Rep-Uk 2015 5 16129-16141

8. Capezzone M, Cantara S, Marchisotta S, Filetti S, De Santi MM, Rossi B, Ronga G, Durante C, Pacini F. Short telomeres, telomerase reverse transcriptase gene amplification, and increased telomerase activity in the blood of familial papillary thyroid cancer patients. J Clin Endocrinol Metab 2008 93 3950-3957

9. Kunkel TA. DNA replication fidelity. J Biol Chem 2004 279 16895-16898

10. Preston BD, Albertson TM, Herr AJ. DNA replication fidelity and cancer. Paper presented at: Seminars in cancer biology 2010

11. Garg P, Burgers PMJ. DNA polymerases that propagate the eukaryotic DNA replication fork. Crit Rev Biochem Mol 2005 40 115-128

12. Elsayed FA, Kets CM, Ruano D, van den Akker B, Mensenkamp AR, Schrumpf M, Nielsen M, Wijnen JT, Tops CM, Ligtenberg MJ, et al. Germline variants in POLE are associated with early onset mismatch repair deficient colorectal cancer. European journal of human genetics : EJHG 2015 23 1080-1084

13. Simon M, Giot L, Faye GITEj. The 3’ to 5’ exonuclease activity located in the DNA polymerase delta subunit of Saccharomyces cerevisiae is required for accurate replication. 1991 10 2165-2170

14. Albertson TM, Ogawa M, Bugni JM, Hays LE, Chen Y, Wang Y, Treuting PM, Heddle JA, Goldsby RE, Preston BD. DNA polymerase epsilon and delta proofreading suppress discrete mutator and cancer phenotypes in mice. Proc Natl Acad Sci U S A 2009 106 17101-17104

15. Venkatesan RN, Hsu JJ, Lawrence NA, Preston BD, Loeb LA. Mutator phenotypes caused by substitution at a conserved motif A residue in eukaryotic DNA polymerase delta. Journal of Biological Chemistry 2006 281 4486-4494

16. Goldsby RE, Lawrence NA, Hays LE, Olmsted EA, Chen X, Singh M, Preston BD. Defective DNA polymerase-delta proofreading causes cancer susceptibility in mice. Nat Med 2001 7 638-639

17. Goldsby RE, Hays LE, Chen X, Olmsted EA, Slayton WB, Spangrude GJ, Preston BD. High incidence of epithelial cancers in mice deficient for DNA polymerase delta proofreading. Proc Natl Acad Sci U S A 2002 99 15560-15565
18. Palles C, Cazier JB, Howarth KM, Domingo E, Jones AM, Broderick P, Kemp Z, Spain SL, Guarino E, Salguero I, et al. Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. Nat Genet 2013 45 136-144
19. Church DN, Briggs SE, Palles C, Domingo E, Kearsey SJ, Grimes JM, Gorman M, Martin L, Howarth KM, Hodgson SV, et al. DNA polymerase epsilon and delta exonuclease domain mutations in endometrial cancer. Hum Mol Genet 2013 22 2820-2828
20. Imboden S, Nastic D, Ghaderi M, Rydberg F, Rau TT, Mueller MD, Epstein E, Carlson JW. Phenotype of POLE-mutated endometrial cancer. PloS one 2019 14 e0214318
21. Mo S, Li Y, Peng J. Somatic POLE proofreading domain mutations by next generation sequencing to predict outcomes in stage II colorectal cancer. American Society of Clinical Oncology; 2019 559
22. Agrawal N, Akbani R, Aksoy BA, Ally A, Arachchi H, Asa SL, Auman JT, Balasundaram M, Balu S, Baylin SB. Integrated genomic characterization of papillary thyroid carcinoma. Cell 2014 159 676-690
23. Khoo US, Ozcelik H, Cheung AN, Chow LW, Ngan HY, Done SJ, Liang AC, Chan VW, Au GK, Ng WF, et al. Somatic mutations in the BRCA1 gene in Chinese sporadic breast and ovarian cancer. Oncogene 1999 18 4643-4646
24. Abubaker J, Jehan Z, Bavi P, Sultana M, Al-Harbi S, Ibrahim M, Al-Nuaim A, Ahmed M, Amin T, Al-Fehaily M, et al. Clinicopathological analysis of papillary thyroid cancer with PIK3CA alterations in a Middle Eastern population. J Clin Endocrinol Metab 2008 93 611-618
25. Siraj AK, Masoodi T, Bu R, Beg S, Al-Sobhi SS, Al-Dayel F, Al-Dawish M, Alkuraya FS, Al-Kuraya KS. Genomic profiling of thyroid cancer reveals a role for thyroglobulin in metastasis. The American Journal of Human Genetics 2016 98 1170-1180
26. Liu H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics 2010 26 589-595
27. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. Nature methods 2014 11 361-362
28. Siraj AK, Parvathareddy SK, Bu R, Iqbal K, Siraj S, Masoodi T, Concepcion RM, Ghazwani LO, AlBadawi I, Al-Dayel F. Germline POLE and POLD1 proofreading domain mutations in endometrial carcinoma from Middle Eastern region. Cancer Cell International 2019 19 334-343
29. Siraj AK, Parvathareddy SK, Bu R, Iqbal K, Siraj S, Masoodi T, Concepcion RM, Ghazwani LO, AlBadawi I, Al-Dayel F. Germline POLE and POLD1 proofreading domain mutations in endometrial carcinoma from Middle Eastern region. Cancer Cell International 2019 19 334-343
36. Camp RL, Dolled-Filhart M, Rimm DL. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. Clinical cancer research 2004 10 7252-7259

37. Siraj AK, Prabhakaran S, Bavi P, Bu R, Beg S, Hazmi MA, Al-Rasheed M, Al-Assiri M, Sairafi R, Al-Dayel F. Prevalence of Lynch syndrome in a Middle Eastern population with colorectal cancer. Cancer 2015 121 1762-1771

38. Buchanan DD, Stewart JR, Clendenning M, Rosty C, Mahmood K, Pope BJ, Jenkins MA, Hopper JL, Southey MC, Macrae FA, et al. Risk of colorectal cancer for carriers of a germ-line mutation in POLE or POLD1. Genet Med 2018 20 890-895

39. Vande Perre P, Siegfried A, Corsini C, Bonnet D, Toulas C, Hamzaoui N, Selves J, Chipoulet E, Hoffmann JS, Uro-Coste E, et al. Germline mutation p.N363K in POLE is associated with an increased risk of colorectal cancer and giant cell glioblastoma. Fam Cancer 2019 18 173-178

40. McConechy MK, Talhouk A, Leung S, Chiu D, Yang W, Senz J, Reha-Krantz LJ, Lee CH, Huntsman DG, Gilks CB, et al. Endometrial Carcinomas with POLE Exonuclease Domain Mutations Have a Favorable Prognosis. Clin Cancer Res 2016 22 2865-2873

41. Spier I, Holzapfel S, Altmuller J, Zhao B, Horpaopan S, Vogt S, Chen S, Morak M, Raeder S, Kayser K, et al. Frequency and phenotypic spectrum of germline mutations in POLE and seven other polymerase genes in 266 patients with colorectal adenomas and carcinomas. Int J Cancer 2015 137 320-331

42. Bellido F, Pineda M, Aiza G, Valdes-Mas R, Navarro M, Puente DA, Pons T, Gonzalez S, Iglesias S, Darder E, et al. POLE and POLD1 mutations in 529 kindred with familial colorectal cancer and/or polyposis: review of reported cases and recommendations for genetic testing and surveillance. Genet Med 2016 18 325-332

43. Guerra J, Pinto C, Pinto D, Pinheiro M, Silva R, Peixoto A, Rocha P, Veiga I, Santos C, Santos R. POLE somatic mutations in advanced colorectal cancer. Cancer medicine 2017 6 2966-2971

44. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, Lee W, Yuan J, Wong P, Ho TS, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science 2015 348 124-128

45. Bouffet E, Larouche V, Campbell BB, Merico D, de Borja R, Aronson M, Durno C, Krueger J, Cabric V, Ramaswamy V, et al. Immune Checkpoint Inhibition for Hypermutant Glioblastoma Multiforme Resulting From Germline Biallelic Mismatch Repair Deficiency. J Clin Oncol 2016 34 2206-2211

46. Cancer Genome Atlas Research N. Integrated genomic characterization of papillary thyroid carcinoma. Cell 2014 159 676-690

47. Siraj AK, Bu R, Iqbal K, Parvathareddy SK, Masoodi T, Siraj N, Al-Rasheed M, Kong Y, Ahmed SO, Al-Obaissi KAS, et al. POLE and POLD1 germline exonuclease domain pathogenic variants, a rare event in colorectal cancer from the Middle East. Molecular genetics & genomic medicine 2020 e1368

48. Elsayed FA, Kets CM, Ruano D, Van Den Akker B, Mensenkamp AR, Schrumpf M, Nielsen M, Wijnjen JT, Tops CM, Ligtelenberg MJ. Germline variants in POLE are associated with early onset mismatch repair deficient colorectal cancer. European Journal of Human Genetics 2015 23 1080-1084

49. Campbell BB, Light N, Fabrizio D, Zatzman M, Fuligni F, de Borja R, Davidson S, Edwards M, Elvin JA, Hodel KP, et al. Comprehensive Analysis of Hypermutation in Human Cancer. Cell 2017 171 1042-1056
Figure legends

Figure 1: Boxplot showing difference in expression between normal and tumor tissue for POLE and POLD1. (A) POLE expression was significantly higher in normal tissue compared to tumor (Mean 138.4±48.5 vs 111.1±55.4; p < 0.0001). (B) POLD1 expression was significantly higher in normal tissue compared to tumor (Mean 165.7±45.4 vs 137.0±50.7; p < 0.0001).

Figure 2: Tissue microarray based immunohistochemistry analysis of POLE and POLD1 in PTC patients. PTC TMA spots showing overexpression of POLE (A) and POLD1 (C). In contrast, another set of TMA spots showing reduced expression of POLE (B) and POLD1 (D). 20 X/0.70 objective on an Olympus BX 51 microscope. (Olympus America Inc., Centre Valley, PA, USA) with the inset showing a 40X 0.85 aperture magnified view of the same TMA spot.

Figure 3: Kaplan-Meier survival analysis for the prognostic significance of POLE and POLD1 expression in PTC. (A) There was no significant difference in disease-free survival with POLE expression (p=0.1590). (B) There was no significant difference in disease-free survival with POLD1 expression (p=0.5197).
Table 1. Clinico-pathological variables for the patient cohort (n=300)

| Variable                     | n (%)          |
|------------------------------|----------------|
| **Age**                      |                |
| Median                       | 39.0           |
| Range (IQR)^                 | 29.0 – 50.0    |
| **Gender**                   |                |
| Female                       | 233 (77.7)     |
| Male                         | 67 (22.3)      |
| **Histopathology**           |                |
| Classical Variant            | 145 (48.4)     |
| Follicular Variant           | 60 (20.0)      |
| Tall Cell Variant            | 43 (14.3)      |
| Others                       | 52 (17.3)      |
| **Extra Thyroidal Extension**|                |
| Absent                       | 172 (57.3)     |
| Present                      | 118 (39.4)     |
| Unknown                      | 10 (3.3)       |
| **pT**                       |                |
| T1                           | 89 (29.7)      |
| T2                           | 60 (20.0)      |
| T3                           | 131 (43.7)     |
| T4                           | 18 (6.0)       |
| Unknown                      | 2 (0.6)        |
| **pN**                       |                |
| N0                           | 175 (58.3)     |
| N1                           | 122 (40.7)     |
| Unknown                      | 3 (1.0)        |
| **pM**                       |                |
| M0                           | 286 (95.4)     |
| M1                           | 13 (4.3)       |
| Unknown                      | 1 (0.3)        |
| **Stage**                    |                |
| I                            | 205 (68.3)     |
| II                           | 20 (6.7)       |
| III                          | 42 (14.0)      |
| IV                           | 31 (10.4)      |
| Unknown                      | 2 (0.6)        |

Abbreviations: ^ Inter quartile range
Table 2. Characteristics of POLE p. Thr457Met identified

| Gene                        | POLE                                      |
|-----------------------------|-------------------------------------------|
| Mutation                    | c.1370C>T: p Thr457Met                   |
| PolyPhen-2                  | Probably Damaging                        |
| Mutation Assessor           | Low                                       |
| SIFT                        | Tolerated                                 |
| CADD                        | 25.1                                      |
| Mutation Taster             | Disease Causing                           |
| Family History              | Negative                                  |
| Conservation between species| 4 out of 6                                |
| Age at diagnosis            | 18                                        |
| Frequency in ExAC           | absent                                    |
Table 3: Clinico-pathologic associations of POLE protein expression in Papillary Thyroid Carcinoma

|                          | Total  | POLE Over-expression | POLE Low expression | p value |
|--------------------------|--------|-----------------------|---------------------|---------|
|                          | No.    | %                     | No.                 | %       |         |
| No. of patients          | 284    |                       | 109                 | 38.3    | 175     | 61.7    |
| Age (Yrs)                |        |                       |                     |         |         |
| ≤ 45                     | 176    | 62.2                  | 67                  | 38.1    | 109     | 61.9    | 0.8427  |
| > 45                     | 107    | 37.8                  | 42                  | 39.3    | 65      | 60.7    |
| Sex                      |        |                       |                     |         |         |
| Female                   | 222    | 78.2                  | 91                  | 41.0    | 131     | 59.0    | 0.0825  |
| Male                     | 62     | 21.8                  | 18                  | 29.0    | 44      | 71.0    |
| Extrathyroidal extension |        |                       |                     |         |         |
| Absent                   | 164    | 59.6                  | 60                  | 36.6    | 104     | 63.4    | 0.4175  |
| Present                  | 111    | 40.4                  | 46                  | 41.4    | 65      | 58.6    |
| pT                       |        |                       |                     |         |         |
| T1                       | 85     | 30.1                  | 27                  | 31.8    | 58      | 68.2    | 0.2926  |
| T2                       | 58     | 20.6                  | 24                  | 41.4    | 34      | 58.6    |
| T3                       | 122    | 43.3                  | 53                  | 43.4    | 69      | 56.6    |
| T4                       | 17     | 6.0                   | 5                   | 29.4    | 12      | 70.6    |
| pN                       |        |                       |                     |         |         |
| pN0                      | 166    | 59.1                  | 65                  | 39.2    | 101     | 60.8    | 0.7647  |
| pN1                      | 115    | 40.9                  | 43                  | 37.4    | 72      | 62.6    |
| pM                       |        |                       |                     |         |         |
| pM0                      | 272    | 96.1                  | 107                 | 39.3    | 165     | 60.7    | 0.1364  |
| pM1                      | 11     | 3.9                   | 2                   | 18.2    | 9       | 81.8    |
| Stage                    |        |                       |                     |         |         |
| I                        | 195    | 69.1                  | 73                  | 37.4    | 122     | 62.6    | 0.2346  |
| II                       | 17     | 6.0                   | 5                   | 29.4    | 12      | 70.6    |
| III                      | 40     | 14.2                  | 21                  | 52.5    | 19      | 47.5    |
| IV                       | 30     | 10.7                  | 10                  | 33.3    | 20      | 66.7    |
| Histology Type           |        |                       |                     |         |         |
| Classical variant        | 137    | 48.2                  | 49                  | 35.8    | 88      | 64.2    | 0.5692  |
| Follicular variant       | 58     | 20.5                  | 22                  | 37.9    | 36      | 62.1    |
| Tall-Cell variant        | 43     | 15.1                  | 16                  | 38.1    | 27      | 61.9    |
| Other variants           | 46     | 16.2                  | 22                  | 47.8    | 24      | 52.2    |
| Disease Free Survival    |        |                       |                     |         |         |
| 5 years                  | 84.3   |                       | 76.4                |         | 0.1590  |
Table 4: Clinico-pathologic associations of POLD1 protein expression in Papillary Thyroid Carcinoma

|                           | Total          | POLD1 Over-expression | POLD1 Low expression | p value |
|---------------------------|----------------|-----------------------|----------------------|---------|
|                           | No. | %   | No. | %   | No. | %   |       |
| No. of patients           | 286 |     | 153 | 53.5 | 133 | 46.5 |       |
| Age (Yrs)                 |      |     |     |     |     |     |       |
| ≤ 45                      | 179  | 62.8 | 102 | 57.0 | 77  | 43.0 | 0.1085 |
| > 45                      | 108  | 37.2 | 50  | 47.2 | 56  | 52.8 |         |
| Sex                       |      |     |     |     |     |     |       |
| Female                    | 223  | 78.0 | 123 | 55.2 | 100 | 44.8 | 0.2900 |
| Male                      | 63   | 22.0 | 30  | 47.6 | 33  | 52.4 |         |
| Extrathyroidal extension  |      |     |     |     |     |     |       |
| Absent                    | 165  | 57.7 | 85  | 51.5 | 80  | 48.5 | 0.2488 |
| Present                   | 111  | 42.3 | 65  | 48.6 | 46  | 51.4 |       |
| pT                        |      |     |     |     |     |     |       |
| T1                        | 85   | 29.9 | 43  | 50.6 | 42  | 49.4 | 0.1013 |
| T2                        | 59   | 20.8 | 28  | 47.5 | 31  | 52.5 |         |
| T3                        | 123  | 43.3 | 75  | 61.0 | 48  | 39.0 |         |
| T4                        | 17   | 6.0  | 6   | 35.3 | 11  | 64.7 |         |
| pN                        |      |     |     |     |     |     |       |
| pN0                       | 166  | 58.7 | 82  | 49.4 | 84  | 50.6 | 0.1112 |
| pN1                       | 117  | 41.3 | 69  | 50.6 | 48  | 41.0 |         |
| pM                        |      |     |     |     |     |     |       |
| pM0                       | 272  | 95.4 | 151 | 55.5 | 121 | 44.5 | 0.0033 |
| pM1                       | 13   | 4.6  | 2   | 15.4 | 11  | 84.6 |         |
| Stage                     |      |     |     |     |     |     |       |
| I                         | 195  | 68.6 | 109 | 55.9 | 86  | 44.1 | 0.0081 |
| II                        | 19   | 6.7  | 9   | 47.4 | 10  | 52.6 |         |
| III                       | 40   | 14.1 | 26  | 65.0 | 14  | 35.0 |         |
| IV                        | 30   | 10.6 | 8   | 26.7 | 22  | 73.3 |         |
| Histology Type            |      |     |     |     |     |     |       |
| Classical variant         | 138  | 48.3 | 75  | 54.3 | 63  | 45.7 | 0.0006 |
| Follicular variant        | 59   | 20.6 | 20  | 33.9 | 39  | 66.1 |         |
| Tall-Cell variant         | 43   | 15.0 | 32  | 74.4 | 11  | 25.6 |         |
| Other variants            | 46   | 16.1 | 26  | 56.5 | 20  | 43.5 |         |
| Disease Free Survival     |      |     |     |     |     |     |       |
| 5 years                   | 81.3 | 77.9 |     |     |     |     | 0.5197 |
Figure 1: Boxplot showing difference in expression between normal and tumor tissue for POLE and POLD1. (A) POLE expression was significantly higher in normal tissue compared to tumor (Mean 138.4±48.5 vs 111.1±55.4; p < 0.0001). (B) POLD1 expression was significantly higher in normal tissue compared to tumor (Mean 165.7±45.4 vs 137.0±50.7; p < 0.0001).
Figure 2: Tissue microarray based immunohistochemistry analysis of POLE and POLD1 in PTC patients. PTC TMA spots showing overexpression of POLE (A) and POLD1 (C). In contrast, another set of TMA spots showing reduced expression of POLE (B) and POLD1 (D). 20 X/0.70 objective on an Olympus BX 51 microscope. (Olympus America Inc., Centre Valley, PA, USA) with the inset showing a 40X 0.85 aperture magnified view of the same TMA spot.

338x190mm (96 x 96 DPI)
Figure 3: Kaplan-Meier survival analysis for the prognostic significance of POLE and POLD1 expression in PTC. (A) There was no significant difference in disease-free survival with POLE expression ($p=0.1590$). (B) There was no significant difference in disease-free survival with POLD1 expression ($p=0.5197$).