Signatures of defective DNA repair and replication in early-onset renal cancer patients referred for germline genetic testing.

Running title: Constitutional DNA repair defects in early-onset renal cancer.

Elena V. Demidova[1,2], Ilya G. Serebriiskii[2,3], Ramilia Vlasenkova[2], Simon Kelow[4], Mark D. Andrake[3], Tiffiney R. Hartman[1,5], Tatiana Kent[6], Richard T. Pomerantz[6], Roland L. Dunbrack Jr.[3], Erica A. Golemis[3,7], Michael J. Hall[1,8], David Y.T. Chen[9], Mary B. Daly*[1,8] and Sanjeevani Arora*[1,10]

[1]Cancer Prevention and Control Program, Fox Chase Cancer Center, Philadelphia PA, 19111; [2]Kazan Federal University, Russia, 420008; [3]Molecular Therapeutics Program, Fox Chase Cancer Center, Philadelphia PA, 19111; [4]Department of Biochemistry and Molecular Biophysics, University of Pennsylvania, Philadelphia, PA,19104; [5]Arcadia University, Glenside, PA; [6]Department of Biochemistry & Molecular Biology, Sidney Kimmel Cancer Center, Thomas Jefferson University, Philadelphia, PA, 19107; [7]Department of Cancer and Cellular Biology, Lewis Katz School of Medicine, Philadelphia, PA 19140; [8]Department of Clinical Genetics, Fox Chase Cancer Center, Philadelphia PA, 19111; [9]Department of Surgical Oncology, Fox Chase Cancer Center, Philadelphia PA, 19111; [10]Department of Radiation Oncology, Fox Chase Cancer Center, Philadelphia PA, 19111

Correspondence:
Sanjeevani Arora, Ph.D.,
Cancer Prevention and Control Program
Fox Chase Cancer Center
333 Cottman Avenue, P2149
Philadelphia, PA 19111-2497
Email: Sanjeevani.Arora@fccc.edu

OR
Mary B. Daly, M.D., Ph.D., F.A.C.P.,
Professor, Department of Clinical Genetics
Director, Risk Assessment Program
Timothy R. Talbot Jr. Chair in Cancer Research
Cancer Prevention and Control Program
Fox Chase Cancer Center
333 Cottman Avenue
Philadelphia, PA 19111-2497
Email: Mary.Daly@fccc.edu

Author contributions. E.V.D. performed most studies, data analysis and interpretation, figure and table generation, writing & editing; T.R.H., I.G.S. data analysis, discussion; R.V. TCGA analysis; S.K. & M.A. & R.L.D. structure modelling, discussion; T.K., R.T.P. Pol delta purification & methods; E.A.G. discussion; M.J.H. & D.Y.T.C. & M.B.D. patients, discussion; M.B.D. & S.A., discussion, designing, planning studies, writing & editing.

Disclosures/Conflict of Interest: M.J.H. performs collaborative research (with no funding) with the following: Myriad Genetics, Invitae Corporation, Ambry Genetics, Foundation Medicine, Inc. He also performs collaborative research (with no funding) and is part of a Precision Oncology Alliance funded by Caris Life Sciences (cover travel and meals at meetings). S.A. performs collaborative research (with no funding) with Caris Life Sciences, Foundation Medicine, Inc., Ambry Genetics, and Invitae Corporation. S.A.'s spouse is employed by Akoya Biosciences and has stocks in Akoya Biosciences, HTG Molecular Diagnostics, Abcam Plc., and Senzo Health. S.A., M.J.H., E.A.G., I.G.S. have patents and/or pending patents related to cancer diagnostics/treatment. All other authors declare no competing interests.

Classification: Biological Sciences, Genetics.

Keywords. early-onset cancer, renal cancer, DNA replication, DNA repair, germline.

Abstract.

Early-onset renal cell carcinoma (eoRCC) is typically associated with pathogenic germline variants (PGVs) in RCC familial syndrome genes. However, most eoRCC patients lack PGVs in familial RCC genes and their genetic risk remains undefined. Here, we analyzed biospecimens from 22 eoRCC patients that were seen at our institution for genetic counseling and tested negative for PGVs in RCC familial syndrome genes. We performed whole-exome sequencing (WES) and found enrichment of candidate pathogenic germline variants in DNA repair and replication genes, including multiple DNA polymerases. Induction of DNA damage in peripheral blood monocytes (PBMCs) significantly elevated numbers of γH2AX foci, a marker of double-stranded breaks, in PBMCs from eoRCC patients versus PBMCs from matched cancer-free controls. Knockdown of candidate PGVs in Caki RCC cells increased γH2AX foci. Immortalized patient-derived B cells bearing candidate PGVs in DNA polymerase genes (*POLD1, POLH, POLE, POLK*) had DNA replication defects compared to control cells. Renal tumors carrying these DNA polymerase variants were microsatellite stable but had a high mutational burden. Direct biochemical analysis of the variant Pol δ and Pol η polymerases revealed defective enzymatic activities. Together, these results suggest that constitutional defects in DNA repair such as DNA replication repair underlie a subset of eoRCC cases. These findings may provide opportunities for use of the DNA repair targeting agents for eoRCC treatment. Screening patient lymphocytes to identify these defects may provide insight into mechanisms of carcinogenesis in a subset of genetically undefined eoRCCs.

Significance Statement. Screening for DNA repair variation may provide a more comprehensive risk assessment for eoRCC patients. Evaluation of DNA repair defects may also provide insight into the cancer initiation mechanisms for subsets of eoRCCs and lay the foundation for targeting DNA repair vulnerabilities in eoRCC.

Introduction.

Early onset renal cell carcinoma (eoRCC) in patients under the age of 60 has been increasing in frequency over the past decade (1). In the United States alone, the most recent analyses report a range of 3.0% annual increase in RCC incidence among individuals aged 45-49 years to as high as a 6.2% increase in incidence among those aged 25-29 years (1). EoRCC is in some cases linked to pathogenic germline variants (PGVs) in genes associated with RCC familial syndromes (*VHL, MET, FLCN, TSC1, TSC2, FH, SDHx, PTEN, BAP1*) (1-3); these genes are also often somatically mutated in sporadic RCC cases (2-5). Identification of a PGV in defined RCC familial syndrome genes guides clinical recommendations for surveillance, often improving survival due to early diagnosis of eoRCC. However, in recent work we found that only ~3.7% of eoRCC patients undergoing cancer risk assessment report a PGV in the currently defined RCC familial syndrome genes (6), reflecting the fact that the majority of eoRCC cases remain genetically not well characterized. Currently, there are no National Comprehensive Cancer Network (NCCN) guidelines for detection, prevention, or risk reduction in individuals who present with an eoRCC but lack a PGV in a familial RCC gene (7).

Recently, we reported that a significant subset of eoRCC patients undergoing cancer risk assessment carry PGVs in DNA damage response and repair genes (~8.55% vs. 3.7% in familial RCC genes) (6). Similarly, Carlo *et al*. reported an increased prevalence of PGVs in DNA repair genes in advanced clear cell and non-clear cell renal cancer patients (3, 8). Although PGVs in DNA repair genes are not currently defined by clinical testing guidelines as increasing risk of RCC, these recent studies suggest a potential role of defective DNA repair pathways in eoRCC carcinogenesis that could also lead to novel therapeutic options for RCC patients. Owing to the rising incidence of eoRCC and limited genetic data in younger RCC patients, we performed germline whole exome sequencing (WES) and functional assays on biospecimens from high-risk eoRCC patients diagnosed before 60 years of age, who were negative for PGVs in familial RCC syndrome genes and had a family history of RCC and/or other familial cancers. Our results suggest that constitutional defects in DNA repair, and specifically in function of DNA polymerases, underlie at least a subset of eoRCC cases. Screening patient lymphocytes to identify genotype-phenotype associations via functional assays may provide insight into the mechanism of carcinogenesis for a subset of genetically undiagnosed eoRCCs.

Results.

eoRCC patients at the Fox Chase Cancer Center (FCCC) and family history of cancer.

We analyzed the personal and family history of the probands in a cohort of 22 eoRCC patients. Multiple probands (6/22, 27%) had a second primary cancer, with breast cancer diagnosed in 3 probands (3/22, 14%) prior to diagnosis of RCC (**Table 1**). Here, 73% (n=16/22) of probands had a family history of RCC, with 50% (n=11/22) of probands having a first-degree relative with RCC. Intriguingly, 64% (n=14/22) of probands had a family history of cancers of the prostate, bladder, and thyroid, and melanoma, which have been associated with an RCC diagnosis (9).

Analysis of whole-exome sequencing data reveals enrichment of germline variation in DNA repair and replication genes in eoRCC patients.

We performed WES on lymphocyte DNA from the 22 eoRCC probands, detecting candidate PGVs based on a candidate gene list that was developed in our prior studies ((10); see (**Supplementary Table 1**) and Supplementary Materials and Methods). Briefly, the candidate gene list was developed by a comprehensive hypothesis-driven framework with the following assumptions: 1) genes involved in genome stability (using Gene Ontology terms such as DNA repair, DNA replication, DNA damage checkpoints, cell cycle, mitotic machinery, replication stress, DNA damage response, chromatin remodeling) would be important for hereditary cancer risk (5, 11-13); and 2) an expanded network of genes relevant to renal biology (such as cellular metabolism) and genes somatically mutated in RCC that might be relevant for eoRCC-predisposition (5, 11, 13).

Novel *candidate* PGVs were stringently defined as those that are predicted to disrupt protein function by the consensus of at least 4 protein predictor algorithms; are rare (gnomAD allele frequency<0.01); and are nonsynonymous variants (frameshifts, stop gains, and splicing) (see Supplementary Methods for variant prioritization). We identified candidate PGVs in 18/22 eoRCC patients in the study, yielding a total of 44 variants in 39 genes (**Table 1, and Supplementary Table 2**). Gene Ontology analysis confirmed that the candidate PGVs were enriched in DNA repair and replication pathway genes (**Figure 1A and Supplementary Figure 1,** WebGestalt) (14). Here, 10 patients (46%; 10/22) had 17 candidate PGVs in 14 genes currently associated with hereditary cancers across major organ systems (*ATM*, *BRCA2*, *POLD1*, *POLE*, *FH*, *MITF*, *MSH3*, *MUTYH*, *PDGFRA*, *RET*, *SDHB*, *SMARCA4*, *SMARCE1*, *TSC2*). Only 4 patients had candidate PGVs in RCC familial syndrome genes (18%, 4/22 – *FH*, *MITF*, *SDHB*, *TSC2*). Finally, a total of 14 patients (64%; 14/22) had candidate PGVs from our expanded candidate gene list, from genes not currently defined as RCC-predisposing.

Among the DNA repair-associated genes, candidate PGVs were found in *BRCA2* (2/22 patients; 9%, **Table 1, Supplementary Table 2**) and in *ATM* (2/22 patients; 9%, **Table 1, Supplementary Table 2**). In addition, 5 candidate variants in DNA replication-repair genes (4/22 patients; 18%, **Table 1, Supplementary Table 2,** *POLD1* and *POLH* (Pt #1), *POLE* (Pt #2), *POLK* (Pt #16), and *RRM2B* (Pt #4)). Pt #1 had candidate missense variants in PolD1, a catalytic subunit of the replicative DNA polymerase, Pol δ, and in the translesion synthesis DNA polymerase, Pol η. *POLD1* G2275A p.V759I is in a highly conserved region of PolD1 subunit of the Pol δ (coded by *POLD1* gene) (15) and occurs at a high allele frequency in the Ashkenazi Jewish population (0.0213) reported in gnomAD (versus 0.0018 in the complete gnomAD dataset (16)). *POLH* G626T (p.G209V) is in the Pol η catalytic core (17, 18). Pt #2 has a candidate stop-gain G4872A (p.W1624X) variant in the *POLE* gene, coding replicative DNA polymerase Pol ε, in the conserved C-terminal domain (19). Pt #4 had a splice site (intronic) variant in *RRM2B*, coding a subunit of p53-inducible ribonucleotide reductase, which performs *de novo* conversion of ribonucleotide diphosphates into the corresponding deoxyribonucleotide diphosphates for DNA synthesis (20), in genome position #103237248. Pt#16 has a missense variant in the highly conserved N-terminus domain of another translesion synthesis polymerase, *POLK* G85A (p.E29K); this variant has been

described previously as compromising enzyme activity (21). ClinVar classifications of the variants are presented in **Supplementary Table 2**, with most variants currently classified as variants of uncertain significance (VUS).

Primary lymphocytes from eoRCC patients have reduced capacity to suppress DNA double strand breaks (DSBs).

To begin to assess the functional effect of candidate PGVs in genes linked to DNA replication and repair, we assessed the numbers of γ(phospho)-H2AX foci (a marker of DSBs, (22)) in patient peripheral blood monocytes (PBMCs) at baseline and after treatment with the DNA polymerase inhibitor aphidicolin (**Figure 2A**). In PBMCs from both matched cancer-free controls (by age and gender) and eoRCC patients, aphidicolin significantly elevated the number of γH2AX foci; however, aphidicolin-treated cells from eoRCC patients had markedly higher numbers of γH2AX foci than those from similarly treated controls on treatment, indicating reduced DSB repair mechanism in eoRCC patient cells (**Figure 2A,** $P<0.001$). In complementary work, we tested whether the genes bearing candidate PGVs were specifically needed to suppress DNA DSBs in RCC cells. For this, we used siRNA to deplete the *POLD1, POLE, POLH, POLK, RRM2B,* and *ATM* genes in the Caki RCC cell line. For each gene, knockdown significantly increased γH2AX foci relative to control (**Supplementary Figure 2**) further supporting a role for these proteins in DSB repair in renal cells.

Patient-derived cell lines with candidate PGVs in DNA polymerases have DNA replication defects.

We prepared EBV-transformed cell lines from the primary lymphocytes of 3 patients bearing candidate PGVs (henceforth referred to as the *POLD1/POLH* cell line, *POLE* cell line, and *POLK* cell line) and from several age- and gender-matched cancer-free controls. The *POLD1/POLH* cell line had significantly reduced levels of the Pol η; for the other candidate PGVs, polymerase level was not affected (**Figure 2B)**. Cell Titer Blue (CTB) cellular assays showed significantly better metabolic capacity or cellular viability than control-derived cell lines when treated either with aphidicolin, or with ultraviolet light (which causes bulky adducts in DNA), suggesting that cell lines from patients had better ability to tolerate DNA damage (**Figures 2C-E**). Analysis of cell cycle did not show any significant differences in patients and matched control cell lines (**Supplementary Figure 3A**).

To further expand on these cell-based findings, we used a DNA fiber assay (see Supplementary Methods) and directly compared DNA replication in patient-derived versus control cells, either untreated or following treatment with aphidicolin or with a DNA-alkylating agent, 1-Methyl-3-nitro-1-nitrosoguanidine (MNNG) (**Figure 2F-G Supplementary Figure 4A-D)**. The *POLD1*/*POLH* cell line and the *POLE* cell line exhibited a significantly lower rate of DNA replication in untreated cells (~1.4-fold decrease, $p<0.001$ for the *POLD1*/*POLH* cell line, and ~1.9-fold decrease, $p<0.05$ for the *POLE* cell line versus controls). We also observed a significantly lower replication fork recovery after 2h treatment with aphidicolin (~1.44-fold decrease, $p<0.001$ for the *POLD1*/*POLH* cell line, and ~1.88-fold decrease, $p<0.01$ for the *POLE* cell line versus controls) (**Figure 2G-H).** Intriguingly, the *POLD1*/*POLH* cell line showed defective replication fork restoration (~1.2-fold decrease, $p<0.001$ versus control line) 2h post-treatment with MNNG (**Figure 2G**). The *POLK* cell line did not show any defects in DNA replication

and replication recovery under the conditions tested (**Supplementary Figure 4A)**. A complete summary of results for the DNA polymerase variants is provided in **Supplementary Table 3.**

Altered enzymatic activity of Pol δ and Pol η variant proteins.

Among the candidate PGVs detected in polymerases, the Pol κ variant E29K has previously been biochemically shown to possess not only a significantly reduced catalytic efficiency but also reduced replication fidelity (21). E29K is in a conserved region of the Pol κ N-terminus, the N-clasp subdomain (1-32 aa), which is essential to maintaining the stability of the open conformation of the Pol κ active site (23). Intriguingly, a previous study showed that deletion of the first 67 amino acids reduces Pol κ activity during translesion synthesis (TLS, i.e., replication by efficient bypass of bulky lesions in DNA) (24).

To directly test effects of the other candidate PGVs on polymerase activity, we first purified the polymerase delta (Pol δ) protein complex, with the PolD1 (*POLD1*), PolD2 (*POLD2*), PolD3 (*POLD3*), and PolD4 (*POLD4*) subunits from recombinant protein co-expressed in *E. coli*, and with preparations containing either wild type (wt) PolD1 or PolD1 V759I variant (**Figure 3**). Both the wt and the variant-containing Pol δ complexes extended a Cy3-labeled DNA primer-template; however, the V579I variant complex had significantly less robust polymerase activity than the wt complex (**Figure 3A**, $p<0.001$). Furthermore, when Pol δ complexes containing PolD1 wt or PolD1 V759I proteins were mixed in a ratio of 1:1, the appearance of the extended primer-template was significantly more robust than the variant alone but significantly less robust than the wt alone. This result suggests that the variant is not only impaired for function but has a partial dominance over the wt in this assay (**Supplementary Figure 5**, $p<0.001$).

Pol η is a low fidelity polymerase, which contributes to its ability to perform TLS (25). Hence, G209V variant and wt Pol η prepared in *E. coli* were assessed for their ability to extend labeled DNA primer-template duplexes (**Figures 3B-C**). In the absence of DNA damage (e.g., in a normally base-paired template), the wt and variant proteins both extended the template (**Figure 3B**, $p>0.05$), but the observed bands suggest higher processivity for the variant on template without lesions compared to wt (**Figure 3B**). To evaluate repair of DNA damage, TLS activity was also tested using a template containing an 8-oxoGuanine (8-oxoG) DNA lesion (25). Pol η wt bypassed the 8-oxoG lesion robustly compared to the Pol η variant ($p<0.05$, **Figure 3C**), suggesting better processivity for the wt protein on template with DNA lesion (25).

Finally, biochemical analysis of a purified Pol ε variant, W1624X, was not performed as it is a stopgain variant in the C-terminal domain or CTD, truncating 662 amino acids of the protein. The CTD region is not well-studied, but is thought to be essential for stability of the Pol ε holoenzyme (26).

Structural modeling of DNA polymerase variants in eoRCC suggests impact on polymerase function.

The PolD1 V759I variant is targets a residue two amino acids away from residue D757 (**Figure 3D**, in dark green). In the PolD1 active site, D757 coordinates the Mg^{2+} ions (neon green spheres) required for DNA synthesis and plays a direct role in the catalytic mechanism and binding of DNA (15). Structural modeling indicates that a substitution of the valine (V) 759 to isoleucine (I) could plausibly alter the position of D757 and

disrupt the efficiency of DNA polymerization. To further understand the structural changes that each polymerase variant might induce, we calculated the change in stability of each amino acid substitution as described in Supplemental Methods and **Supplementary Table 4**. Interestingly, the PolD1 I759 yielded a mild stabilization (**ΔΔG** of -1.36 kcal/mole) relative to the wildtype V759. The I759 residue makes twice the number of hydrophobic contacts as the wildtype V759 with the long helix below. The variant could lock this strand in an overly rigid position that compromises DNA polymerization steps that involve flexibility, consistent with the biochemical results observed in **Figure 3A**.

The Pol η G209 residue is in the catalytic core of the polymerase (**Figure 3E,** residues 1-432, colored green) at a position often called the C-cap, i.e. the residue in this position is proximal to the C-terminal end of the α-helix (in orange) (27). Typically, valine, threonine, and isoleucine residues are not preferred in the C-cap, due to poor solvation at the C-terminus of the helix when the side chains are bulky (27). Structural modeling of the G209V substitution showed that the valine with a bulkier side chain could not only alter the stability of the α-helix, but also the nearby β-strands of the catalytic active site (in pink). Rosetta modeling shows the variant G209V is capable of making 5 hydrophobic contacts across the cleft with R24, which is just downstream from the key residue D13, which coordinates active site metals that are central to the catalytic mechanism and the binding of the incoming NTP. The predicted change in stability of the Pol η G209V variant revealed a significant destabilization (**ΔΔG** of 5.67 kcal/mole) relative to the wildtype G209, consistent with lower levels observed in protein purification (**Figure 2B**, **Figure 3B** (left gel), and **Supplementary Table 4**). Interestingly, the activity of the G209V compares well with wt for a normal primer template and might even be more processive (**Figure 3B**, middle gel). However, the variant appears defective for TLS when the template contains an 8oxoG (**Figure 3C**). Consideration of these data together show that changes in stability and conformation may be subtle and even result in an alteration of substrate preference. Thus, a careful combination of both computational and biochemical methods is required to gain a clear understanding of the role a given polymerase variant might have in DNA replication, repair, and cancer initiation and/or progression.

EoRCCs carrying candidate PGVs in DNA polymerases are hypermutant and microsatellite stable (MSS).

To extend these functional tests, we next explored tumor mutation burden (TMB) in tumors from RCC patients in TCGA and from the FCCC eoRCC patients. Previous studies have shown that colorectal and endometrial tumors carrying mutations in *POLE* exonuclease domain (ExoD) and in *POLD1* exhibit a high burden of mutations, are typically MSS but few cases with microsatellite instability (MSI) have been reported, and do not exhibit loss of heterozygosity (LOH) (15, 28-36). RCCs are typically non-hypermutated, with an average TMB of ~1 mut/Mb (37); however, rare hypermutated ($\geq$10 mut/Mb) and rarer ultra-hypermutated (>100 mut/Mb) RCCs carrying polymerase mutations, with or without MSI, have been reported (37). Analysis of renal tumors in TCGA found that several candidate PGV genes from this study (**Figure 4A**), including specifically DNA polymerase genes were mutated in these hypermutant clear cell RCCs (ccRCCs) s (**Figure 4A**). We analyzed the TMB and MSI/MSS status of tumors from the FCCC eoRCC patients and found hypermutation without MSI (*POLD1/POLH* tumor with 12.85 mut/Mb; *POLE* tumor 14.44 mut/Mb) (**Figure 4B**). The tumors from the FCCC eoRCC patients

did not exhibit LOH of the polymerase genes, as has been observed in other polymerase-mutated tumors (37, 38) (**Figure 4C**).

To expand the analysis of DNA polymerases in RCC, we next modeled the structural consequences of the PolD1 and Pol ε variants in hypermutated ccRCCs from TCGA. **Supplementary Table 5** shows the predicted changes in stability for 20 different PolD1 variants and 23 Pol ε variants from hypermutated ccRCCs in TCGA. A broad range of both stabilizing and destabilizing variants was found in all domains of each of these polymerases. **Figure 5** shows these variants in relation to known pathogenic variants in *POLD1* (**Figure 5A-E**) and *POLE* (**Figure 5F-I**). PolD1 R823G/L is in a β-sheet of the polymerase domain close to the DNA (**Figure 5C**). The substitution of a positively charged arginine (R) to a hydrophobic glycine (G) or leucine (L) could destabilize the β-sheet and impact DNA binding. PolD1 D893N is positioned close to the DNA and a variant in this region may destabilize DNA binding or position for DNA-protein interactions (**Figure 5D**). PolD1 P151S is in a β-sheet in the ExoD and the change from proline (P) to serine (S) could destabilize the β-sheet geometry (**Figure 5E**). P151 is close to a known cancer driver mutation (37, 39), E245K, in the unfolded region of the ExoD (**Figure 5A**).

Pol ε P696R is in the palm region of the polymerase domain, which is highly conserved among replicative polymerases (**Figure 5G)**. Arginine (R) has a very large positively charged side chain when compared to smaller proline (P), suggesting this variant may disrupt the polymerase structure and impact DNA synthesis. Pol ε S803L and F753L are in the flexible region of the polymerase domain or the fingers (in cyan, **Figure 5I**). This finger region shifts (27° tilt) on DNA binding (40), and thus plays an essential role in polymerase function. S803 is close to the positively charged lysine (K) 431 in the ExoD, and serine (S) is polar and a smaller residue than the hydrophobic leucine (L). Pol ε S803L is near a site of known cancer driving mutations, C810 (37), suggesting that specific alterations in this α-helix could impact polymerase functioning. F753L is on the border of the fingers and the polymerase domain, close to the ExoD and could be important in the coordination of these regions with or without DNA binding. Finally, several variants were found in the C-terminal domain (**Figure 5H**, in light grey), which is currently not well-studied, but is known for stabilizing the Pol2 (human Pol ε) complex in yeast (26).

Discussion.

In this study, we focus on analysis of candidate PGVs in DNA repair and replication genes in probands with RCC diagnosed prior to 60 years of age undergoing cancer risk assessment at our cancer center who tested negative for RCC familial syndrome genes. We applied a well-curated pipeline of candidate genes in genome stability, metabolism, metabolic stress, normal renal function, RCC biology, and chromatin remodeling to germline WES data from eoRCC patients. Gene Ontology analysis confirmed that the identified candidate PGVs were enriched in DNA repair and replication genes. Intriguingly, we found that many eoRCC patients exhibit defects in suppression of DSBs in their primary PBMCs, with PBMCs from eoRCC patients exhibiting higher γH2AX foci than matched cancer-free controls in response to DNA damage. Direct knockdown of some of these candidate variant genes in Caki RCC cell line also led to increased γH2AX foci. Genes with candidate PGVs were found to be mutated in sporadic RCCs, with specific enrichment of alterations in *BRCA2*, and in DNA polymerase (*POLE, POLD1, POLH1,* and *POLK*) genes in hypermutant RCCs in TCGA. Importantly, detailed analysis of candidate

PGVs in DNA polymerase genes from the FCCC eoRCC patients confirmed the damaging nature of the candidate PGVs and suggested a mechanistic basis for association of these variants with observed defects in DNA repair and replication. Several PolD1 and Pol ε variants from the hypermutant RCCs in TCGA were proximal to the catalytic center or substrate binding regions and will benefit from similar future biochemistry experiments shown in this study.

This work complements a number of recent studies indicating inherited defects in DNA replication machinery may increase cancer risk. PGVs in the ExoD of *POLE* and *POLD1* predispose to cancer and exhibit a strong mutagenic effect, however, the role of non-ExoD variants in cancer risk has been controversial (29, 31, 41). A recent study reported that the *POLD1* candidate PGV (p.V759I, in Pt #1) is frequently present in the Ashkenazi Jewish population and proposed this gene as a founder mutation (2). Mertz et. al. have demonstrated a strong mutator effect of the PolD1 polymerase domain variant R689W in human cells (42). Several TLS polymerases including Pol η and Pol κ, are important for preventing accumulation of single strand DNA gaps, and the replication of DNA fragile sites. Owing to their high error-propensity, TLS polymerases are likely to contribute to oncogene-induced mutagenesis (43, 44).

It is likely that the candidate PGVs in DNA replication and repair genes detected here interact with other germline variants to impact eoRCC risk. In this study, Pt #1 also harbored candidate PGVs in *MTOR, PARP1, FH, MCM2,* suggesting the *POLD1* and *POLH* variants assessed here may act together with other variants to augment the DNA repair defects observed. Pt#2 with a *POLE* candidate PGV also harbored candidate PGVs in *BRCA2, PDGFRA.* In other studies, hypermutant ccRCCs also carried mutations in *TP53, PTEN, VHL,* and *UNC5C* (45, 46). It is also possible that defects in polymerase genes affect cancer risk by affecting biological processes beyond DNA replication and repair. We found that the immortalized cell lines from Pt#1, 2, and 16 (*POLK*) showed better metabolic capacity compared to control cell lines. Intriguingly, patients with mandibular hypoplasia, deafness and progeroid features with concomitant lipodystrophy, characterized by germline PGVs in *POLD1,* also present with mitochondrial dysfunction and metabolic abnormalities (15, 47, 48). Conversely, some familial RCC genes (such as *FH, VHL, PBMR1,* and *SDHx*) have also been implicated in suppression of DSBs and in replication stress (49-52), based on mechanisms that are not well understood.

It is important to note that while the family history of the high-risk probands in this study is suggestive of underlying genetics (53), the clinical testing using a RCC familial syndrome genes or panel did not yield any actionable PGVs according to current NCCN recommended guidelines. Here, 64% of probands had an extensive family history of cancers of the prostate, bladder, and thyroid, and melanoma, all of which have previously been associated with RCC diagnosis (9). In fact, PGVs in DNA repair genes have been reported as risk factors for bladder, skin, thyroid, and prostate cancers (54-59). Our results suggest that in the absence of PGVs in RCC familial syndrome genes or phenotypic features of familial RCC or family history of RCC, a comprehensive assessment of general cancer predisposition genes, including DNA repair genes, may be beneficial. Here, multiple probands (27%) had at least one additional primary cancer with breast cancer being the most common additional primary cancer (14%). A recent retrospective analysis of the Swedish Cancer Registry showed that

~10% of RCC patients develop another second primary cancer, and this is currently thought to be independent of the primary RCC, suggesting broader cancer predisposition (60), compatible with PGVs in genes affecting DNA repair.

Besides genetic screening, these data suggest the value of functional assessment for the families of individuals with eoRCC. In this study, the majority of eoRCC PBMC biospecimens samples exhibited elevated γH2AX levels and candidate variants in DNA repair genes. Our data support a potential role of germline variation in DNA repair/replication leading to suboptimal encoding protein activity, and genome instability. Overall, these data suggest that assays of γH2AX foci in normal cells, supporting germline variation in DNA repair/replication genes, could be a potential tool for the identification of individuals with genetically unexplained eoRCC. As these defects could be detected in normal cells, it could lead to the identification of individuals in need of cancer risk assessment. This is especially relevant because case-control studies suggest that an elevated familial RCC risk may be multifactorial, and or due to an interaction of the heritable genetics and the shared environment (53). It is possible that defective DNA repair in the heterozygous state could be a recessive heritable factor that when combined with other RCC risk factors may jointly increase the risk of eoRCC.

Currently, the therapeutic significance of DNA repair genes is not clinically defined for RCC. Evidence is emerging that PARP inhibitors could be therapeutics of choice in RCCs that may not carry mutations in the classical *BRCA* genes, but which have other defects in DNA repair, with recent clinical trials assessing the use of PARP inhibitors in RCC (61, 62). Hence, there is a critical need to not only understand the biological impact of defective DNA repair in renal tissue but to also define risk of RCC due to a germline defect in DNA repair genes. Further exploration in a larger and more diverse (by race/ethnicity) patient population is clearly of interest for future work. In sum, study of PBMCs offers a useful approach to resolve the biological and clinical significance of rare gene variants identified by exome sequencing and may improve clinical approaches to risk assessment and medical management.

Materials and Methods.

eoRCC patient population, and peripheral blood DNA analysis

Case-only eoRCC probands that underwent clinical germline genetic testing between 2010-2016 were included in this study (n=22). Patients were followed by the Genitourinary Program at the Fox Chase Cancer Center and had undergone evaluation for inherited cancer risk at the FCCC Family Risk Assessment Program (RAP). Each participant had a strong family cancer history as shown in **Table 1**, with either multiple first-degree or second-degree relatives with RC, RC-associated cancers, or other cancers. The mean age at eoRCC diagnosis was 48 years (range 36–59 years). No pathogenic mutations were identified from sequencing the following RC-specific genes: *VHL, MET, FLCN, TSC1, TSC2, FH, SDHx, PTEN* and *BAP1*. The patients had consented to the FCCC RAP Registry, which allowed further research, genomic sequencing, and banking of their biospecimens in the BioSample Repository Facility. The patients reported here were self-reported white,

non-Hispanic. Family histories were obtained by trained licensed genetic counselors and verified by attending physicians. The analyses performed and publication of deidentified information was under the approval of the FCCC Institutional Review Board Committee protocol number 14-831. See **Supplementary Methods** for more methods.

Abbreviations. 8-oxoG, 8-oxoGuanine; BWA, Burrows-Wheeler aligner; ccRCC, clear cell renal cell carcinoma; CTB, Cell Titer Blue; CldU, 5-Chloro-2'-deoxyuridine; DSBs, double-strand breaks; eoRCC, early-onset renal cell carcinoma; DDR, DNA damage response; EBV, Epstein-Barr virus; ExAC, Exome Aggregation Consortium; ExoD, exonuclease domain; FCCC, Fox Chase Cancer Center; FDR, false discovery rate; GATK, genome analysis toolkit; GnomAD, Genome Aggregation Database; GO, gene ontology; IdU, 5-Iodo-2′-deoxyuridine; LOH, loss of heterozygosity; MMR, mismatch repair; MNNG, 1-Methyl-3-nitro-1-nitrosoguanidine; MSI, microsatellite instability; MSS, microsatellite stability; NCCN, National Comprehensive Cancer Network; PBMCs, peripheral blood monocytes; PGVs, pathogenic germline variants; Pt, patient; RAP, risk assessment program; RCC, renal cell carcinoma; SNPs, single nucleotide polymorphisms; TLS, translesion synthesis; TMB, tumor mutation burden; VUS, variant of uncertain significance; WebGestalt, WEB-based GEne SeT AnaLysis Toolkit; WES, whole-exome sequencing; wt, wild type.

Acknowledgments. The work in this grant was supported by the resources and expertise provided by the FCCC Cell Culture, Biosample Repository, Genomics, Biostatistics, Population Sciences, Molecular Modelling, and High Throughput Screening Facilities. We acknowledge Nina Shah (summer undergraduate at FCCC) for assistance with knockdown experiments, Emmanuelle Nicolas (Genomics Facility, FCCC) and Waleed Iqbal (Drexel University, PA) for assistance with genomic analysis. We express gratitude to Lisa Bealin, Andrea F. Forman, and Kim Rainey (Department of Clinical Genetics, FCCC) for valuable assistance on the study.

Grant Support. All Fox Chase Cancer Center affiliated authors are in part supported by the NCI Core Grant, P30 CA006927, to the Fox Chase Cancer Center. S.A. was supported by the DOD W81XWH-18-1-0148, and a CEP Award from the Yale Head and Neck Cancer SPORE. M.J.H. was supported by funding from the American Cancer Society. M.B.D was supported by the NIH U01 CA164920, R01 CA207365 grants. E.A.G. was supported by NIH R01 DK108195. E.V.D. and I.G.S. were partially supported by the Kazan Federal University Strategic Academic Leadership Program (PRIORITY-2030). R.J.D. was supported by NIH R35 GM122517. R.T.P. was supported by NIH grants 1R01GM130889 and 1R01GM137124.

References.

1. H. Sung, R. L. Siegel, P. S. Rosenberg, A. Jemal, Emerging cancer trends among young adults in the USA: analysis of a population-based cancer registry. *The Lancet Public Health* **4**, e137-e147 (2019).
2. G. Rosner *et al.*, POLD1 and POLE Gene Mutations in Jewish Cohorts of Early-Onset Colorectal Cancer and of Multiple Colorectal Adenomas. *Dis Colon Rectum* **61**, 1073-1079 (2018).
3. M. I. Carlo *et al.*, Prevalence of Germline Mutations in Cancer Susceptibility Genes in Patients With Advanced Renal Cell Carcinoma. *JAMA Oncol* **4**, 1228-1235 (2018).

4. W. M. Linehan *et al.*, Molecular diagnosis and therapy of kidney cancer. *Annu Rev Med* **61**, 329-343 (2010).
5. W. M. Linehan, R. Srinivasan, L. S. Schmidt, The genetic basis of kidney cancer: a metabolic disease. *Nat Rev Urol* **7**, 277-285 (2010).
6. T. R. Hartman *et al.*, Prevalence of pathogenic variants in DNA damage response and repair genes in patients undergoing cancer risk assessment and reporting a personal history of early-onset renal cancer. *Scientific Reports* **10**, 13518 (2020).
7. R. J. Motzer *et al.*, NCCN Guidelines Insights: Kidney Cancer, Version 2.2020: Featured Updates to the NCCN Guidelines. *Journal of the National Comprehensive Cancer Network J Natl Compr Canc Netw* **17**, 1278-1285 (2019).
8. H. Truong *et al.*, Germline Variants Identified in Patients with Early-onset Renal Cell Carcinoma Referred for Germline Genetic Testing. *Eur Urol Oncol* 10.1016/j.euo.2021.09.005 (2021).
9. H. Liu, J. Sundquist, K. Hemminki, Familial renal cell carcinoma from the Swedish Family-Cancer Database. *Eur Urol* **60**, 987-993 (2011).
10. E. Nicolas *et al.*, Interaction of germline variants in a family with a history of early-onset clear cell renal cell carcinoma. *Mol Genet Genomic Med* **7**, e556 (2019).
11. M. Berti, A. Vindigni, Replication stress: getting back on track. *Nat Struct Mol Biol* **23**, 103-109 (2016).
12. J. A. Coleman, P. Russo, Hereditary and familial kidney cancer. *Curr Opin Urol* **19**, 478-485 (2009).
13. J. C. Henegan, Jr., C. R. Gomez, Heritable Cancer Syndromes Related to the Hypoxia Pathway. *Front Oncol* **6**, 68 (2016).
14. Y. Liao, J. Wang, E. J. Jaehnig, Z. Shi, B. Zhang, WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res* **47**, W199-w205 (2019).
15. E. Nicolas, E. A. Golemis, S. Arora, POLD1: Central mediator of DNA replication and repair, and implication in cancer and other pathologies. *Gene* **590**, 128-141 (2016).
16. K. J. Karczewski *et al.*, The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434-443 (2020).
17. P. Tomczyk, E. Synowiec, D. Wysokiński, K. Woźniak, Eukaryotic TLS polymerases. *Postepy Hig Med Dosw (Online)* **70**, 522-533 (2016).
18. N. Acharya *et al.*, Multifaceted activities of DNA polymerase η: beyond translesion DNA synthesis. *Curr Genet* **65**, 649-656 (2019).
19. P. M. J. Burgers, T. A. Kunkel, Eukaryotic DNA Replication Fork. *Annual review of biochemistry* **86**, 417-438 (2017).
20. W. Iqbal *et al.*, RRM2B Is Frequently Amplified Across Multiple Tumor Types: Implications for DNA Repair, Cellular Survival, and Cancer Therapy. *Front Genet* **12**, 628758 (2021).
21. S. Yadav, S. Mukhopadhyay, M. Anbalagan, N. Makridakis, Somatic Mutations in Catalytic Core of POLK Reported in Prostate Cancer Alter Translesion DNA Synthesis. *Human mutation* **36**, 873-880 (2015).
22. S. Arora *et al.*, Genetic Variants That Predispose to DNA Double-Strand Breaks in Lymphocytes From a Subset of Patients With Familial Colorectal Carcinomas. *Gastroenterology* **149**, 1872-1883 e1879 (2015).
23. V. Jha, H. Ling, 2.0 Å resolution crystal structure of human polκ reveals a new catalytic function of N-clasp in DNA replication. *Sci Rep* **8**, 15125 (2018).
24. S. Lone *et al.*, Human DNA polymerase kappa encircles DNA: implications for mismatch extension and lesion bypass. *Mol Cell* **25**, 601-614 (2007).
25. S. C. Suarez, R. A. Beardslee, S. M. Toffton, S. D. McCulloch, Biochemical analysis of active site mutations of human polymerase η. *Mutat Res* **745-746**, 46-54 (2013).
26. Z. Yuan, R. Georgescu, G. D. Schauer, M. E. O'Donnell, H. Li, Structure of the polymerase ε holoenzyme and atomic model of the leading strand replisome. *Nature Communications* **11**, 3156 (2020).
27. A. J. Doig, R. L. Baldwin, N- and C-capping preferences for all 20 amino acids in alpha-helical peptides. *Protein Sci* **4**, 1325-1336 (1995).
28. S. R. Barbari, P. V. Shcherbakova, Replicative DNA polymerase defects in human cancers: Consequences, mechanisms, and implications for therapy. *DNA Repair (Amst)* **56**, 16-25 (2017).
29. D. N. Church *et al.*, DNA polymerase ε and δ exonuclease domain mutations in endometrial cancer. *Hum Mol Genet* **22**, 2820-2828 (2013).

30. K. P. Hodel *et al.*, Explosive mutation accumulation triggered by heterozygous human Pol ε proofreading-deficiency is driven by suppression of mismatch repair. *Elife* **7** (2018).
31. C. Palles *et al.*, Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nat Genet* **45**, 136-144 (2013).
32. S. Zhao *et al.*, Landscape of somatic single-nucleotide and copy-number mutations in uterine serous carcinoma. *Proc Natl Acad Sci U S A* **110**, 2916-2921 (2013).
33. E. E. Henninger, Z. F. Pursell, DNA polymerase ε and its roles in genome stability. *IUBMB Life* **66**, 339-351 (2014).
34. E. Shinbrot *et al.*, Exonuclease mutations in DNA polymerase epsilon reveal replication strand specific mutation patterns and human origins of replication. *Genome Res* **24**, 1740-1750 (2014).
35. A. Shlien *et al.*, Combined hereditary and somatic mutations of replication error repair genes result in rapid onset of ultra-hypermutated cancers. *Nat Genet* **47**, 257-262 (2015).
36. S. Arora *et al.*, The landscape of POLE variants in colorectal and endometrial tumors: Correlation with microsatellite instability (MSI) and tumor mutation burden (TMB). *Journal of Clinical Oncology* **38**, e13538-e13538 (2020).
37. B. B. Campbell *et al.*, Comprehensive Analysis of Hypermutation in Human Cancer. *Cell* **171**, 1042-1056.e1010 (2017).
38. V. S. Park, Z. F. Pursell, POLE proofreading defects: Contributions to mutagenesis and cancer. *DNA Repair (Amst)* **76**, 50-59 (2019).
39. J. He *et al.*, Distinctive genomic characteristics in POLE/POLD1-mutant cancers can potentially predict beneficial clinical outcomes in patients who receive immune checkpoint inhibitor. *Ann Transl Med* **9**, 129-129 (2021).
40. C. Lancey *et al.*, Structure of the processive human Pol δ holoenzyme. *Nat Commun* **11**, 1109 (2020).
41. P. Mur *et al.*, Role of POLE and POLD1 in familial cancer. *Genetics in Medicine* **22**, 2089-2100 (2020).
42. T. M. Mertz, A. G. Baranovskiy, J. Wang, T. H. Tahirov, P. V. Shcherbakova, Nucleotide selectivity defect and mutator phenotype conferred by a colon cancer-associated DNA polymerase δ mutation in human cells. *Oncogene* **36**, 4427-4433 (2017).
43. C. Vaziri, I. B. Rogozin, Q. Gu, D. Wu, T. A. Day, Unravelling roles of error-prone DNA polymerases in shaping cancer genomes. *Oncogene* **40**, 6549-6565 (2021).
44. S. S. Lange, K. Takata, R. D. Wood, DNA polymerases and cancer. *Nat Rev Cancer* **11**, 96-110 (2011).
45. V. S. Park *et al.*, Mouse model and human patient data reveal critical roles for Pten and p53 in suppressing POLE mutant tumor development. *NAR Cancer* **4** (2022).
46. I. G. Serebriiskii *et al.*, Comprehensive characterization of PTEN mutational profile in a series of 34,129 colorectal cancers. *Nature Communications* **13**, 1618 (2022).
47. F. Reinier *et al.*, Mandibular hypoplasia, deafness, progeroid features and lipodystrophy (MDPL) syndrome in the context of inherited lipodystrophies. *Metabolism* **64**, 1530-1540 (2015).
48. M. Murdocca *et al.*, Mitochondrial dysfunction in mandibular hypoplasia, deafness and progeroid features with concomitant lipodystrophy (MDPL) patients. *Aging (Albany NY)* **14**, 1651-1664 (2022).
49. E. V. Demidova, P. Ghatalia, S. Arora, Combination Strategies for Immune Checkpoint Inhibitors in PBRM1-mutant Renal Cell Carcinoma: To PARP or Not To PARP? *European Urology* https://doi.org/10.1016/j.eururo.2021.10.028 (2021).
50. P. L. Sulkowski *et al.*, Krebs-cycle-deficient hereditary cancer syndromes are defined by defects in homologous-recombination DNA repair. *Nat Genet* **50**, 1086-1092 (2018).
51. S. E. Scanlon, D. C. Hegan, P. L. Sulkowski, P. M. Glazer, Suppression of homology-dependent DNA double-strand break repair induces PARP inhibitor sensitivity in VHL-deficient human renal cell carcinoma. *Oncotarget* **9**, 4647-4660 (2018).
52. R. M. Chabanon *et al.*, PBRM1 Deficiency Confers Synthetic Lethality to DNA Repair Inhibitors in Cancer. *Cancer Res* **81**, 2888-2902 (2021).
53. R. J. Hung *et al.*, Family history and the risk of kidney cancer: a multicenter case-control study in Central Europe. *Cancer Epidemiol Biomarkers Prev* **16**, 1287-1290 (2007).
54. E. Nicolas *et al.*, Systematic evaluation of underlying defects in DNA repair as an approach to case-only assessment of familial prostate cancer. *Oncotarget* **6**, 39614-39633 (2015).
55. S. M. Shah *et al.*, Therapeutic implications of germline vulnerabilities in DNA repair for precision oncology. *Cancer Treat Rev* **104**, 102337 (2022).

56. E. M. V. Allen *et al.*, Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* **350**, 207-211 (2015).
57. A. Pemov *et al.*, Identification of Genetic Risk Factors for Familial Urinary Bladder Cancer: An Exome Sequencing Study. *JCO Precision Oncology* 10.1200/PO.21.00115, 1830-1839 (2021).
58. J. E. Sandler *et al.*, Germline Variants in DNA Repair Genes, Diagnostic Radiation, and Risk of Thyroid Cancer. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **27**, 285-294 (2018).
59. P. J. Vlachostergios *et al.*, The emerging landscape of germline variants in urothelial carcinoma: Implications for genetic testing. *Cancer Treatment and Research Communications* **23**, 100165 (2020).
60. K. Czene, K. Hemminki, Kidney cancer in the Swedish Family Cancer Database: familial risks and second primary malignancies. *Kidney Int* **61**, 1806-1813 (2002).
61. NCT03786796 (Study of Olaparib in Metastatic Renal Cell Carcinoma Patients With DNA Repair Gene Mutations (ORCHID).
62. NCT04337970 (Talazoparib and Axitinib for People With Previously Treated Advanced Kidney Cancer.

Figure Legends.

Figure 1. Enrichment of predicted pathogenic variants in DNA repair genes in the cohort. A. Summary of variants in genes and pathways, identified in the cohort. In color - number of variants identified for each gene. For detailed information, see Supplementary tables 1 and 2.

Figure 2. Cell-based functional analysis revealed defects in DNA repair and DNA replication in lymphocytes from eoRCC patients. A. γH2AX foci immune fluorescence staining in primary PBMCs from eoRCC patients versus matched controls, at baseline or post treatment with aphidicolin (2h). PBMCs from patients showed statistically significant elevation of γH2AX foci post treatment with aphidicolin. Data were normalized and are presented as percent of positive γH2AX foci. **B.** Representative Western blots showing expression of PolD1, Pol η, Pol ε and Pol κ in EBV-transformed cell lines carrying variants versus matched controls (without the variants). Data quantification was performed based on 3 independent biological repeats, technical repeats are presented on gels. **C-E.** Relative viability of EBV-transformed cell lines was assessed by CTB assay at baseline or after treatment with aphidicolin or UV. Data were normalized to CTB values for controls and are presented as percent cellular viability for *POLD1*/*POLH* (**C**), *POLE* (**D**), and *POLK* (**E**) cell lines. Data from 3 independent biological repeats are presented. **F-G.** Difference in DNA replication fork elongation/restoration in EBV-transformed cell lines (**F** - *POLD1*, *POLH*; **G** - *POLE* lines) at the baseline and post replications stress was assessed using DNA fiber assay. At baseline the EBV-transformed cells were labeled with IdU for 20 min, for fork restoration cells then were treated with 100 μM aphidicolin or 1 uM MNNG for 2 h, and then labeled with CldU for 40 min. For all conditions, post labeling, cells were lysed, and DNA fibers stretched onto glass-slides, fixed, denatured, blocked, and stained with corresponding antibodies. Fiber images were captured using the Nikon TS2R Inverted Microscope and analyzed in ImageJ software. Data for 3 independent repeats are presented as IdU tract length or CldU/IdU tract length ratio. For all graphs: *** for $p<0.001$, ** for $p<0.01$, * for $p<0.05$ and NS for $p>0.05$, unpaired, non-parametric t-test, Mann-Whitney criteria.

Figure 3. Structural and biochemical assays revealed altered enzymatic activities of the PolD1 and Pol η variants. A. Pol δ complex and primer extension assay. On the left - representative gel image of purified wt and variant Pol δ protein complexes, containing 4 subunits: PolD1 (125 kDa), PolD2 (50 kDa), PolD3 (66 kDa) and PolD4 (12.5 kDa). Center and right – Pol δ complex primer extension assay with quantification. Representative gel image showing reactions performed with 20 nM Cy-3 labeled DNA-duplex template (SA#1), 20 nM of indicated proteins and 500 uM dNTPs. PolD1 V759I complex extended DNA-template less efficiently comparing to wt protein complex. Data for 3 independent repeats are presented**. B.** Pol η and primer extension assay. On the left - representative gel image of purified wt and variant Pol η catalytic cores (432 amino acids), molecular weight ~56 kDa. Center and right – Pol η primer extension assay with quantification. Representative gel image showing reactions performed with 20 nM Cy-3 labeled DNA-duplex template (SA#1), 20 nM of indicated proteins and 500 uM dNTPs. Data for 3 independent repeats are presented. **C.** Pol η lesion (8-oxoG) bypass assay with quantification. Representative gel image showing reactions performed with 20 nM Cy-3 labeled DNA-duplex template with 8-oxoG in the position, opposite to 3`-OH group (SA#4), and template of the same sequence without lesion (SA#3), 20 nM of indicated proteins and 500 uM dNTPs. Data for 3 independent repeats are presented. For **A-C**: *** for p<0.001, ** for p<0.01, * for p<0.05 and NS for p>0.05, unpaired, non-parametric t-test, Mann-Whitney criteria. All template sequences may be found in **Supplementary Table 4**. **D-E.** Homology modeling structures using yeast protein templates for human PolD1 V759I (**D**) and for human Pol η G209V (**E**).

Figure 4. Renal tumors carrying polymerase variants showed high TMB, MSS, and no LOH. A. Percent alteration frequency in 897 tumors from TCGA in different histological types of RCC: chromophobe (n=66), ccRCC - clear cell renal cell carcinoma (n=538), ccRCC (hyper) - hypermutated samples (n=12), papillary (n=293). **B.** TMB and MSS data are presented for Pt #1 (*POLD1* V759I, *POLH* G209V) and Pt #2 (*POLE* W1624X). **C.** Tumor and normal Sanger sequencing for variants in Pt #1 (*POLD1* V759I, *POLH* G209V) and Pt #2 (*POLE* W1624X) showing no LOH. Arrows show variants of interest on sequencing tracks.

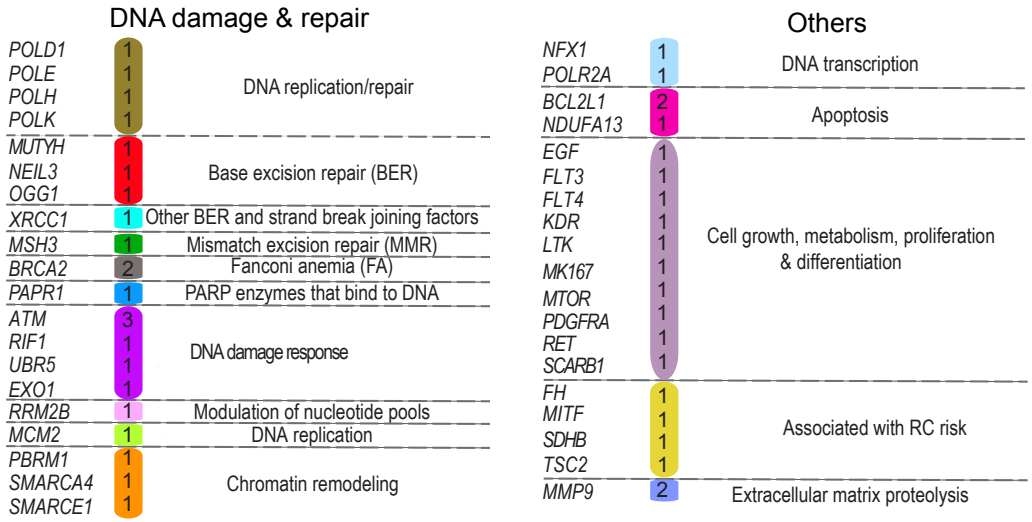
Figure 5. Structure mapping of the novel PolD1 and Pol ε variants from hypermutated ccRCCs in TCGA. A-E. DNA-bound PolD1 3D-model was refined from PDB:3IAY. The colored functional domains are exonuclease (light blue, residues 131-477) and polymerase (green, residues 550-978). **A.** Red spheres represent known cancer drivers. **B.** Blue spheres represent variants of uncertain significance in ccRCC. **C-E.** Fragments of PolD1 model showing variants: A810S and R823G (**C**), D893N, A810S, and R978C (**D**), P151S and E245K (**E**). **F.** DNA-bound N-terminal domain of Pol ε was refined from PDB:4M8O. The colored functional domains are N-terminal subdomain (dark grey, residues 31-281), exonuclease (wheat, residues 282-527), polymerase (light pink, palm: 528-950; cyan, fingers: 769-833; lime, thumb: 951-1186). Red spheres represent known cancer drivers (structure above). Blue spheres represent variants of uncertain significance in ccRCC (structure below). **G.** A fragment of Pol ε model showing variant P696L. **H.** 3D-model of whole-length Pol ε (without DNA). Structure

was refined as described in Methods based on (26). The colored functional domains are N-terminal subdomain (dark grey, residues 31-281), exonuclease (wheat, residues 282-527), polymerase (light pink, palm: 528-950; cyan, fingers: 769-833; lime, thumb: 951-1186), C-terminal domain (light grey, residues 1308-2222). Blue spheres represent variants of uncertain significance in ccRCC. **I.** Fragments of Pol ε model showing variants S803L and F753.

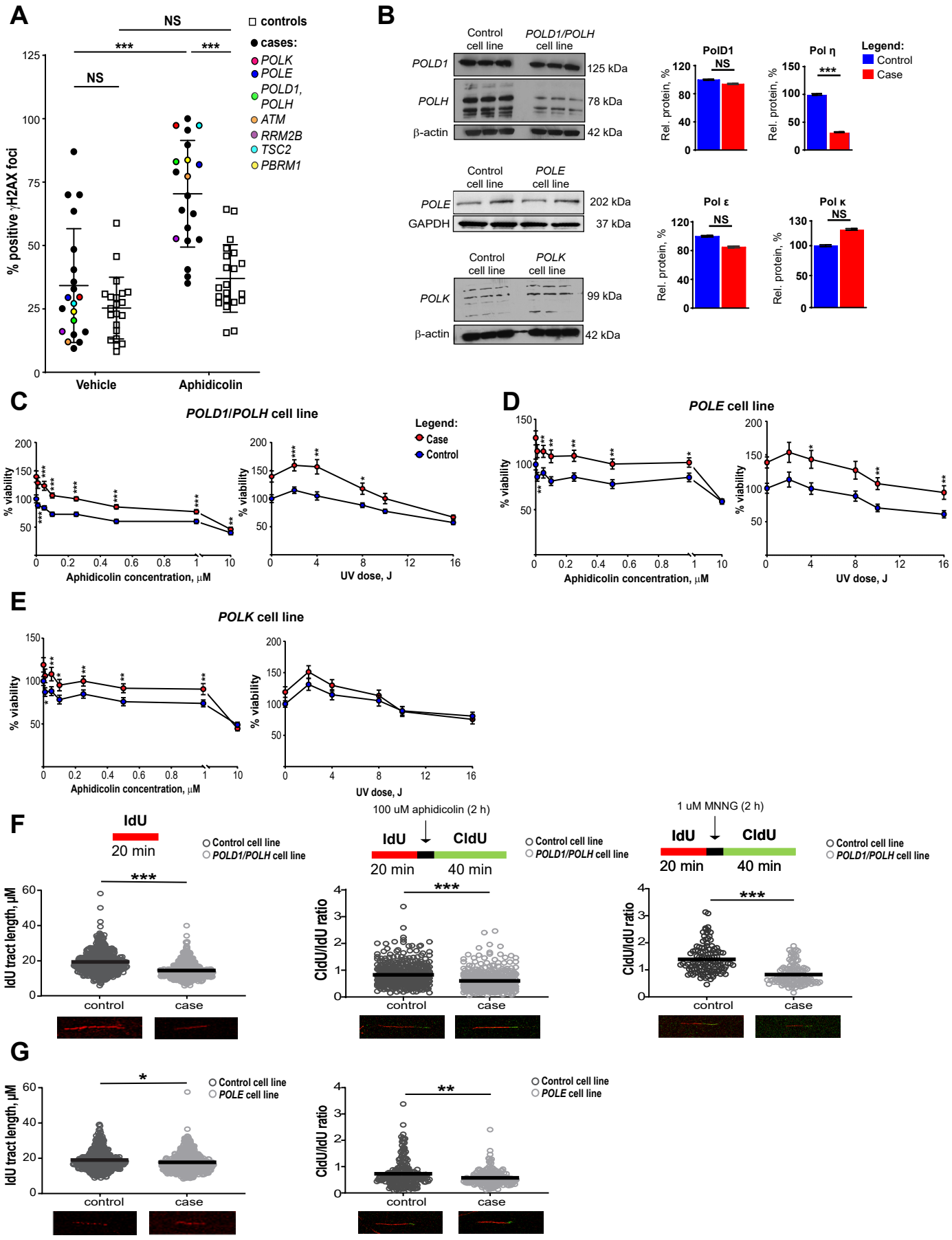
Table Legends.

Table 1. eoRCC patient characteristics, genomic findings, and family history.

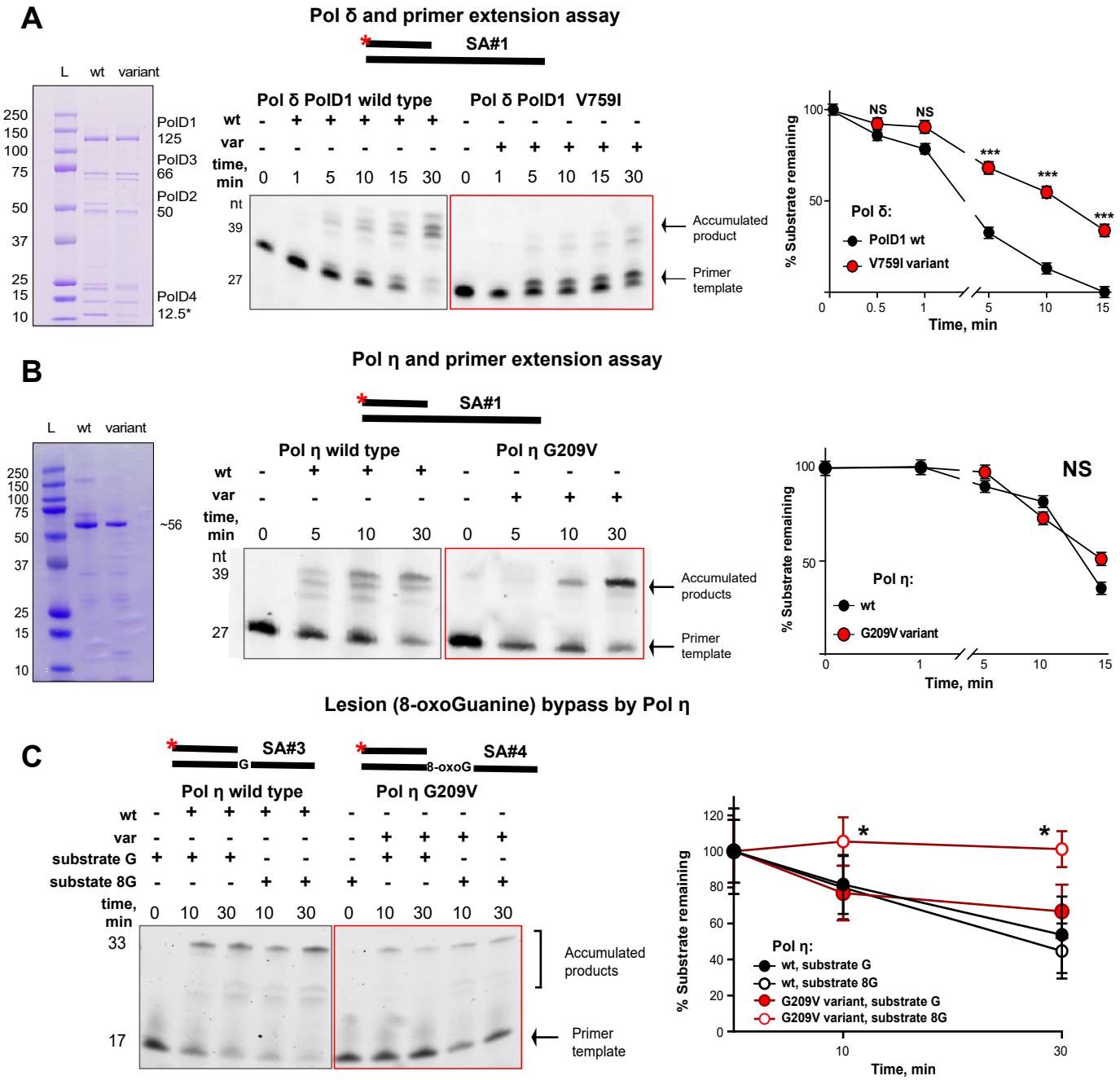
A



Demidova et al., Figure 1

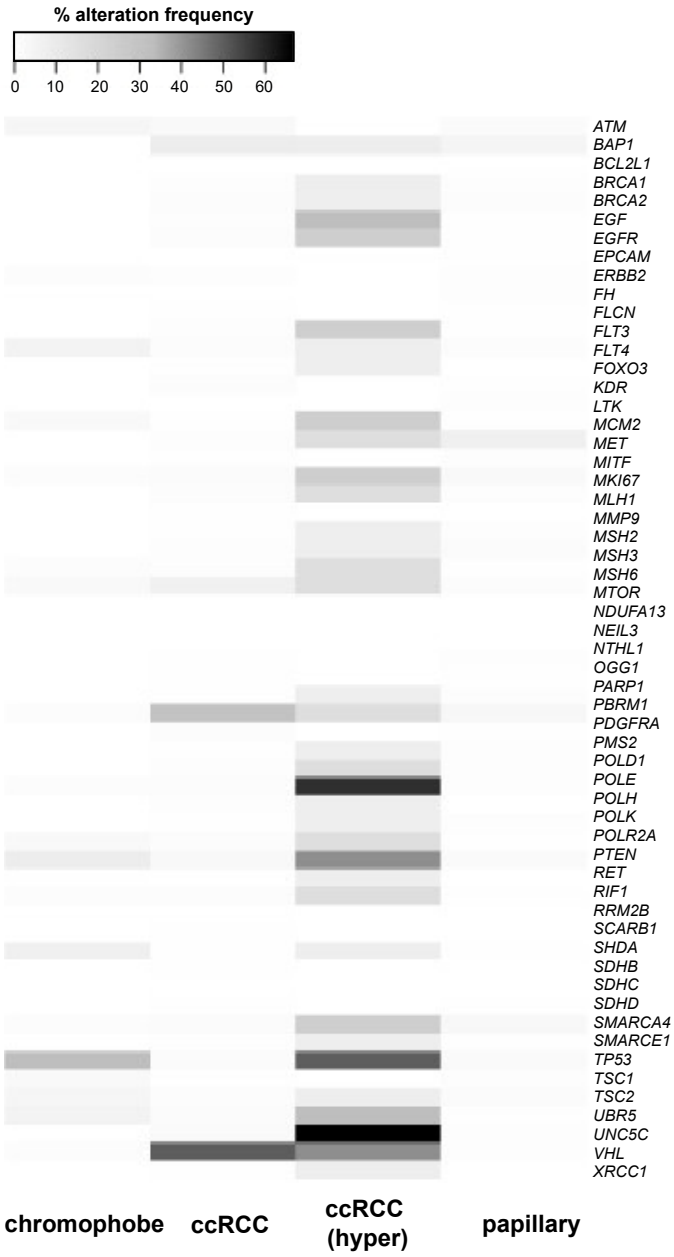


Demidova et al., Figure 2



Demidova et al., Figure 3

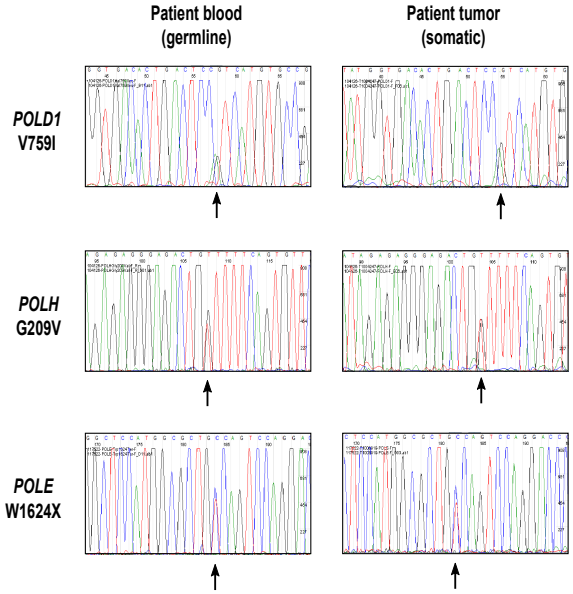
A



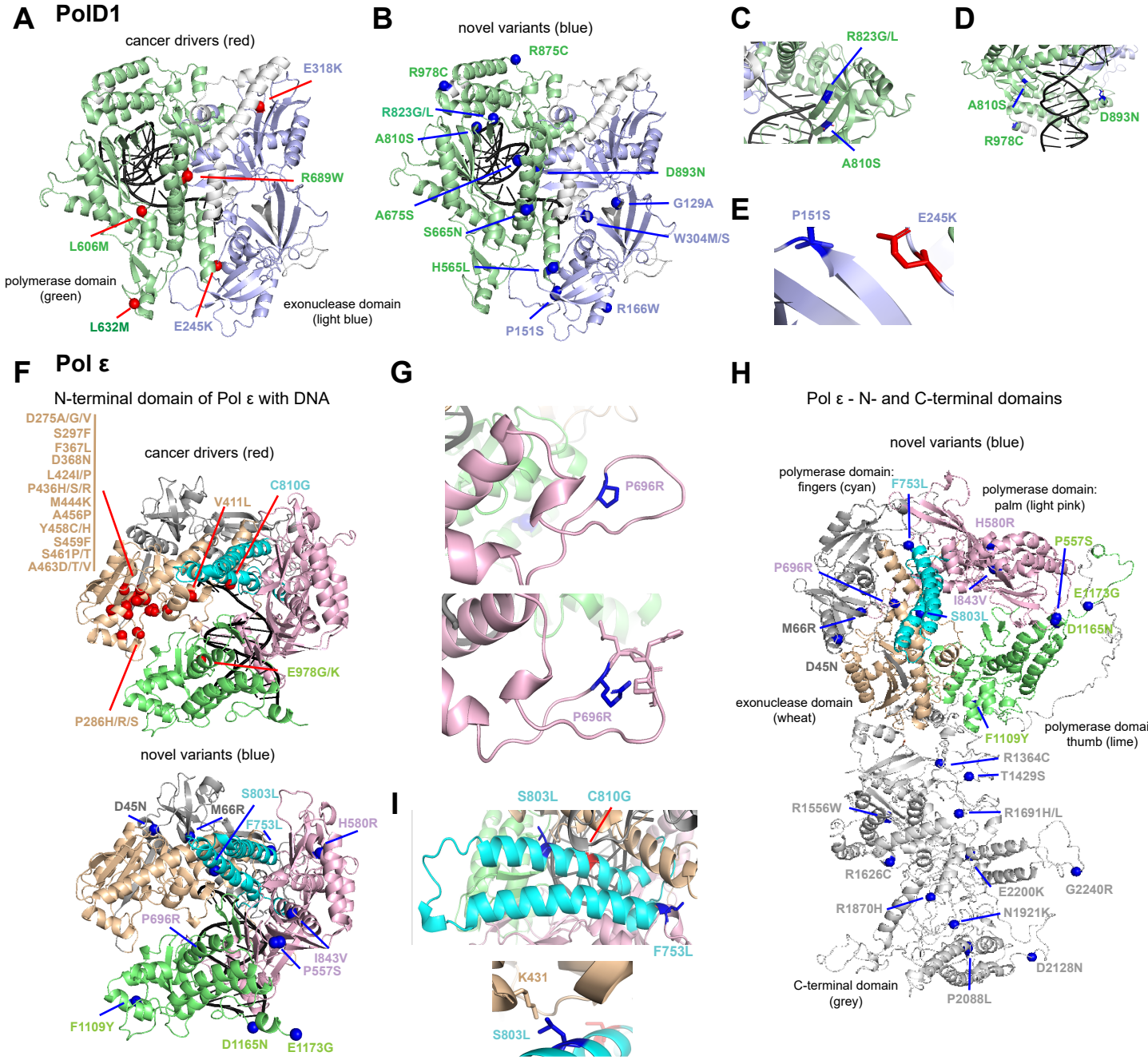
B

Patient #	Tumor variants	TMB, mut/MB	MSS
P1	PolD1 V759I, Pol η G209V	12.85	+
P2	Pol ε W1624X	14.44	+

C



Demidova et al., Figure 4



Demidova et al., Figure 5

Table 1. eoRCC patient characteristics, genomic findings and family history.

Patient #	Sex	Genetic variants	Other cancers	Cancers in relatives			
				1st degree	2nd degree	3rd degree	other
Pt 1	M	*POLD1 POLH MTOR PARP1 FH MCM2*	sarcoma, thyroid, Hodgkins lymphoma	SKIN, PANCREAS	COLON/RECTUM, THYROID, LUNG, PROSTATE, BREAST, BONE	SKIN, STOMACH	
Pt 2	F	*POLE PDGFRA BRCA2*		COLON	BREAST, COLON	RCC, BREAST	
Pt 3	F	*ATM MITF*	breast	RCC, LUNG	BREAST, STOMACH		BLADDER, PANCREAS
Pt 4	F	*RRM2B BCL2L1*	breast	UTERUS	RCC, LEUKEMIA		
Pt 5	F	*OGG1 NEIL3 UBR5*	breast	SKIN		PROSTATE	
Pt 6	F	*RIF1 KDR XRCC1*	neuroblastoma		MELANOMA, LUNG		
Pt 7	F	*MK167*	2 primary RCC	COLON			
Pt 8	F	*RET BCL2L1*		RCC			
Pt 9	F				RCC, BLADDER, TESTICLE/LIVER		
Pt 10	F	*PBRM1*		RCC, BREAST, PROSTATE			BREAST, COLON
Pt 11	F			BREAST, STOMACH	BREAST, BLADDER, RCC		
Pt 12	M	*SCARB1*			RCC		
Pt 13	M	*TSC2*		LUNG, PANCREAS, SKIN			
Pt 14	M	*ATM (2) FLT3 SMARCA4*		RCC, BLADDER, LUNG			
Pt 15	M	*EGF*		RCC, BLADDER, LIVER, PROSTATE			
Pt 16	M	*POLK EXO1 BRCA2 MUTYH*		BLADDER			RCC
Pt 17	M			RCC			
Pt 18	M	*NFX1*		RCC, PROSTATE			
Pt 19	M	*SDHB NDUFA13 MMP9*		RCC, BLADDER, COL.POLYPS	BLADDER		
Pt 20	M	*MMP9 MSH3 LTK POLR2A*		RCC, LUNG, NON-HODGKINS LYMPH			
Pt 21	M			BREAST, RCC			
Pt 22	M	*FLT4 SMARCE1*		RCC	LUNG		

F – female, M – male, Pt – patient, RCC – renal cell carcinoma