Polymerase-mediated ultramutagenesis in mice produces diverse cancers with high mutational load

Hao-Dong Li,[1] Ileana Cuevas,[1] Musi Zhang,[1] Changzheng Lu,[1] Md Maksudul Alam,[1] Yang-Xin Fu,[1] M. James You,[2] Esra A. Akbay,[1] He Zhang,[3] and Diego H. Castrillon[1]

[1]Department of Pathology and Simmons Comprehensive Cancer Center, University of Texas Southwestern (UTSW) Medical Center, Dallas, Texas, USA. [2]Department of Hematopathology, Division of Pathology and Laboratory Medicine, University of Texas MD Anderson Cancer Center, Houston, Texas, USA. [3]Lyda Hill Department of Bioinformatics, UTSW Medical Center, Dallas, Texas, USA.

Mutations underlie all cancers, and their identification and study are the foundation of cancer biology. We describe what we believe to be a novel approach to mutagenesis and cancer studies based on the DNA polymerase ε (POLE) ultramutator phenotype recently described in human cancers, in which a single amino acid substitution (most commonly *P286R*) in the proofreading domain results in error-prone DNA replication. We engineered a conditional *Pole^P286R* allele in mice. *Pole^P286R/+* embryonic fibroblasts exhibited a striking mutator phenotype and immortalized more efficiently. *Pole^P286R/+* mice were born at Mendelian ratios but rapidly developed lethal cancers of diverse lineages, yielding the most cancer-prone monoallelic model described to date, to our knowledge. Comprehensive whole-genome sequencing analyses showed that the cancers were driven by high base substitution rates in the range of human cancers, overcoming a major limitation of previous murine cancer models. These data establish polymerase-mediated ultramutagenesis as an efficient in vivo approach for the generation of diverse animal cancer models that recapitulate the high mutational loads inherent to human cancers.

Introduction

Eukaryotic DNA replication is astoundingly accurate, with an error rate of only 1 per 10^9 to 10^{10} bases, or approximately 1 mutation per cell division. Replication proceeds at DNA forks, where leading-strand and discontinuous lagging-strand synthesis places different demands on the polymerases responsible for DNA synthesis. The DNA polymerases δ (Pol δ) and ε (Pol ε), which replicate the bulk of the nuclear genome, act on opposite DNA strands at replication forks. Pol δ is the principal lagging-strand replicase, while Pol ε replicates the leading strand (1). Pol ε is a heterotetramer composed of 3 accessory subunits and a single catalytic subunit encoded by the *POLE* locus (2–4).

The remarkable fidelity of DNA replication by Pol δ and Pol ε is determined by 3 mechanisms acting in series. First, nucleotide incorporation into the nascent strand is highly accurate. If an incorrect nucleotide enters the nucleotide-binding pocket of the polymerase, the nucleotide leaves the pocket. The intrinsic nucleotide fidelity of the replicative DNA polymerases is approximately 1 error per 10^4 to 10^5 nucleotides (5, 6). Second, Pol δ and Pol ε are unique among mammalian nuclear DNA polymerases, in that they contain 3′ exonuclease ("proofreading") domains that remove most misincorporated nucleotides (3, 7, 8). This endogenous proofreading activity lowers the error rate to approximately 1 error in 10^6 to 10^7 nucleotides (4, 9). Third, errors escaping proof-reading are usually repaired via postreplicative mismatch repair (MMR) pathways, whereby base-base mismatches are recognized, excised, and repaired (10). The remarkably low intrinsic DNA replication error rate underscores the relentless selective pressure maintaining high overall DNA replication fidelity. And yet the residual error is a fundamental driver of human cancer, with strong selection for rare mutations that promote genome instability and accelerate the acquisition of mutations by diverse mechanisms, i.e., the mutator phenotype (11, 12).

Although mutation is the basis of all cancer, mutation incidences vary by orders of magnitude among individual cancers. Most carcinomas have base substitution rates in the range of 1 to 100 per million bases (Mb). MMR-deficient carcinomas are at the high end of this range (≥10/Mb), termed hypermutation. Cancers associated with chronic mutagen exposure, such as lung cancer and melanoma, also have high base substitution rates (13). Much more recently, cancers with even higher base substitution rates of 100/Mb or higher (ultramutation) have been attributed to somatically acquired *POLE* missense mutations leading to single amino acid substitution in the proofreading domain (e.g., P286R or V411L) (14). *POLE*-driven ultramutation represents a recently described mutator phenotype that is common among endometrial and colorectal cancers but also occurs in diverse carcinomas, sarcomas, and hematopoietic malignancies. Some *POLE*-driven cancers have base substitution rates in the hypermutated range (10–100 Mb), while some MMR-deficient cancers have base substitution rates of greater than 100/Mb. Thus, there is overlap with respect to base substitution rates among MMR-deficient and *POLE*-driven human cancers. Analogous exonuclease domain mutations in *POLD1* (the catalytic subunit of Pol δ) occur less commonly and in fewer cancer types than do those in *POLE* (15).

The biochemical impact of the recurrent POLE amino acid substitutions such as P286R is not well understood, but the result is misincorporation of nucleotides during DNA replication, with an error rate much higher than would result from mere inactivation of the exonuclease domain (16–18). The frequency of POLE amino acid substitutions among ultramutated human cancers correlates with the strength of the resulting mutator phenotype; P286R, the most common substitution, results in the strongest mutator phenotype. The $POLE^{P286R}$ mutation behaves in a genetically dominant manner in cancers, as the WT allele is retained (16). This genetic dominance reflects the fact that $POLE^{P286R}$ is a neomorphic mutation that does not result merely in gain or loss of function but produces a new function not intrinsic to the normal allele (active misincorporation of nucleotides) (19, 20). The overall replication error rate of $POLE^{P286R/+}$ cells is thus driven by the presumptive 50% of Pol ε holoenzymes harboring P286R.

Genetically engineered mice are essential tools for the study of cancer. Typically in such models, 1–3 mutations that occur in a human cancer are recapitulated via conditional or other genetic approaches. While such models have proven remarkably effective for studies of diverse aspects of oncogenic signaling pathways to carcinogenesis, there has also been growing concern that genetically engineered mouse models fail to recapitulate essential aspects of human tumor biology. Perhaps foremost among these is the fact that mouse models of cancer—e.g., *Egfr*-, *Kras*-, or *Myc*-driven models of lung cancer—have dramatically lower average mutational frequencies than do human lung carcinomas. The same limitations are likely to be encountered with cancer models based on newer genome-editing methods (21), pointing to the need for different approaches to optimize mouse models with respect to mutational load, which defines many aspects of tumor biology, clinical behavior, and treatment response (22–24).

In this study, we recapitulate $POLE^{P286R}$-driven ultramutation in mice with a conditional knockin allele to permit control of the initiation of mutagenesis. Activation of the allele was efficient and resulted in striking phenotypes in both embryonic fibroblasts and live mice. The incidence and spectrum of malignancies point to $Pole^{P286R}$ as the most potent oncogenic driver mutation described to date, to our knowledge. Whole-genome sequencing (WGS) showed that the cancers harbored base substitutions in the range of up to 100/Mb and in the range of human carcinomas, including hypermutant and ultramutant carcinomas (25). We believe our approach represents a new, efficient, and widely applicable route to the creation of mouse models that recapitulate the mutational loads inherent to human cancer.

Results

Generation of a conditional LSL-$Pole^{P286R}$–knockin allele. The human *POLE* and mouse *Pole* genes both contain 49 exons encoding proteins of 262 kDa (Figure 1A). Proline 286 is encoded by exon 9, and the 10 amino acids flanking this residue are perfectly conserved in mice and humans (Figure 1B). An 8.0-kb mouse *Pole* genomic fragment spanning exon 9 was cloned; the c.857C>G point mutation was introduced by site-directed mutagenesis, and a *Lox-Stop-Lox* (*LSL*) cassette (26) was inserted into a native intronic *XhoI* site to generate the targeting construct. Following targeting of embryonic stem (ES) cells and the births of chimeric animals, germline trans-

mission was achieved and a LSL-$Pole^{P286R/+}$ mouse line was established. The LSL cassette prevents the expression of active protein; Cre-mediated recombination excises the LSL cassette, resulting in $Pole^{P286R}$ expression (Figure 1C). LSL-$Pole^{P286R/+}$ heterozygous mice were healthy and fertile. We developed a multiplex genotyping protocol to distinguish between the +, LSL-$Pole^{P286R}$, and $Pole^{P286R}$ alleles (Supplemental Figure 1A; supplemental material available online with this article; https://doi.org/10.1172/JCI122095DS1). cDNA sequencing with intron-spanning primers confirmed equal expression of the mutant and WT alleles at the RNA level, as well as absent expression from the LSL-$Pole^{P286R}$ allele (Supplemental Figure 1B).

$Pole^{P286R/+}$ mouse embryonic fibroblasts immortalize more efficiently. Mouse embryonic fibroblasts (MEFs) are useful for the characterization of cell growth and immortalization phenotypes (27). To study how $Pole^{P286R}$ influences these fundamental cell-growth properties, LSL-$Pole^{P286R/+}$ heterozygous males were bred with heterozygous *Ddx4-Cre*$^{/+}$ females. *Ddx4-Cre* is expressed only in germ cells and induces global recombination of floxed loci within oocytes, without mosaicism in 100% of the progeny, even in progeny that do not inherit *Ddx4-Cre* (maternal effect) (28). $Pole^{P286R/+}$ and sibling $Pole^{+/+}$ (hereafter referred to as +/+) embryos were harvested at E13.5 and used to establish MEF lines via 3T3 assays, in which 3×10^5 cells were serially passaged and their growth kinetics (i.e., cell doublings) measured every 3 days. The culture of murine fibroblasts induces p16^{INK4a} and p19^{ARF}, causing cell-cycle arrest in most cells within a few passages. The immortalization of WT MEFs requires serial passage and stochastic mutational events leading to p53 or p19^{ARF} loss, among others (27). Initially, $Pole^{P286R/+}$ and +/+ MEFs proliferated at similar rates, but the $Pole^{P286R/+}$ MEFs bypassed tissue culture–induced senescence earlier (Figure 2A). This could also be clearly seen for a subset of 6 MEF lines obtained from the same litter (Figure 2B). By passage 20 (P20), $Pole^{P286R/+}$ MEFs had undergone an average of 26.2 doublings versus 15.6 for +/+ MEFs (Figure 2A, $P = 0.005$, unpaired Student's t test). Also, all $Pole^{P286R/+}$ MEFs were successfully immortalized (10 of 10), whereas this was not the case for +/+ MEFs (7 of 10) (Figure 2A). The earlier, more frequent, and consistent immortalization of $Pole^{P286R/+}$ MEFs was suggestive of an increased incidence of immortalizing mutations (see below). Of note, we found that markers of DNA damage were not elevated, indicating that the underlying process did not trigger DNA breaks or a general DNA damage response (Figure 2C).

We then performed a simple forward genetic screen to further explore these inferences. MEFs from P20 were subjected to 6-thioguanine (6-TG) selection. Within cells, 6-TG is converted by hypoxanthine-guanine phosphoribosyltransferase (HPRT) into cytotoxic metabolites. The *HPRT* gene is X-linked, and treatment with 6-TG selects for rare cells with inactivating *HPRT* mutations. Following standard determinations of plating efficiencies, male MEF lines were subjected to 6-TG selection. Consistent with early passage, the +/+ MEFs failed to give rise to resistant colonies, whereas 6-TG–resistant colonies arose consistently among $Pole^{P286R/+}$ MEFs (10^6 per plate, $n = 5$ replicates per line) (Supplemental Figure 2A). Cloning and sequencing of *HPRT* cDNAs from the resistant colonies revealed a number of missense mutations and also some deletions (Supplemental Figure 2B).
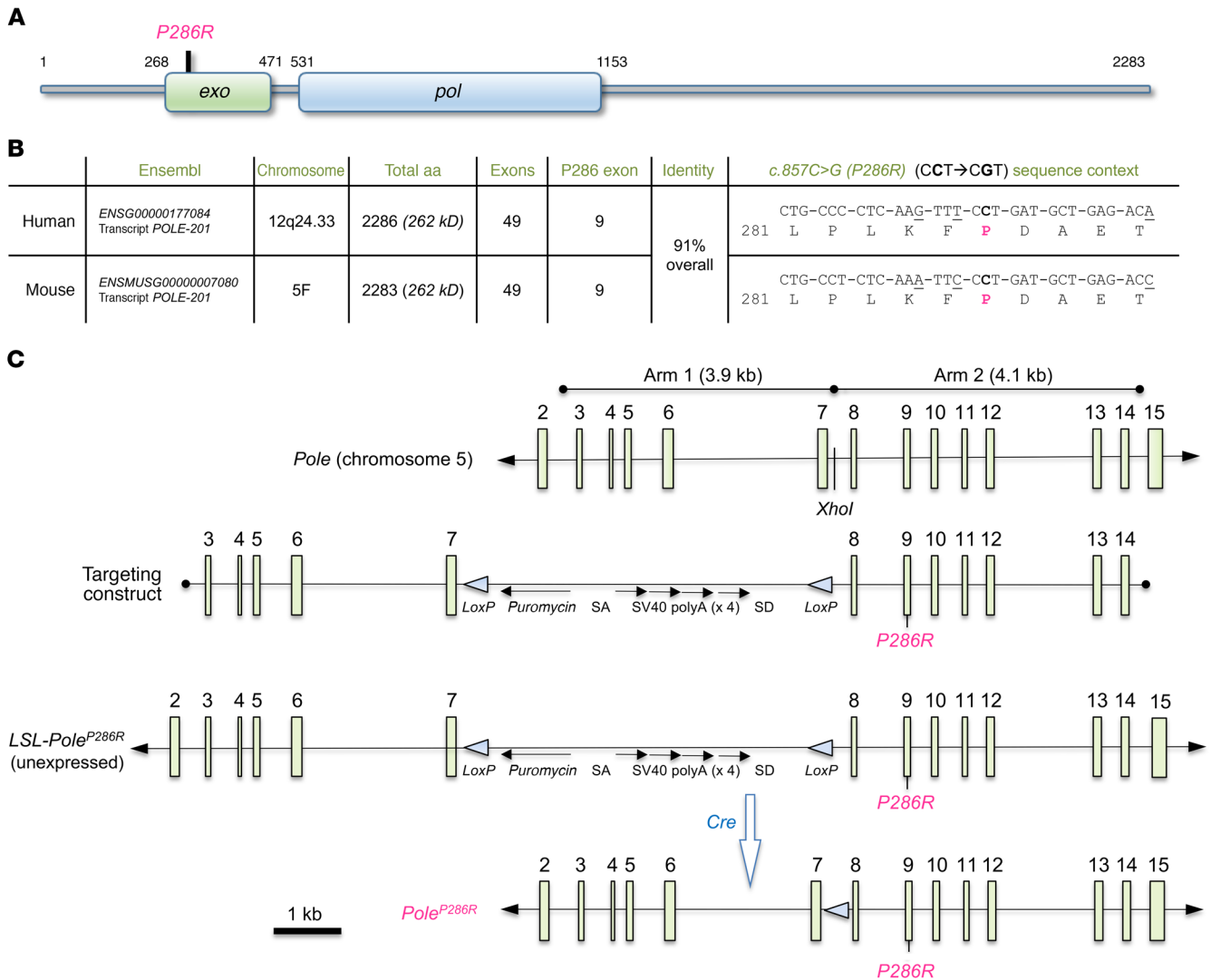
Figure 1. Generation of the LSL-Pole^P286R conditional knockin allele. (**A**) POLE protein domain structure and position of *P286R*. The C-terminal half of the protein is not essential for polymerase activity but serves as a protein-protein interaction domain. *pol*, polymerase domain; *exo*, exonuclease domain. (**B**) Human and mouse gene information and sequence context for the *c.857C>G* p.pro286arg (*P286R*) mutation. *c.857C>G* p.pro286arg is equivalent in humans and mice (i.e., the position of the base and amino acid are the same in human and mouse cDNA and polypeptide reference sequences). (**C**) Schematic showing a portion of the mouse *Pole* locus used for generating the targeting construct. The *XhoI* site was used for insertion of the *LSL* cassette.

This experiment (together with the data below) demonstrates that *Pole^P286R* leads to an increase in the overall mutation rate, with functional repercussions.

Polymerase-mediated ultramutagenesis in mice elicits malignant cancers of diverse lineages. To attempt to study the impact of *Pole^P286R* across all cell lineages in vivo, cohorts of *Pole^P286R/+* and +/+ sibling controls were generated by breeding *LSL-Pole^P286R/+* males with *Ddx4-Cre* females. *Pole^P286R/+* heterozygous mice were born at the expected Mendelian ratios, and the animals appeared healthy, demonstrating that expression of *Pole^P286R* in all cells of a living organism does not interfere with essential physiologic functions and is well tolerated, at least for a period of time. To identify age-related phenotypes, *Pole^P286R/+* mice (*n* = 50) were permitted to age naturally. While there were no deaths in the control cohort (*n* = 22) past 350 days, *Pole^P286R* mice exhibited striking age-dependent mortality. The first deaths occurred at only 72 days of age, and the

median age at death was 138 days, with all animals succumbing by 293 days (*P* < 0.00001 +/+ vs. *P286R^/+*, Figure 3A).

The animals died of cancers (Figure 3, B–G). The cancer-prone phenotype was remarkable in that at least 1 malignant neoplasm was identified in all but 1 *Pole^P286R/+* mouse, and most had multiple distinct malignancies of diverse lineages, even within the same organ. The most common malignancies were lung adenocarcinomas and aggressive CD3+CD19− thymus-based T cell lymphomas that typically infiltrated the lungs and were widely disseminated, with some animals in the leukemic phase and with bone marrow infiltration (Figures 3 and 4). Other malignancies included diverse sarcomas, most commonly angiosarcomas, and diverse carcinomas including neuroendocrine, mammary gland, uterine, colon, and squamous cell carcinomas (SCCs), among others (Figures 3 and 4, and Supplemental Table 1A). The vast majority of tumors were highly invasive and obviously malignant, but some tumors
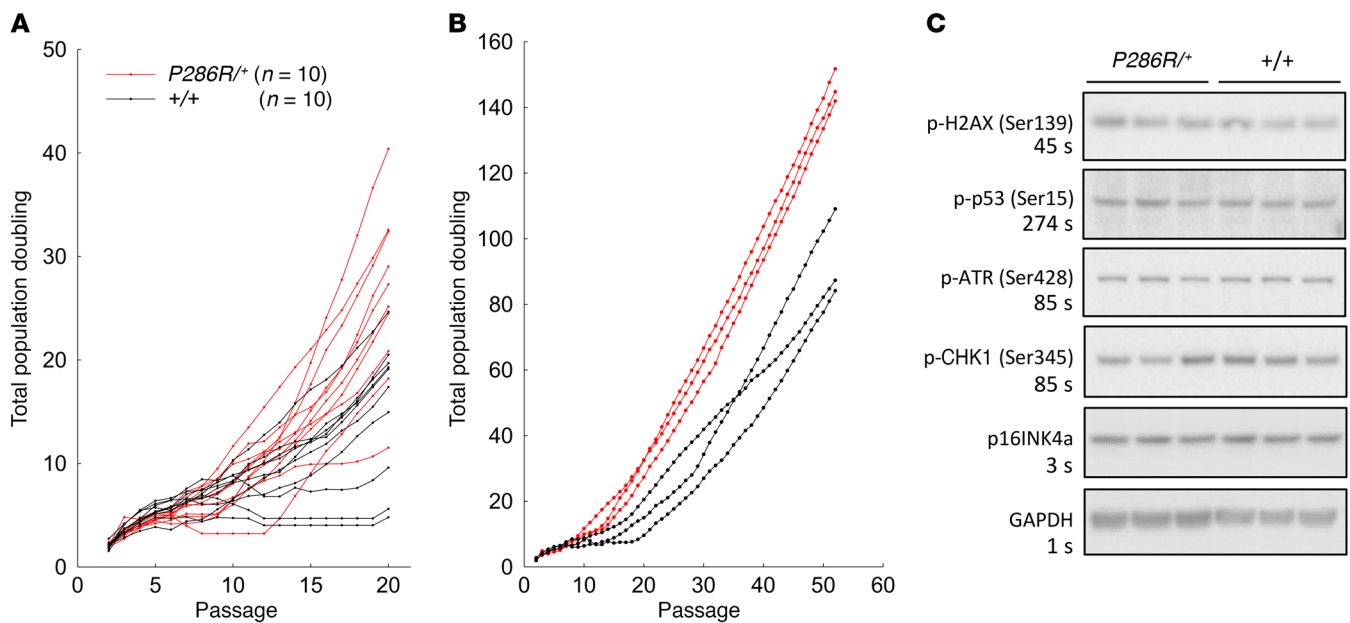
Figure 2. Efficient immortalization and absence of DNA damage responses in *Pole^P286R/+* MEFs. (**A**) 3T3 assays of 10 *Pole^P286R/+* and 10 control +/+ MEF lines. (**B**) 3T3 assays of a select subset of MEF lines (embryos from a single litter) performed up to P52. (**C**) Western blot analysis of DNA damage markers and p16INK4a performed at P15, showing no constitutional abnormalities in *Pole^P286R/+* MEFs. Each lane corresponds to one of the MEF lines shown in **B**. Exposure durations are shown for each panel; long exposure durations were used for some markers to permit the visualization of bands and lack of differences between experimental and control samples. p-, phosphorylated.

appeared benign (e.g., hemangiomas; see Figure 5E for 1 example and Supplemental Table 1A). Some mice also exhibited dysplastic (in situ) epithelial precancers, e.g., in the colon and mammary gland (Figure 5B), but these were not systematically analyzed. In total, we identified 97 tumors in 47 necropsied mice (average of 2.1 tumors/mouse), with mice showing as many as 8 distinct malignant neoplasms. These results show that (a) *Pole^P286R* is active across cell lineages, (b) *Pole^P286R* acts in a genetically dominant manner in mice, (c) *Pole^P286R* alone efficiently drives the formation of cancers of diverse epithelial, hematopoietic, and mesenchymal lineages, with the result that (d) *Pole^P286R/+* is among the most tumor-prone animal models described to date, to our knowledge — which is all the more remarkable for being based on a monoallelic point mutation leading to a single heterozygous amino acid substitution.

The presence of the *LSL* cassette renders the *LSL-Pole^P286R* allele functionally null with respect to *Pole* function. *Pole^P286R LSL-Pole^P286R* mice would thus be genetically hemizygous and express only *Pole^P286R* protein. To determine whether such mice might be viable, we conducted crosses between *Pole^P286R/+* and *LSL-Pole^P286R/+* mice. Of 164 live-born progeny, only 2 were *Pole^P286R LSL-Pole^P286R* versus the expected 41 ($P < 0.00001$), indicating embryonic lethality. Thus, *Pole^P286R* hemizygosity resulted in a more severe phenotype than did heterozygosity, although rare escaper animals survived. The hemizygous live-born animals were minute and rapidly succumbed (both at 66 days of age) to aggressive T cell lymphomas, with each mouse also harboring multiple independent malignancies ($P < 0.00001$, *P286R/LSL* vs. *P286R/+*; Figure 3A and Supplemental Table 1B). Consistent with prior studies in yeast, these results show that the P286R amino acid substitution does not block DNA replication per se, but imply that the correspondingly higher levels of mutation asso-

ciated with hemizygosity are incompatible with normal organismal development (16).

WGS of 20 samples reveals a high frequency of base substitutions in the range of human cancers. First, we performed WGS at an average depth of 21.7× for the 6 MEF lines shown in Figure 2B; representative quality control metrics are shown in Supplemental Figure 3. SNPs present in the parental mouse strains, which were independently sequenced ($n = 2$ samples per the 2 parental [FVB/n and 129/SvEvTac] strains), as well as known mouse SNPs were filtered out (29). To facilitate the accurate identification of de novo mutations in the setting of very high anticipated clonal variation and low allelic frequencies for the large number of variants, DNA was obtained from colonies cloned from single flow-sorted MEF cells. Such single-cell cloning of MEFs is possible only following immortalization, which occurred by P15 in the +/+ and *Pole^P286R/+* MEF lines. At P15, *Pole^P286R/+* MEF clones consistently showed high overall base substitution rates relative to those of WT controls (Figure 6A). To estimate the mutation rate per cell division, the experiment was repeated with P30 *Pole^P286R/+* clones from the same MEF lines. The P30 clones consistently showed higher base substitution rates. The *Pole^P286R/+* mutation rate was then "clocked" by calculating the mutation rate per cell division as determined by the 3T3 assays, which measure the number of cell doublings at each passage (Figure 2). This analysis yielded an estimate of 1.6 nucleotide substitutions per Mb per cell division, at least 3 orders of magnitude higher than the intrinsic DNA replication error rate (Figure 6B).

We also conducted WGS on 7 *Pole^P286R/+* primary cancers, including 3 lung adenocarcinomas, 3 T cell lymphomas, 1 cutaneous SCC, and 4 cancers from the 2 *Pole^P286R LSL-Pole^P286R* (hemizygous) mice. Despite the presence of multiple tumors in some mice, the tumors could be inferred to be clonal, because they arose within otherwise
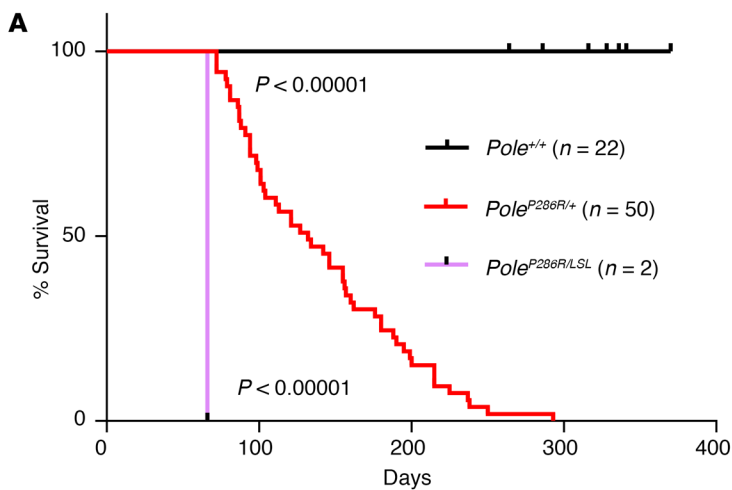
Figure 3. *Pole^P286R/+* mice rapidly succumb to diverse cancers. (**A**) Survival analysis of 50 *Pole^P286R/+*, 22 sibling control (+/+) mice, and 2 *Pole^P286R LSL-Pole^P286R* mice. *P* values were determined by log-rank test. (**B–G**) Examples of malignant neoplasms. (**B**) Large thymic T cell lymphoma expanding the chest cavity, with infiltration into the lungs (lg) and below the heart (h). (**C**) Angiosarcoma of the hind leg, with hemorrhage. (**D**) Angiosarcoma of the shoulder (different mouse from that shown in **C**). (**E**) Histiocytic "sarcoma" involving 1 ovary. These tumors are not true sarcomas, but rather malignant hematopoietic neoplasms with histiocytic differentiation. The tumor was disseminated and present in other tissues. (**F**) Aggressive SCC of the snout. (**G**) Adenocarcinoma of the lung. Scale bars: 1 cm.
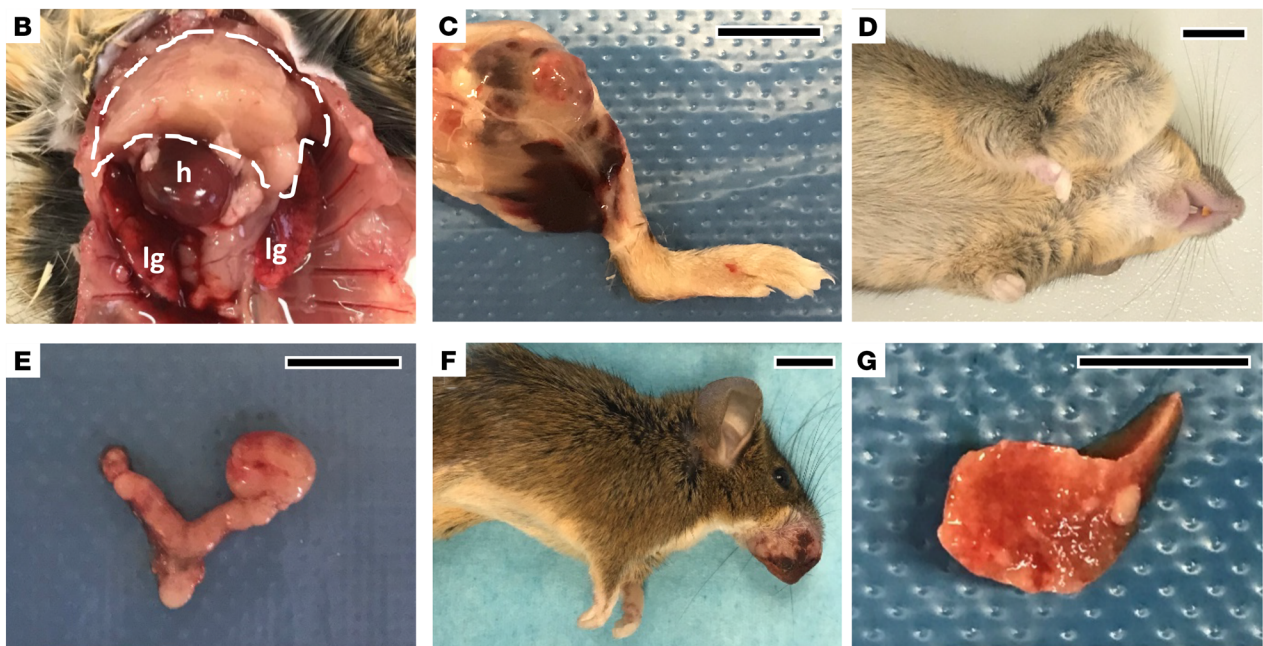
histologically normal tissues, demonstrating that only very rare cells became neoplastic. Neoplastic transformation in this model can be anticipated to require unique and statistically improbable combinations of mutations in oncogenes and/or tumor suppressors. Thus, although clonal variation is expected to continue unabated, each tumor should contain a large number of mutations identifiable by WGS, as with human *POLE*–driven cancers, which are also clonal (15). WGS revealed a very high frequency of mutations, in the range of 10 to 100/Mb (Figure 6C). Notably, this is in the range for human malignancies such as lung cancers (11, 15). In contrast, lung carcinomas from diverse genetically engineered mouse models (*Trp53/Rb1*-, *Egfr*-, *Myc*-, and *Kras*-driven) of adenocarcinoma have consistently shown exceedingly low mutation burdens (~50 times lower than human lung cancers) (22–24). Thus, to our knowledge, *Pole^P286R/+* is the first genetically engineered mouse model to recapitulate the high mutation burden and therefore the clonal variation and heterogeneity intrinsic to human cancer.

The most common mutations were of the missense type, with significant numbers of nonsense and splicing mutations. Indels

were correspondingly rare (Figure 7), and *Pole^P286R* did not produce detectable large-scale rearrangements in most tumors (Supplemental Figure 4). We then tabulated base substitutions in the context of flanking bases to determine whether there was a recurring *Pole^P286R*-related mutational signature with respect to trinucleotide context (Supplemental Figure 5). *Pole^P286R/+* cancers and MEFs showed distinct and consistent signatures across samples, with a high incidence of C>A and C>T substitutions and correspondingly rarer C>G and T>A substitutions. Of note, C>A substitutions occurred most frequently in a TCT flanking base context, as in *POLE*-mutant human cancers, although our data show a greater dependence on the 3′ base (i.e., a higher incidence of substitutions in NCT than in TCN trinucleotides). We also noted that T>G substitutions occurred with a bias toward a TTT context in some samples, as previously observed for *POLE*-driven human cancers, but with a bias against this context in other samples; the biological basis of these differences is uncertain. In contrast, the +/+ control MEF clones did not have this signature (Supplemental Figure 5). Thus, we document a mutational signature in murine *Pole^P286R/+* MEFs and tumors of diverse lin-
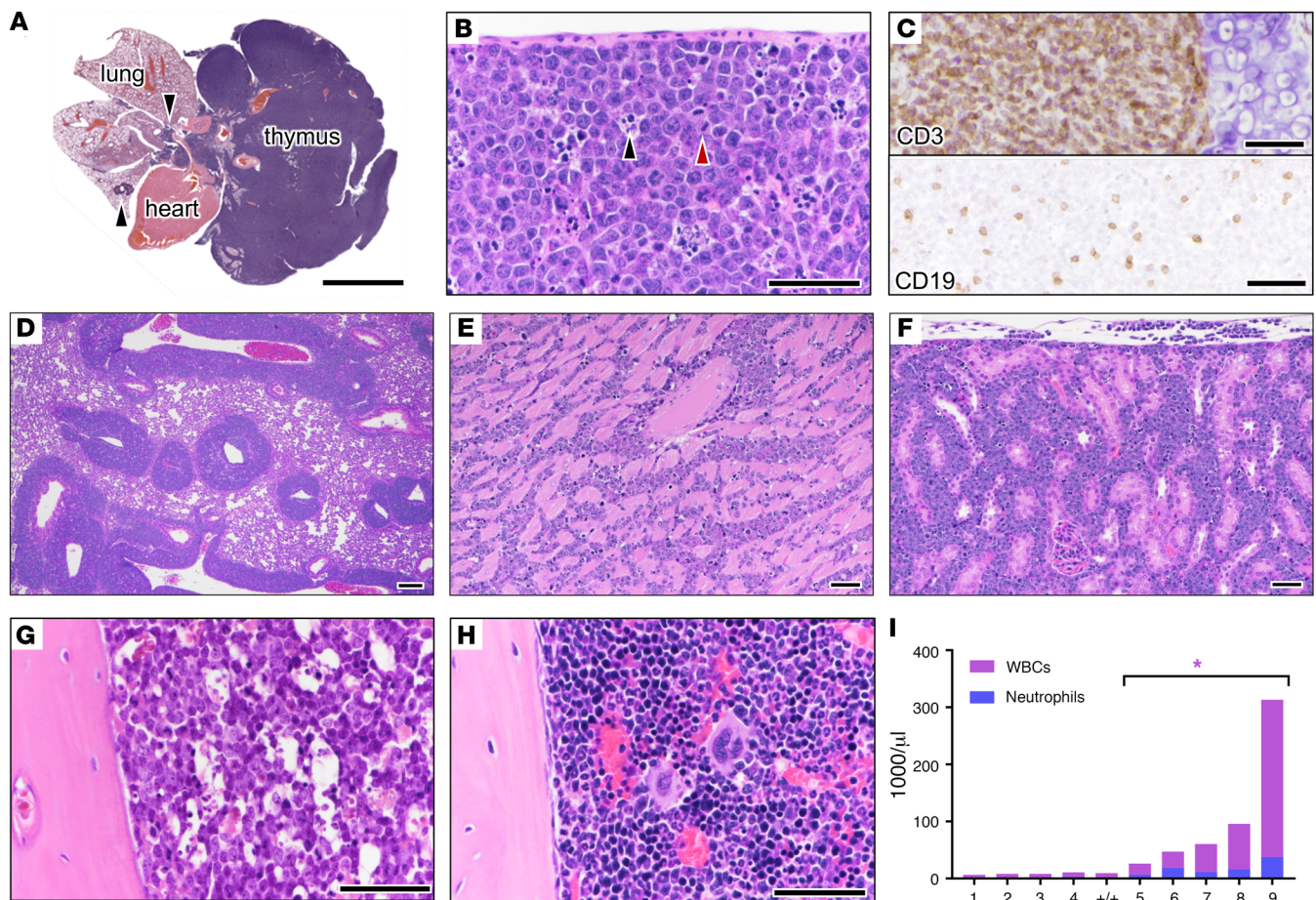
Figure 4. Characterization of thymus-based high-grade T cell lymphomas in *Pole^{P286R/+}* mice. (**A**) Low-magnification scan of H&E-stained slide showing large thymus-based T cell lymphoma, with infiltration between the lobes of the lung and pericardial and lung parenchymal infiltration (arrowheads). (**B**) Lymphomas had high-grade features, numerous mitoses (red arrowhead), and apoptotic bodies (black arrowhead). (**C**) IHC of paraffin-embedded, formalin-fixed tissue shows that tumor cells were CD3⁺CD19⁻. Admixed are rare CD19⁺ B lymphocytes. (**D**–**F**) Different mouse tissues illustrating an aggressive T cell lymphoma phenotype. (**D**) Extensive infiltration into pulmonary parenchyma in a characteristic perivascular pattern. (**E**) Infiltration within myocardium. (**F**) Infiltration within kidney. (**G**) Femur showing extensive infiltration. (**H**) Femur (bone marrow) from a +/+ mouse showing normal trilineage maturation for comparison. (**I**) CBCs of 9 mice with obvious thymic lymphomas at necropsy, also shown is 1 reference (+/+) sample. Asterisk indicates mice in the blast phase; their neutrophil counts were only mildly elevated, showing that leukocytosis was not due to sepsis/neutrophilia. Scale bars: 5 mm (**A**) and 50 μm (**B**–**H**).

eages that resembles signature 10, which was previously described in *POLE* exonuclease domain–mutant tumors (30–32).

This signature should skew toward specific codon substitution patterns. Indeed, as previously reported for human *POLE^{P286R}* tumors, arginine was the most frequently substituted amino acid, and replacements by nonsense codons (most frequently from glutamate or arginine) were among the most frequent codon substitutions (Supplemental Table 2 and ref. 30). However, as stated above, missense mutations were the most common subtype of mutation in *Pole^{P286R}* tumors (Figure 7). These findings reveal that *Pole^{P286R}* is particularly efficient at generating missense and nonsense mutations capable of inactivating tumor suppressor loci, although in some cases, nonsense mutations can also turn on protein activity (i.e., by removing regulatory C-terminal domains). Inspection of mutations in the WGS data of cancer driver genes selected on the basis of their relevance to the observed tumors (*Notch1*, *Kras*, *Trp53*, and *Pten*) illustrated this point. For example, all *Notch1* mutations occurred in the T cell lymphomas, consistent

with the presence of frequent *Notch1* mutations in human T cell acute lymphoblastic leukemia. One of the mutations was a truncating *c.7395C>A* p.Tyr2465→STOP mutation, a well-known hot spot for *Notch1* hyperactivating mutations that remove the negative regulatory C-terminal PEST domain (33). Truncating mutations were also identified in *Pten* and *Trp53* (Supplemental Table 3). Canonical *Kras* activating mutations *c.182A>G* p.Gln61Arg and *c.436G>A* p.Ala146Thr were also observed in 2 of the cancers, further emphasizing that *Pole^{P286R}* drives carcinogenesis through gain-of-function mutations in oncogenes as well as inactivating loss-of-function mutations in tumor suppressors (34).

Discussion

This study establishes polymerase-mediated ultramutagenesis as a genetic approach to the generation and study of cancers in animal models. The anticipated deleterious effect of *Pole^{P286R}* on DNA replication fidelity called for a conditional strategy. We engineered the conditional knockin allele *LSL-Pole^{P286R}*, in which expression of
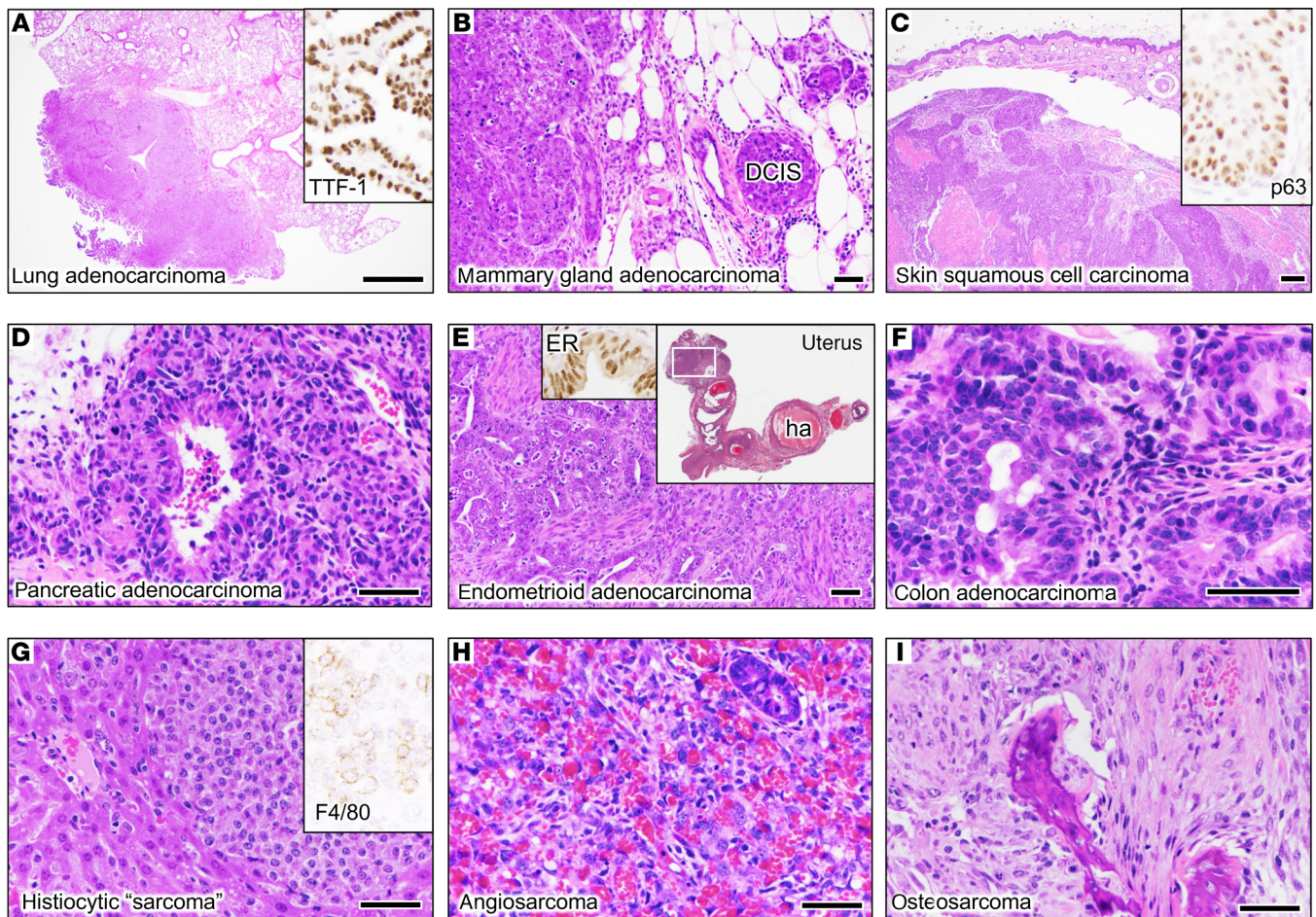
Figure 5. Characterization of other lethal malignancies observed in *Pole^P286R/+* mice. (**A**) Lung adenocarcinoma with extensive infiltration (inset) and invasion through pleura. (**B**) Mammary gland carcinoma, with focus of ductal carcinoma in situ (DCIS) and invasive carcinoma (left). (**C**) Skin with invasive SCC, ranging from well differentiated to poorly differentiated. (**D**) Pancreatic adenocarcinoma; area shown is well differentiated. (**E**) Invasive endometrioid adenocarcinoma (left); inset shows low-magnification scan (actual size) of H&E-stained slide demonstrating endometrial adenocarcinoma in the left uterine horn and a benign hemangioma (ha) on the right horn. The endometrioid adenocarcinoma invaded through myometrium and uterine serosa into adjacent structures such as oviducts. (**F**) Colonic adenocarcinoma. (**G**) Histiocytic "sarcoma" in the liver (of hematopoietic origin/histiocytic differentiation). This tumor was widely disseminated and replaced more than 50% of the liver. (**H**) Angiosarcoma of the colon; entrapped colonic gland is visible in the upper right-hand corner. (**I**) Osteosarcoma with mineralized osteoid matrix. IHC stains are shown as insets in **A**, **C**, and **G** for selected lineage-specific markers confirming histotypes (TTF-1, ER, p63, and F4/80; ×50 magnification for all). Scale bars: 200 μm (**A** and **C**), 50 μm (**B** and **D**–**I**).

Pole^P286R is blocked by the *LSL* cassette (26). Cre-mediated recombination efficiently deleted the cassette, converting the *LSL-Pole^P286R* allele to *Pole^P286R*, resulting in *Pole^P286R* expression at the same level as the WT allele. The *LSL-Pole^P286R* allele is functionally null with respect to *Pole* function; *LSL-Pole^P286R/+* mice thus only had 1 functional *Pole* allele, but were nonetheless healthy and fertile. This result is consistent with the rare human hereditary condition known as FILS syndrome (<u>f</u>acial dysmorphism, <u>i</u>mmunodeficiency, <u>l</u>ivedo, and <u>s</u>hort stature) described in a consanguineous kindred. Affected individuals carry a homozygous nucleotide substitution that strongly affects *POLE* splicing, leading to a 90% decrease in overall POLE protein levels. However, heterozygous carriers are asymptomatic, demonstrating that POLE protein levels approximately 50% below normal do not have a discernable physiologic impact (35).

A mouse *Pole*–knockin allele (*Pole^e*) has been described previously, in which the conserved and functionally essential D272 and

E274 residues in the 3′ exonuclease (proofreading) domain were replaced with alanines (36). These substitutions preserve normal polymerase activity, but selectively inactivate the proofreading function. Heterozygous *Pole^e/+* mice did not exhibit any phenotypes or decreased longevity, demonstrating that heterozygous loss of *Pole* proofreading activity has minor functional effects, if any. In contrast, homozygous *Pole^e/e* mice had a median survival of 16 months and progressively succumbed to cancers beginning at 9 months of age. The most frequent tumors were intestinal adenomas/adenocarcinomas (45% incidence), followed by histiocytic sarcomas (36%), lymphomas (24%), and lung adenocarcinomas (12%) (36). Although primary tumors were not analyzed, mutation rates in *Pole^e/e* MEF lines were elevated, consistent with a general mutator phenotype due to defective proofreading. Some interesting and notable distinctions can be drawn between *Pole^P286R* and *Pole^e*. First, *Pole^P286R* is a genetically dominant neomorphic allele (19, 20), whereas *Pole^e* is recessive (loss of func-
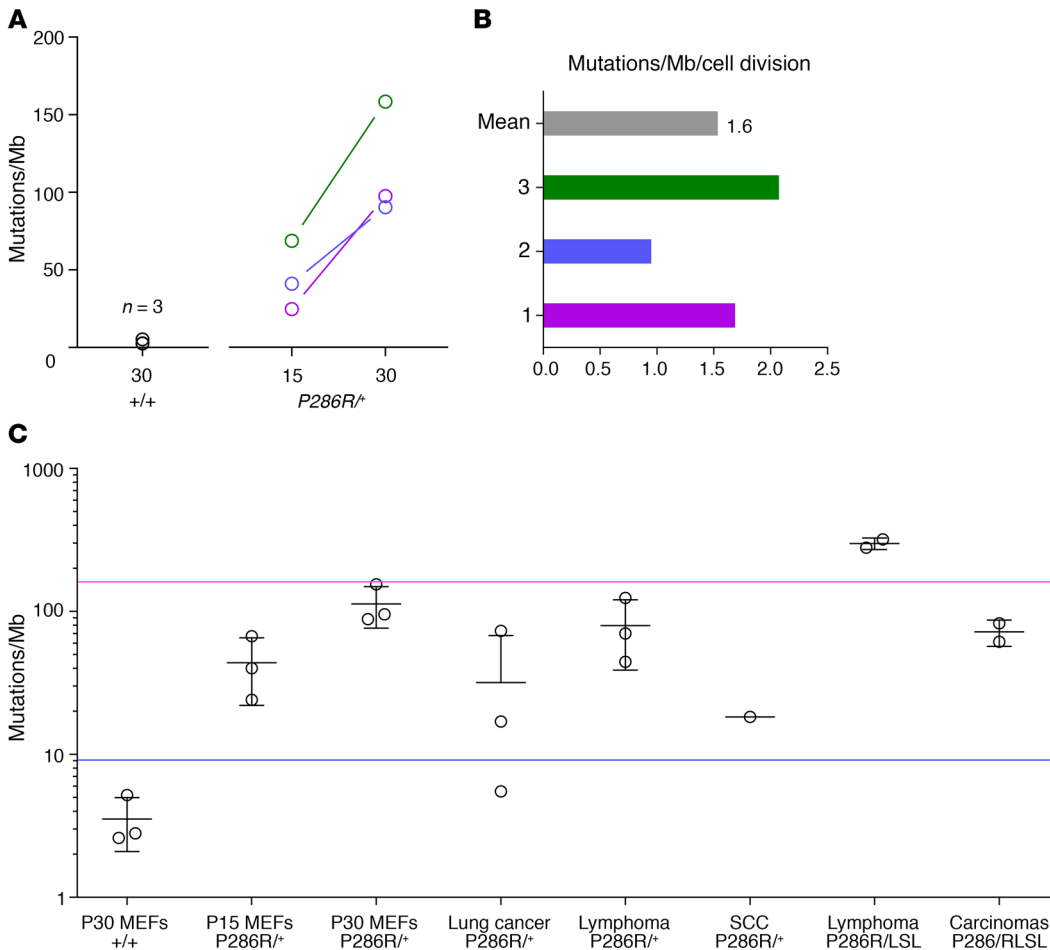
Figure 6. WGS analyses of 20 samples to determine genome-wide base substitution rates expressed as mutations per Mb. (**A**) WGS analysis of 9 MEF clones (same lines as shown in Figure 2B). DNA was prepared from single cell–derived clones isolated at P30 for the three +/+ MEF lines and at P15 and P30 for the three *Pole*$^{P286R/+}$ MEF lines. Clones obtained from the same MEF lines at P15 and P30 are shown in the same color. (**B**) Data in **A** were used to calculate the number of mutations per cell division between P15 and P30. (**C**) Overall base substitution rates for 9 MEF lines shown in **A** and **B**, together with 11 primary tumors from *Pole*$^{P286R/+}$ and *Pole*$^{P286R/LSL}$ mice. The 2 carcinomas from *Pole*$^{P286R/LSL}$ mice correspond to 1 lung adenocarcinoma and 1 cutaneous SCC. The blue and red lines indicate the average mutation rates in human lung and ultramutated cancers, respectively (13, 15).

tion). This simplifies the breeding schema considerably, as only a single allele is needed for strong phenotypic expression. Second, the much greater severity of the heterozygous *Pole*$^{P286R/+}$ phenotype versus the homozygous *Pole*$^{e/e}$ phenotype (median survival of 5 months vs. 16 months) is in keeping with prior studies in yeast showing that *Pole*^{P286R} causes a much stronger mutator phenotype than does mere inactivation of proofreading activity (16, 18), as well as prior analyses of human tumors, in which *Pole*^{P286R} mutations were heterozygous with retention of the WT allele (15, 16, 30).

Another notable feature of the cancers in *Pole*$^{P286R/+}$ mice was their aggressive clinical features and histology. For example, the T cell lymphomas were large, and many were widely disseminated and in the leukemic phase with bone marrow infiltration, a markedly aggressive phenotype for a mouse model of T cell lymphoma (37, 38). Also, few benign tumors were identified, with most tumors exhibiting striking tissue invasion and thus representing obvious malignancies. We anticipate that the *LSL-Pole*^{P286R} allele could be used to generate a wide range of murine cancers through the use of tissue-specific Cre drivers or through other approaches such as the use of Adeno-Cre to recapitulate *Pole* ultramutator phenotypes (i.e., in cancers in which this phenomenon is frequent) or the high mutational loads that are common in many human cancers. It will be of interest to determine whether conditional expression of *Pole*^{P286R} can induce malignant transformation of any cell lineage in the mouse, although the observed exception-

ally rapid onset of a wide spectrum of malignancies encompassing diverse cell lineages (mesenchymal, epithelial, hematopoietic), together with the potent mutator phenotype documented in MEFs, suggests that this will be the case. It is also noteworthy that such mice can be efficiently and rapidly generated in a single generation (i.e., by breeding *LSL-Pole*$^{P286R/+}$ with any Cre-driver line). It should also be possible to combine *LSL-Pole*^{P286R} with other alleles that drive malignant transformation to efficiently create high base substitution rates in any genetically engineered mouse model of cancer.

WGS revealed high base substitution rates, in line with those described for most human carcinomas and with a similar mutational signature previously described for *POLE*-driven human tumors (30), consistent with strong sequence and functional conservation among the human and mouse polymerases. We observed similar mutational signatures among MEF clones and tumors, arguing that the signatures are not dependent on selective pressures unique to tumor progression in vivo. It is also not clear why some samples (including both MEFs and tumors) showed a T>G bias in a TTT context, while others did not, although such variability has been documented in human tumors (15). While a specific mutational signature associated with *POLE* proofreading domain mutations (signature 10) has been described (29–32), it should be noted that the impact of other biological factors (such as chemotherapy or radiation therapy, among many other variables)
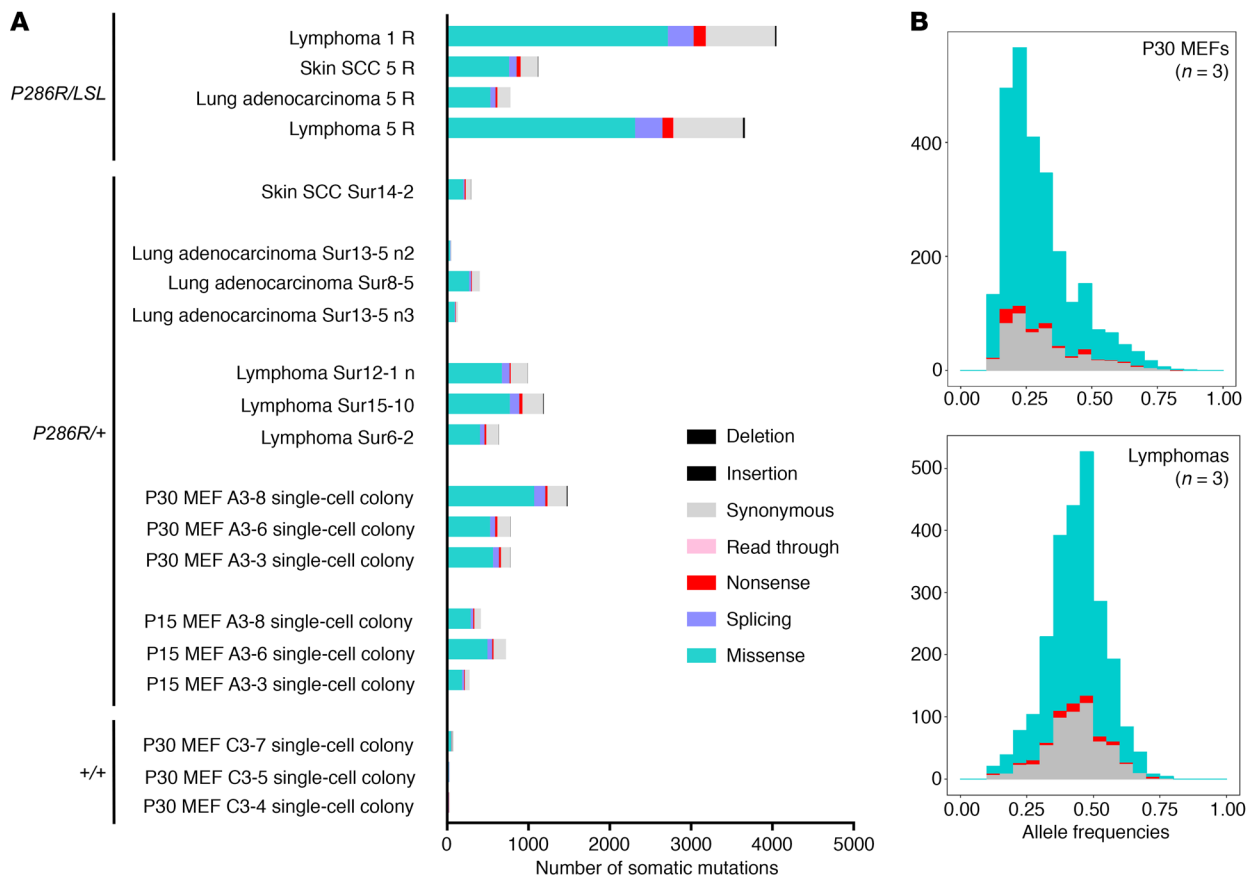
Figure 7. WGS-derived mutation types in *Pole^P286R/+* (heterozygous) MEFs and tumors and (hemizygous) *Pole^P286R/LSL* tumors. (**A**) Deletions and insertions were rare, consistent with a preponderance of single base substitutions. Somatic mutations located within coding sequences were selected. Mutations were assigned to different categories using the SnpEff tool with GENECODE M16 as a reference. The *x* axis represents the number of somatic mutations in each sample. (**B**) Distributions of allele frequencies for representative samples. Mutations were classified into 3 categories: missense, synonymous, and nonsense.

on signature 10 remains incompletely defined. The role of MMR deficiency in *Pole^P286R/+* mouse tumors remains to be formally determined. Although the rapid kinetics of mutation in the MEFs and in vivo argues against the necessity of superimposed MMR deficiency, several studies have found evidence for synergism between MMR deficiency and *POLE* proofreading domain mutations, as could be expected, given that the MMR pathway repairs most DNA replication errors (15, 39–42).

One interesting open question is the impact of this signature on the specific range of mutations that serve as effective cancer drivers in diverse lineages. On one hand, as we and others have shown, *Pole^P286R* is particularly effective at generating stop codons, implying that inactivation of tumor suppressors should be frequent in such tumors (30). Indeed, we identified nonsense mutations in *Notch1*, *Trp53*, and *Pten*. At the same time, diverse amino acid substitutions were also dramatically increased, creating a rich assortment of codon substitutions with potential oncogenic impact. This is illustrated by the 2 *KRAS* activating mutations identified in this study, one in a lung adenocarcinoma (c.182A>G; Q61R) and the other in a T cell lymphoma (c.436G>A; A146T). Both are known oncogenic *KRAS* mutations previously described in human cancer, although Q61R is much more common in human cancers and in carcinogen-induced mouse lung tumors (24). However, while A146T occurs in a favored AGC/GCT context, Q61R occurs in a strongly disfavored

CAA/TTG context, demonstrating that both favored and unfavored substitutions can occur in *Pole^P286R*-driven cancers.

One limitation of classic genetically engineered mouse models of cancer — and one to which increasing attention has been drawn — is their surprisingly low overall mutation frequencies, which are far below what is observed in the vast majority of human cancers (22, 23). These properties make most such models particularly useful for certain studies, such as those of the functional or signaling consequences of specific tumor suppressor or oncogene mutations, but greatly limit their utility for others, such as studies of tumor heterogeneity and clonal variation or the impact of mutational load/base substitution rates on tumor behavior and response to therapy. The latter has become particularly relevant with the advent of immune checkpoint therapies, as there is growing evidence that treatment success correlates with a high incidence of somatic mutations irrespective of tumor type (43). That knockout/knockin mouse models of cancer have not proven particularly useful for testing immune checkpoint therapies probably relates to mutational burdens too low to recapitulate human tumor immunology or trigger robust immune responses following immune checkpoint therapy (44, 45). In this context, *Pole^P286R* could be used to "humanize" a wide range of cancer models with respect to mutational load, opening new avenues of investigation.

Pole^{P286R}-mediated ultramutagenesis may also prove useful for a variety of studies such as forward genetic screens for diverse phenotypes such as drug resistance, either in vivo or in vitro, as illustrated by the 6-TG studies.

In summary, we present an approach to the generation of mouse models of cancer based on the *POLE* ultramutator phenotype recently described in human cancers. Mutations accumulate initially in a biologically silent manner, without promoting cell death or triggering DNA damage responses. Mice expressing heterozygous *Pole^{P286R}* mutations died of multiple aggressive malignancies of diverse lineages, resulting in one of the most cancer-prone animal models described to date. WGS of the tumors revealed very high mutation frequencies, in the range of 10 to 100/Mb described for human malignancies such as lung cancers. This *Pole*-based mouse model is the first genetic model to our knowledge to recapitulate the very high mutational burdens and therefore the clonal variation and heterogeneity intrinsic to human cancer. Our data demonstrate that this approach will enable a wide range of genetic approaches in the study of cancer or other biological processes.

Methods

Generation and validation of the Pole^{P286R(L)} (LSL-P286R) conditional knockin allele. First, a 8.0-kb genomic fragment including exons 3–14 was generated by PCR from 129/SvEvTac DNA using the primers L3 (5′-GGT<u>GTCGAC</u>GATGAGTGATAGAGCAGAGGAAGGCACG-3′) and R3 (5′-G<u>GCGGCCGC</u>GTCTCCAGAATCATTCCACCT-CAAGCATC-3′). *SalI* and *NotI* sites (underlined) were added to the 5′ ends of the L and R primers. PCR was performed with Phusion High Fidelity Taq (New England BioLabs) under the following conditions: initial denaturation at 98°C for 3 minutes followed by 24 cycles of denaturation at 98°C for 1 minute, annealing at 65°C for 30 seconds, and amplification at 72°C for 6 minutes. Following addition of 3′ A overhangs to the gel-purified PCR product by incubating with ATP and Taq polymerase at 37°C for 30 minutes, the 8.0-kb fragment was cloned into pCR2.1TOPO (Invitrogen, Thermo Fisher Scientific). The 8.0-kb insert was Sanger sequenced to confirm the absence of PCR mutations. Second, the *CCT→CGT* mutation resulting in the P286R amino acid substitution (exon 9, underlined, see also Figure 1A) was generated by in vitro mutagenesis with the oligonucleotide 5′-CCAAACTGCCTCTCAAATTC<u>CGT</u>GATGCTGAGACCGAT-CAG-3′ using the QuickChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies). Site-specific mutagenesis and absence of other undesired mutations were confirmed by Sanger sequencing of all exons. Third, the *Pole* genomic fragment containing the mutation was gel purified following digestion with *SalI* and *Not* and then subcloned into the pKOII targeting construct (contains a pGKNeo positive selection cassette flanked by frt sites and a diphtheria toxin negative selection cassette) digested with *XhoI* and *NotI* (46). This procedure removes the pGKNeo cassette from pKOII, which is not needed, as this functionality is provided by LSL (see below); however, the pKOII diphtheria toxin negative selection marker is retained. Fourth, the LSL cassette (5.8 kb) was gel purified following *XhoI* digestion of plasmid LSL TOPO (Addgene no. 11584) and inserted into the naturally occurring *XhoI* site within intron 7–8 by conventional cloning methods including phosphatase treatment of the vector fragment (Figure 1C). A clone with the desired orientation was identified by Sanger sequencing. In addition to the tetrameric tandem array of SV40 polyadenyl-

ation sequences (the "stopper") and additional sequences including a strong splice acceptor site to prevent splicing around the stopper, the LSL cassette contains a puromycin selection marker (26). The insertion of the LSL cassette created targeting arms 1 (3.9 kb) and 2 (4.1 kb) from the original 8.0-kb cloned *Pole* genomic fragment. The targeting construct was sequenced in its entirety to confirm the desired structure; also, the functionality of the *LoxP* sites (and deletion of the intervening sequences) was confirmed by Sanger sequencing following transformation into the Cre-expressing *E. coli* strain 294-Cre (47).

Following *NotI* linearization and column purification of 300 μg of the targeting construct (18.0 kb), DNA was electroporated into SM-1 ES cells (derived from the 129/SvEvTac inbred mouse strain), and subsequent "plus/minus" selection (puromycin/diphtheria toxin) was conducted by standard methods (48). Screening for homologous recombination was then performed by PCR followed by Sanger sequencing to confirm the presence of the *Pole* point mutation. Two clones were selected for blastocyst injection. Chimeric mice were bred with 129/SvEvTac mice, and germline transmission and confirmation of allele function of both lines (e.g., Cre-mediated recombination at the *LoxP* sites) was performed as described in the text. One *LSL-Pole^{P286R}* line was selected and had been maintained on a pure 129/SvEvTac genetic background in the heterozygous state by breeding with 129/SvEvTac mice (Taconic Biosciences).

Mouse husbandry, survival analysis, and clinical assays. Mice were housed in a pathogen-free animal facility in microisolator cages and fed standard chow ad libitum. Survival analyses were conducted on experimental and control animals selected at the time of weaning. Complete blood counts (CBCs) were performed on 50 μl blood collected into K3-EDTA tubes (Sarstadt), with gentle mixing of the sample by gently inverting the tube multiple times immediately after collection. The mouse CBC analyses were performed on a ProCyte Dx Hematology Analyzer (IDEXX).

Genotyping and reverse transcription PCR. Tail DNA was extracted using standard protocols (49). MEF DNA was extracted using the QiaAmp DNA Mini Kit (catalog 51306, QIAGEN). Multiplex genotyping was performed with 10 μM of the primers 220 (5′-ACATTC-GCCTCTCCATTGAC-3′), 222, (5′-ATCATCTCGTCGGGTGATTT-3′) and 5F (5′-GAATTCCGCAAGCTAGCCAC-3′) and GoTaq polymerase in 1× GoTaq buffer in 20 μl of total reaction volumes. The PCR conditions were as follows: initial denaturation step of 95°C for 3 minutes, followed by 35 cycles of 95°C for 30 seconds, 65°C for 30 seconds, and 72°C for 45 seconds, followed by 72°C for 10 minutes. The PCR products were as follows: WT (220 + 222) 413 bp, P286R (220 + 222) 531 bp, and LSL (5F + 222) 512 bp.

For reverse transcription PCR (RT-PCR), RNA extraction was performed with the RNeasy Mini Kit (QIAGEN). The High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific) was used to reverse transcribe 1 μg RNA. PCR was performed on 20 ng cDNA, 0.2 mM each dNTP, and 2 mM MgCl₂, and with an intron-spanning primer pair (forward: 5′-GTGGACATGCGGGAGTATGA-3′; reverse: 5′-TTCTGGCTTAGGGGTGAACTC-3′) and an initial denaturation step of 95°C for 2 minutes, followed by 35 cycles of 95°C for 15 seconds, 60°C for 15 seconds, and 72°C for 30 seconds for 35 cycles, followed by 72°C for 3 minutes. PCR products were purified with the QIAquick PCR Purification Kit (QIAGEN) and then Sanger sequenced.

MEF generation and 6-TG selection. MEF lines were prepared using standard methods (27). Timed matings were established, with the morning of the vaginal plug being dated as E0.5. Embryos were har-

vested at E13.5, and genotyping of mutant and control lines was performed on the unattached cells 1 day after initial plating. 3T3 assays were performed as described previously (27). One MEF line per genotype was used for *HPRT* mutation rate analysis. For 6-TG selection, 1×10^6 cells were plated per 10-cm dish in DMEM and 10% FBS with 60 μM 6-TG (MilliporeSigma, A4882) (50). Cells were subjected to 6-TG selection for 4 weeks, with fresh media containing 6-TG changed every week. Cells (5×10^2) per 10-cm dish were plated to determine plating efficiency. Colonies were stained with 1% methylene blue and counted. Randomly selected colonies were isolated and expanded for *HPRT* RT-PCR. *HPRT* RT-PCR was performed on 20 ng cDNA, with 0.2 mM each dNTP and 2 mM $MgCl_2$ and an intron-spanning primer pair (forward: 5′-ATGCCGACCCGCAGTCC-3′; reverse: 5′-TTAGGCTTTGTATTTGGCTTTTCCA-3′) and an initial denaturation step of 95°C for 2 minutes, followed by 35 cycles of 95°C for 15 seconds, 60°C for 15 seconds, and 72°C for 60 seconds, followed by 72°C for 3 minutes. PCR products were purified with the QIAquick PCR Purification Kit and Sanger sequenced. Graphical representation of data was generated with MultAlin (51).

MEF protein lysates were prepared with RIPA buffer (catalog 89900, Pierce, Thermo Fisher Scientific) with a 1:100 protease inhibitor cocktail (MilliporeSigma, catalog P8340) and 0.5 mM DTT (Invitrogen, Thermo Fisher Scientific, catalog P2325). Lysates were loaded onto NuPAGE 4%–12% Bis-Tris Protein Gels (Invitrogen, Thermo Fisher Scientific, catalog NP0336BOX) and run at 150 V for 2 hours. Proteins were wet transferred onto PVDF membranes (MilliporeSigma, catalog PVH00010). Antibodies used for Western blots were from the DNA Damage Antibody Sampler Kit (Cell Signaling Technology, catalog 9947, all at 1:1,000 dilution), p16INK4a (Abcam, catalog ab189034, 1:500), and GAPDH (Cell Signaling Technology, catalog 2118, 1:10,000). A Bio-Rad ChemiDoc system was used for Western gel chemiluminescence detection for the indicated exposure durations (Figure 2C).

Tissue processing and IHC. Fixation, sectioning, antigen retrieval, blocking, and secondary detection for primary antibody dilutions in 2% BSA were performed as previously described (52). The following antibodies were used for IHC: CD3 (Abcam, catalog ab5690 rabbit polyclonal Ab [pAb], 1:2,000 dilution), CD19 (Cell Signaling Technology, catalog 90176 rabbit mAb, 1:1,600 dilution), ERα (Santa Cruz Biotechnology, catalog sc-532 rabbit pAb, 1:1,000 dilution), p63 (GeneTex, catalog GTX102425 rabbit pAb, 1:1,000 dilution), F4/80 (Abcam, catalog ab6640, rat mAb, 1:100 dilution), and TTF1 (Abcam, catalog ab76013, rabbit mAb, 1:250 dilution) (53).

DNA and library preparation. DNA from all cell and tissue samples used for high-throughput sequencing were extracted with the QiaAmp DNA Mini Kit (QIAGEN, catalog 51306). Concentration was measured with the Qubit fluorometer (Invitrogen, Thermo Fisher Scientific). Sample integrity and purification were confirmed by agarose gel electrophoresis. For the preparation of libraries, 1.5 μg DNA was fragmented with a Covaris ultrasonicator and then analyzed by gel electrophoresis to evaluate the fragmented DNA. The fragmented DNA was combined with End Repair Mix (Thermo Fisher Scientific) and incubated at 20°C for 30 minutes. The end-repaired DNA was purified with the QIAquick PCR Purification Kit (QIAGEN), followed by addition of A-Tailing Mix (New England BioLabs), and incubated at 3°C for 30 minutes. This was combined with the purified Adenylate 3′ ends DNA, adapter and ligation mix, and the ligation reaction was incubated at 20°C for 15 minutes.

Adapter-ligated DNA was selected by electrophoresis on a 2% agarose gel to recover the target fragments. The DNA was gel purified with the QIAquick Gel Extraction Kit (QIAGEN). Several rounds of PCR amplification with PCR Primer Cocktail (Illumina) and PCR Master Mix (Illumina) were performed to enrich the adapter-ligated DNA fragments. Then, the PCR products were selected by electrophoresis on a 2% agarose gel to recover the target fragments, followed by gel purification with the QIAquick Gel Extraction Kit. The final library was analyzed in 2 ways: (a) determination of the average molecule length using the Agilent 2100 Bioanalyzer (Agilent DNA 1000 Reagents), and (b) real-time PCR quantification of the library by TaqMan assay (Thermo Fisher Scientific). The qualified libraries were loaded onto the HiSeq Xten Sequencer (Illumina) for paired-end sequencing, with read lengths of 100 to 150 bp.

Variant calling. Reads were mapped to the mouse reference genome (GRCm38) using BWA 0.7.17 (54). The duplicated reads were then marked using Picard (Broad Institute), and the base quality score recalibration was applied using GATK 4.0 (55). SNP and INDEL discovery across all samples was performed using SAMtools (56). A mutation was considered to exist in a sample only if the alternative allele frequency was greater than 10%. A mutation was considered somatic when it was not a known mouse variant from the Mouse Genomes Project (57), and the mutation was identified in only 1 sample and not in any other samples. All mutations were annotated using the SnpEff tool (58) on the basis of GENCODE M16 annotation (59).

Depth analyses. Average depth in each window (1 Mbp) was estimated using SAMtools. Raw depth was normalized by dividing the median depth across the genome, followed by $\log_2$ transformation.

Data availability. Sequence data supporting the findings of this study have been deposited in the NCBI's Sequence Read Archive (SRA) database (SRA SRP150647).

Statistics. Data are presented as the mean ± SEM unless otherwise indicated. To determine *P* values, a 2-tailed Student's *t* test or Fisher's exact test was performed unless otherwise indicated. A *P* value of less than 0.05 was considered statistically significant. For survival curves, Kaplan-Meier analysis was used, with statistical comparison among curves performed using the log-rank test. Routine statistical analyses were conducted with either GraphPad Prism, version 7 (GraphPad Software) or Microsoft Excel. No statistical method was used to predetermine sample sizes. The experiments were not randomized, and the investigators were not blinded to the treatment allocation during experiments or outcome assessments.

Study approval. The animal studies were approved by the IACUC of the UTSW Medical Center.

Author contributions

IC, HDL, and DHC conceived the study. IC and DHC designed and generated the allele and mice. HDL and IC performed the experiments. MZ, YXF, CL, EAA, and MMA assisted with mouse studies. DHC and MJY conducted the immunohistochemical and histologic characterization of tumors. HZ designed and conducted the computational and data analyses. DHC, HDL, and IC wrote the manuscript, with input from all authors.

Acknowledgments

1. Lujan SA, Williams JS, Kunkel TA. DNA Polymerases Divide the Labor of Genome Replication. *Trends Cell Biol*. 2016;26(9):640–654.
2. Chilkova O, Jonsson BH, Johansson E. The quaternary structure of DNA polymerase epsilon from Saccharomyces cerevisiae. *J Biol Chem*. 2003;278(16):14082–14086.
3. Syväoja J, et al. DNA polymerases alpha, delta, and epsilon: three distinct enzymes from HeLa cells. *Proc Natl Acad Sci U S A*. 1990;87(17):6664–6668.
4. Henninger EE, Pursell ZF. DNA polymerase ε and its roles in genome stability. *IUBMB Life*. 2014;66(5):339–351.
5. Ganai RA, Johansson E. DNA replication — a matter of fidelity. *Mol Cell*. 2016;62(5):745–755.
6. Shcherbakova PV, Pavlov YI, Chilkova O, Rogozin IB, Johansson E, Kunkel TA. Unique error signature of the four-subunit yeast DNA polymerase epsilon. *J Biol Chem*. 2003;278(44):43770–43780.
7. Burgers PMJ, Kunkel TA. Eukaryotic DNA Replication Fork. *Annu Rev Biochem*. 2017;86:417–438.
8. Bauer GA, Heller HM, Burgers PM. DNA polymerase III from Saccharomyces cerevisiae. I. Purification and characterization. *J Biol Chem*. 1988;263(2):917–924.
9. Korona DA, Lecompte KG, Pursell ZF. The high fidelity and unique error signature of human DNA polymerase epsilon. *Nucleic Acids Res*. 2011;39(5):1763–1773.
10. Liu D, Keijzers G, Rasmussen LJ. DNA mismatch repair and its many roles in eukaryotic cells. *Mutat Res*. 2017;773:174–187.
11. Greaves M, Maley CC. Clonal evolution in cancer. *Nature*. 2012;481(7381):306–313.
12. Loeb LA. Human cancers express mutator phenotypes: origin, consequences and targeting. *Nat Rev Cancer*. 2011;11(6):450–457.
13. Lawrence MS, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*. 2013;499(7457):214–218.
14. Rayner E, et al. A panoply of errors: polymerase proofreading domain mutations in cancer. *Nat Rev Cancer*. 2016;16(2):71–81.
15. Campbell BB, et al. Comprehensive analysis of hypermutation in human cancer. *Cell*. 2017;171(5):1042–1056.e10.
16. Barbari SR, Kane DP, Moore EA, Shcherbakova PV. Functional analysis of cancer-associated DNA polymerase ε variants in Saccharomyces cerevisiae. *G3 (Bethesda)*. 2018;8(3):1019–1029.
17. Barbari SR, Shcherbakova PV. Replicative DNA polymerase defects in human cancers: Consequences, mechanisms, and implications for therapy. *DNA Repair (Amst)*. 2017;56:16–25.
18. Kane DP, Shcherbakova PV. A common cancer-associated DNA polymerase ε mutation causes an exceptionally strong mutator phenotype, indicating fidelity defects distinct from loss of proofreading. *Cancer Res*. 2014;74(7):1895–1901.
19. Takiar V, Ip CK, Gao M, Mills GB, Cheung LW. Neomorphic mutations create therapeutic challenges in cancer. *Oncogene*. 2017;36(12):1607–1618.
20. Muller HJ. Further studies on the nature and causes of gene mutations. *Int Congr Genet*. 1932;6(1):213–255.
21. Sánchez-Rivera FJ, Jacks T. Applications of the CRISPR-Cas9 system in cancer biology. *Nat Rev Cancer*. 2015;15(7):387–395.
22. McFadden DG, et al. Genetic and clonal dissection of murine small cell lung carcinoma progression by genome sequencing. *Cell*. 2014;156(6):1298–1311.
23. McFadden DG, et al. Mutational landscape of EGFR-, MYC-, and Kras-driven genetically engineered mouse models of lung adenocarcinoma. *Proc Natl Acad Sci U S A*. 2016;113(42):E6409–E6417.
24. Westcott PM, et al. The mutational landscapes of genetic and chemical models of Kras-driven lung cancer. *Nature*. 2015;517(7535):489–492.
25. Abbott JA. Adenomyosis and Abnormal Uterine Bleeding (AUB-A)-Pathogenesis, diagnosis, and management. *Best Pract Res Clin Obstet Gynaecol*. 2017;40:68–81.
26. Jackson EL, et al. Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. *Genes Dev*. 2001;15(24):3243–3248.
27. Sharpless NE. Preparation and immortalization of primary murine cells. In: Celis JE, ed. *Cell biology: a laboratory handbook*. 3rd ed. Burlington, Vermont, USA: Academic Press; 2006:223–228
28. Gallardo T, Shirley L, John GB, Castrillon DH. Generation of a germ cell-specific mouse transgenic Cre line, Vasa-Cre. *Genesis*. 2007;45(6):413–417.
29. Keane TM, et al. Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature*. 2011;477(7364):289–294.
30. Shinbrot E, et al. Exonuclease mutations in DNA polymerase epsilon reveal replication strand specific mutation patterns and human origins of replication. *Genome Res*. 2014;24(11):1740–1750.
31. Alexandrov LB, et al. Signatures of mutational processes in human cancer. *Nature*. 2013;500(7463):415–421.
32. Haradhvala NJ, et al. Distinct mutational signatures characterize concurrent loss of polymerase proofreading and mismatch repair. *Nat Commun*. 2018;9(1):1746.
33. Weng AP, et al. Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. *Science*. 2004;306(5694):269–271.
34. Chang MT, et al. Accelerating Discovery of Functional Mutant Alleles in Cancer. *Cancer Discov*. 2018;8(2):174–183.
35. Pachlopnik Schmid J, et al. Polymerase epsilon1 mutation in a human syndrome with facial dysmorphism, immunodeficiency, livedo, and short stature ("FILS syndrome"). *J Exp Med*. 2012;209(13):2323–2330.
36. Albertson TM, et al. DNA polymerase epsilon and delta proofreading suppress discrete mutator and cancer phenotypes in mice. *Proc Natl Acad Sci U S A*. 2009;106(40):17101–17104.
37. Warner K, et al. Models for mature T-cell lymphomas--a critical appraisal of experimental systems and their contribution to current T-cell tumorigenic concepts. *Crit Rev Oncol Hematol*. 2013;88(3):680–695.
38. Yuan T, et al. Regulation of PI3K signaling in T-cell acute lymphoblastic leukemia: a novel PTEN/Ikaros/miR-26b mechanism reveals a critical targetable role for PIK3CD. *Leukemia*. 2017;31(11):2355–2364.
39. Cosgrove CM, et al. An NRG Oncology/GOG study of molecular classification for risk prediction in endometrioid endometrial cancer. *Gynecol Oncol*. 2018;148(1):174–180.
40. Elsayed FA, et al. Germline variants in POLE are associated with early onset mismatch repair deficient colorectal cancer. *Eur J Hum Genet*. 2015;23(8):1080–1084.
41. Hodel KP, et al. Explosive mutation accumulation triggered by heterozygous human Pol ε proofreading-deficiency is driven by suppression of mismatch repair. *Elife*. 2018;7:e32692.
42. Billingsley CC, Cohn DE, Mutch DG, Stephens JA, Suarez AA, Goodfellow PJ. Polymerase ε (POLE) mutations in endometrial cancer: clinical outcomes and implications for Lynch syndrome testing. *Cancer*. 2015;121(3):386–394.
43. Le DT, et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science*. 2017;357(6349):409–413.
44. Li QX, Feuer G, Ouyang X, An X. Experimental animal modeling for immuno-oncology. *Pharmacol Ther*. 2017;173:34–46.
45. Zitvogel L, Pitt JM, Daillère R, Smyth MJ, Kroemer G. Mouse models in oncoimmunology. *Nat Rev Cancer*. 2016;16(12):759–773.
46. Bardeesy N, et al. Loss of the Lkb1 tumour suppressor provokes intestinal polyposis but resistance to transformation. *Nature*. 2002;419(6903):162–167.
47. Buchholz F, Angrand PO, Stewart AF. A simple assay to determine the functionality of Cre or FLP recombination targets in genomic manipulation constructs. *Nucleic Acids Res*. 1996;24(15):3118–3119.
48. Behringer RR, Gerstenstein M, Vintersten Nagy K, Nagy A. *Manipulating the mouse embryo: a laboratory manual*. 4th ed. Cold Spring Harbor, New York, USA: Cold Spring Harbor

Laboratory Press; 2014.

49. Truett GE, Heeger P, Mynatt RL, Truett AA, Walker JA, Warman ML. Preparation of PCR-quality mouse genomic DNA with hot sodium hydroxide and tris (HotSHOT). *BioTechniques*. 2000;29(1):52–54.

50. Glaab WE, Tindall KR. Mutation rate at the hprt locus in human cancer cell lines with specific mismatch repair-gene defects. *Carcinogenesis*. 1997;18(1):1–8.

51. Corpet F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res*. 1988;16(22):10881–10890.

52. Nakada Y, et al. The LKB1 tumor suppressor as a biomarker in mouse and human tissues. *PLoS One*. 2013;8(9):e73449.

53. Nucci MR, et al. Biomarkers in diagnostic obstetric and gynecologic pathology: a review. *Adv Anat Pathol*. 2003;10(2):55–68.

54. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–1760.

55. McKenna A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297–1303.

56. Li H, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–2079.

57. Adams DJ, Doran AG, Lilue J, Keane TM. The Mouse Genomes Project: a repository of inbred laboratory mouse strain genomes. *Mamm Genome*. 2015;26(9–10):403–412.

58. Cingolani P, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)*. 2012;6(2):80–92.

59. Wright JC, et al. Improving GENCODE reference gene annotation using a high-stringency proteogenomics workflow. *Nat Commun*. 2016;7:11778.