Deficiency in POLE exonuclease causes synthetic lethality in highly aneuploid cancer cells

Eun Jung Kim[1], Alejandro Chibly[2], Kristel Dorighi[3], Jack Kamm[2], Szymon Juszkiewicz[3], Kevin Cabrera[4], Dexter Jin[4], Rohit Reja[2], Yangmeng Wang[1], Rachel Nakagawa[1], Minyi Shi[3], Yuxin Liang[3], Anwesha Dey[3], Ethan S. Sokol[4,*], Steven Gendreau[1,*], Meng He[1,5,*]

[1]Translational Medicine, Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080, USA
[2]Computational Sciences, Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080, USA
[3]Research Biology, Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080, USA
[4]Foundation Medicine, Inc., 400 Summer Street, Boston, MA 02210, USA
[*]Correspondence: hem8@gene.com, esokol@foundationmedicine.com and gendreas@gene.com
[5]Lead contact

Summary

Aneuploidy is a hallmark of cancer and is associated with drug resistance and poor clinical outcomes across diverse cancer types. However, no therapies have been clinically established to target highly aneuploid tumors. By analyzing a real-world dataset comprising nearly half a million tumor samples subjected to comprehensive genomic profiling, we identified a striking mutual exclusivity between POLE exonuclease domain mutations and high aneuploidy burden. This observation was independently validated using data from The Cancer Genome Atlas (TCGA) and the Cancer Cell Line Encyclopedia (CCLE). Probabilistic modeling revealed that the elevated quantity and unique spectrum of mutations induced by POLE exonuclease deficiency increase the likelihood of inactivating essential genes on chromosome arms harboring losses, leading to a synthetic lethal phenotype in highly aneuploid cells. Functional experiments demonstrated that POLE exonuclease activity is essential for the viability of highly aneuploid cancer cell lines but is dispensable in diploid cells. These findings suggest that selective inhibition of POLE exonuclease may represent a promising therapeutic strategy for targeting highly aneuploid tumors.

Keywords

aneuploidy, exonuclease, POLE, synthetic lethality, essential genes, chromosome, tumor mutational burden, adenine base editor, microsatellite instability

Introduction

Aneuploidy is a cellular state characterized by the acquisition of an abnormal number of chromosomes or chromosome arms [1]. These chromosomal abnormalities are observed in over 90% of solid tumors and approximately 50% of hematological malignancies [2-4]. Aneuploidy has been recognized as both a consequence and a driver of tumorigenesis [5-7]. While aneuploidy often reduces fitness in normal cells, cancer cells tolerate and even exploit chromosomal imbalances for increased fitness [8-10]. Though the molecular mechanisms underpinning the initiation of aneuploidy is largely unknown, recent advances in single-cell genomics, karyotype-resolved sequencing, and functional genetics have illuminated the paradoxical role of aneuploidy in tumorigenesis [11]. Aneuploid genomes in cancer cells offer a substrate for clonal diversification under selective pressures, enabling tumor cells to adapt to microenvironmental stressors, evade immune detection, and resist therapeutic interventions [12,13]. Additionally, while chromosomal alterations can amplify oncogenes or eliminate tumor suppressors, the effects of chromosomal abnormalities can extend to global transcriptomic and proteomic disruption [14]. Previous work has implicated aneuploidy in the induction of proteotoxic and metabolic stress, disruption of mitotic fidelity, and activation of tumor-suppressive checkpoints, such as p53 [15,16].

Clinically, high aneuploidy burden, as well as specific recurrent aneuploidies—such as gain of chromosome 1q or 8q, or loss of 17p or 8p—are associated with poor prognosis, therapeutic resistance, and aggressive disease behavior across numerous cancer types, including breast, colorectal, ovarian, and lung cancers [10,18,19]. Mechanistically, aneuploidy promotes resistance through multiple routes, including increased intratumoral heterogeneity, reprogramming of apoptotic and stress response pathways, altered drug metabolism, and modulation of the tumor immune microenvironment [20]. For instance, tumors with high levels of aneuploidy often exhibit reduced antigen presentation and increased immunosuppressive signaling, features that correlate with diminished response to immunotherapies [21].

Therapeutic strategies targeting aneuploidy and consequential stresses have gained increased attention. Early preclinical efforts focused on amplifying the intrinsic vulnerabilities of aneuploid cells, such as proteotoxic and mitotic stress, through the use of proteasome inhibitors, heat shock protein (HSP) inhibitors, or spindle assembly checkpoint modulators [22-24]. More recently, synthetic lethality screens have identified aneuploidy-specific dependencies, including elevated reliance on autophagy, DNA repair pathways, and the unfolded protein response

[25,26]. For example, cells with whole-chromosome gains are more sensitive to centrosome clustering inhibitors or agents targeting oxidative phosphorylation and energy homeostasis [27]. Despite promising preclinical results, clinical translation remains challenging. Many stress pathway inhibitors lack tumor specificity or induce systemic toxicity [28]. Nonetheless, early-phase trials have evaluated compounds such as the HSP90 inhibitor ganetespib and proteasome inhibitor bortezomib, with mixed results [22,23,29].

In this study, we analyzed a real-world database comprising half a million patient samples across over 400 cancer types and subtypes to identify potential synthetic lethal targets for high aneuploidy burden. We identified and validated the exonuclease domain of POLE as a potential therapeutic target for tumors exhibiting high aneuploidy burden.

Results

Characterization of chromosomal arm-level aneuploidy in a real-world comprehensive genomic profiling database

We examined a real-world database of comprehensive genomic profiling results derived from 499,320 patient samples. This database encompasses over 90 different primary cancer types, including non-small cell lung cancer, colorectal cancer, breast cancer, pancreatic cancer, ovary cancer, prostate cancer, etc. (Figure 1A). Patient and tumor characteristics, including chromosomal sex, histological cancer type, local/metastasis biopsy site and genetic ancestry, are well represented in this database (Figure 1B).

A computational algorithm was run to call chromosomal arm level gains and losses for 40 autosomal chromosomal arms for each sample [30,31]. The arm-level aneuploidy caller revealed unique chromosome alterations across different histological cancer types (Figure 1C-1F). Consistent with previous reports, 8p loss, 17p loss, 1q gain and 8q gain are particularly prevalent in carcinomas, as compared to sarcomas, gliomas and leukemias [3,18,19], whereas whole chromosome 7 gain and whole chromosome 10 loss were observed as unique features in gliomas [32]. On the other hand, leukemia exhibited significantly less arm-level aneuploidy [33]. Upon examining the arm-level aneuploidy across different primary tumor types (Supplemental Figure 1), we also observed distinct distributions of specific chromosomal alterations. For instance, while 1q gain is common across various carcinoma cancer types, it is particularly common in breast cancer with a near 70% incidence rate. 3p loss is particularly enriched in

small cell, head and neck, cholangiocarcinoma and kidney cancers with an incidence rate over 50%. Another notable example is 13q gain which is detected in approximately 70% of colorectal cancer but is rare in most other cancer types.

To estimate the aneuploidy burden in each tumor type, we calculated an aneuploidy score by counting the total number of chromosome arm losses or gains in a tumor sample [3,34,35]. Not surprisingly, we observed pronounced differences in aneuploidy scores across tumor types (Figure 1G). Hematological malignancies such as leukemia, non-Hodgkin lymphoma, and diffuse large B-cell lymphoma exhibit significantly less aneuploidy. Even within solid tumor types, gliomas, mesotheliomas, and prostate cancers tend to have significantly lower aneuploidy scores compared to breast cancer, non-small cell lung cancer and small cell cancer.

Additionally, we observed statistically significant differences in aneuploidy scores among patients with different genetic ancestries (Supplemental Figure 2A). Patients of Admixed American ancestry generally display lower aneuploidy compared to patients of African ancestry. Higher aneuploidy burden has been associated with metastasis [36,37], which was consistent with our observation of significantly higher aneuploidy scores in tumor samples collected from metastatic sites (Supplemental Figure 2B). A strong correlation between aneuploidy scores and patient age was also noted (Supplemental Figure 2C). Chromosomal arm losses and gains contribute equally to aneuploidy scores, and no correlation was observed between these two types of alterations (Supplemental Figure 2D), suggesting these events are likely selected independently during cancer evolution.

Elastic net regression model identifies genetic mutation of POLE as a top negative predictor of aneuploidy score

To systematically evaluate the impact of patient and tumor characteristics, including sex, tumor subtype, genetic ancestry, metastasis status, and genomic alterations on aneuploidy, we deployed an elastic net regression model using aneuploidy score as the dependent variable (see Method for details). The comprehensive results from the elastic net regression model can be found in Supplemental Table S1, with an optimal alpha value of 0.25 (Supplemental Figure 3A). Genomic alterations with non-zero coefficients are summarized in Figure 2A. The model identified *TP53* alteration as the top positive predictor of aneuploidy score, a well-established driver of cancer genomic instability and aneuploidy [4,6,38]. In non-transformed primary cells, *TP53* inactivation is often required for cells to tolerate aneuploidy [8,39,40]. Thus, *TP53*

123 alteration serves as a positive control, demonstrating the robustness of our model. Further
124 examination showed that tumors with homozygous loss (CN) of *TP53* exhibit significantly higher
125 aneuploidy scores than those with *TP53* short variants (SV) or structural rearrangements (RE)
126 (Figure 2B and Supplemental Figure 3B). Apart from *TP53*, our model identified several other
127 genomic alterations as top positive predictors for aneuploidy scores, such as *AR, GID4, AKT3,*
128 *and FGF10*. Although the functional link between these genomic alterations and aneuploidy
129 remains unclear, it is interesting to note that alterations of *GID4, AKT3,* and *FGF10* are
130 predominantly through amplification (Supplemental Figure 3C), and the association between *AR*
131 and aneuploidy score is driven by *AR* amplification in prostate cancer (Supplemental Figure
132 3D).

133 The model also identified several negative predictors for aneuploidy scores, such as *GNAQ,*
134 *BCL2A1, CD74, FOXL2, EWSR1*, and notably *POLE*. Although their functional links with
135 aneuploidy are unclear, the strong signal that alterations in *POLE* is associated with lower
136 aneuploidy scores warranted further investigation. *POLE* encodes the catalytic subunit of the
137 DNA polymerase epsilon complex and is most frequently mutated in endometrial (1.5%) and
138 colorectal cancers (0.5%) (Figure 2C). *POLE* encodes a protein with multiple functional
139 domains, including the exonuclease domain, the DNA polymerase domain, and the less
140 characterized DUF1744 domain [41,42] (Figure 2D). Hotspot mutations in the *POLE*
141 exonuclease domain have been reported [41,43] and separating tumors with mutations within
142 this domain versus outside of the domain revealed that the negative association with high
143 aneuploidy burden is driven predominantly by mutations within the exonuclease domain (Figure
144 2E and Supplemental Figure 3E). This association was validated in The Cancer Genome Atlas
145 (TCGA) (Figure 2F) and the Cancer Cell Line Encyclopedia (CCLE) (Figure 2G). Across all
146 three datasets, tumors with exonuclease domain mutations of *POLE* are near diploid compared
147 to those with mutations outside the exonuclease domain or *POLE* wild-type tumors.

148 We also looked to see if there was an association between aneuploidy score and *POLD1*
149 alterations. *POLD1* encodes the catalytic subunit of DNA polymerase delta. Both *POLE* and
150 *POLD1* are essential B-family DNA polymerases in eukaryotic cells, playing crucial roles in DNA
151 replication and repair [44-47]. Both enzymes possess 5′ to 3′ polymerase and 3′ to 5′
152 exonuclease proofreading activities, ensuring high-fidelity DNA synthesis. However, they differ
153 in their primary functions during replication: POLE is primarily responsible for leading strand
154 synthesis, while POLD1 mainly synthesizes the lagging strand [44]. Mutations in the
155 exonuclease domains of POLE and POLD1 can compromise their proofreading abilities, leading

156 to increased mutation rates [41,42]. Interestingly, our elastic net regression model identified
157 *POLE* alteration as a top negative predictor to aneuploidy scores, but not *POLD1*. We therefore
158 examined the association between aneuploidy scores and *POLD1* mutation status. While there
159 was a statistically significant reduction in aneuploidy score in *POLD1* exonuclease (a.a.304-
160 a.a.533) mutated tumors, the effect size was minimal and less consistent across cancer types
161 compared to *POLE* (Figure 2H and Supplemental Figure 3F).

162 Deficiency in POLE exonuclease causes synthetic lethality in highly-aneuploid 163 cells by inactivating essential genes on chromosome arms with losses

164 We were interested in understanding the underlying mechanism that drives the mutual
165 exclusivity between POLE exonuclease mutations and high aneuploidy burden in tumor cells.
166 Mutual exclusivity in biology has been proposed to stem from either functional redundancy or
167 synthetic lethality [48]. POLE exonuclease mutations are often associated with ultra mutated
168 tumors, characterized by extremely high tumor mutation burdens (TMB), in many cases
169 exceeding 500,000 mutations per haploid genome (Figure 3A and Supplemental Figure 4A)
170 [45]. Although there is no clear association between TMB and aneuploidy scores when
171 examined across all tumors (Supplemental Figure 4C) [3], we observed a strong negative
172 correlation between TMB and aneuploidy scores in *POLE*-mutated tumors (Figure 3B). POLE
173 exonuclease mutations are not only associated with ultrahigh mutations but also associated with
174 a unique mutational signature [49,50], leading us to the hypothesis that the ultrahigh TMB and
175 unique mutation spectrum caused by POLE exonuclease mutations lead to synthetic lethality in
176 tumor cells with high aneuploidy burden (Figure 3C). We proposed that in highly aneuploid
177 tumor cells with numerous chromosome arm losses and gains, the inactivation of POLE's
178 exonuclease domain would lead to ultrahigh mutations across the genome. The unique
179 spectrum associated with these mutations may have a high probability of inactivating essential
180 genes located on chromosome arms with losses, leading to cancer cell death (Figure 3C).
181 Consistently, in POLD1 exonuclease mutated tumors where we did not observe strong mutual
182 exclusivity with high aneuploidy burden (Figure 2H), the TMB is substantially lower in
183 comparison to POLE exonuclease mutated tumors (Supplemental Figure 4B).

184 Consistent with our hypothesis, we observed in the real-world database that chromosomal arm
185 losses and random mutations (limited to variants with unknown significance) are less likely to
186 co-occur (odds ratio < 1) (Figure 3D). Next, we developed a probabilistic model to estimate the

187 likelihood of mutations caused by POLE exonuclease deficiency causing lethality by inactivating
188 essential genes (see Method for details).

189 $$Pr_{cell\ death\ per\ cell\ cycle} \approx 1 - exp\ (-\ n \times P_{arm} \times \alpha \times m)$$

190 In this model, $\mathbf{n}$ stands for the number of chromosome arms with losses in a tumor cell, P_{arm}
191 stands for the average probability of a mutation caused by POLE exonuclease deficiency hitting
192 an essential gene per chromosome arm, $\boldsymbol{\alpha}$ stands for the portion of deleterious mutations hitting
193 essential genes, and it remains a challenge to determine the true value of $\boldsymbol{\alpha}$. Lastly, $\mathbf{m}$ stands
194 for the number of mutations POLE exonuclease deficiency can cause in a cell per cell cycle per
195 haploid genome. Based on this model, we determined that for a typical solid tumor cell with an
196 average of 6 chromosome arm losses (based on the real-world database), if 10% ($\boldsymbol{\alpha}$) of
197 mutations caused by POLE exonuclease deficiency hitting an essential gene are deleterious,
198 the probability of inactivating at least one essential gene, leading to cell death per cell cycle,
199 approaches 100%. If we assume $\boldsymbol{\alpha}$ = 0.01, the probability of cell death per cell cycle under the
200 same condition is approximately 25% (Figure 3E). As the value of $\boldsymbol{\alpha}$ becomes smaller, the
201 probability of cell death per cell cycle is less likely to induce a synthetic lethality or mutual
202 exclusivity phenotype.

203 The elastic net regression model also identified microsatellite instability as another top negative
204 predictor for aneuploidy score (Supplemental Table 1, Supplemental Figure 4F). Mutational
205 signatures related to mismatch repair deficiency also negatively correlate with low aneuploidy
206 burden (Supplemental Figure 4D). In previous reports, a synthetic lethal interaction between
207 microsatellite instability and high aneuploidy burden was also proposed, however, the
208 underlying mechanisms are unknown [51-53]. Interestingly, even though microsatellite unstable
209 tumors exhibit low aneuploidy burden, almost comparable to POLE exonuclease mutated
210 tumors (Supplemental figure 4F), these tumors showed significantly lower TMB compared to
211 POLE exonuclease-mutated tumors (Supplemental Figures 4E and 4G), but more comparable
212 to POLD1 exonuclease-mutated tumors. However, the TMB in POLD1 exonuclease-mutated
213 tumors are likely not sufficient to drive the mutual exclusivity with high aneuploidy burden
214 (Figure 2H). Therefore, it remains to be determined if the mechanism we proposed here for
215 POLE can be applied to microsatellite unstable tumors.

216 POLE exonuclease activity is essential for highly-aneuploid cancer cells but not
217 diploid cells

218 While our probabilistic modeling provided evidence that the mutual exclusivity between POLE
219 exonuclease mutations and high aneuploidy burden in tumor cells are likely driven by synthetic
220 lethality, the unknown value of the key variable α undermines the model's validity. Next, in order
221 to experimentally validate that POLE exonuclease activity is required for survival of highly-
222 aneuploid cancer cells, but not diploid cells, we utilized an inducible adenine base editor to
223 perform a mutagenesis screen using guide RNAs targeting POLE's exonuclease domain (Figure
224 4A) [54]. We selected four cancer cell lines with high aneuploidy scores (EFE184, MDST8,
225 SW480, SW837) and four diploid cell lines (DLD1, HCT116, HEC265, LS180) (Figure 4B). We
226 first validated the inducible expression of the adenine base editor across a few cell lines,
227 including a non-cancerous cell line 293, an aneuploidy cell line SW837 and a diploid cell line
228 DLD1(Supplemental Figure 5A). Then, we performed a mutagenesis screen on the eight cancer
229 cell lines (Figure 4C, see Method for details). Seven of the eight cell lines propagated post-
230 lentiviral transduction, except for MDST8 (Figure 4D). Expression of the adenine base editor
231 was maintained during the screen (Figure 4E), and guide RNAs were amplified at passage 8,
232 sequenced, and quantified. Guide RNAs targeting POLE's exonuclease domain were
233 consistently underrepresented in the three highly aneuploid cell lines (EFE184, SW480,
234 SW837), but not in the diploid cell lines (DLD1, HCT116, HEC265, LS180) (Figure 4F,
235 Supplemental Figure S5). This supports our hypothesis that POLE exonuclease domain
236 inactivation induces synthetic lethality in highly aneuploid cells but not in diploid cells.
237 Consistent with the base editor screen, we showed that, in the highly aneuploid cell line SW480,
238 ectopic expression of wild type POLE rescued the fitness defect caused by *POLE* knockdown,
239 while the two hot-spot mutations in the POLE exonuclease domain (P286R and V411L) failed to
240 rescue the defect (Figure 4G, H). Together, our experimental data across multiple cancer cell
241 line models demonstrated the essentiality of POLE exonuclease activity in highly aneuploid
242 cells.

243 Discussion

244 Developing therapies for cancers with high aneuploidy burden is of high medical importance.
245 Synthetic lethality is an attractive strategy for identifying therapeutic targets, due to the unknown
246 molecular mechanisms causing aneuploidy and the complex biological aberrations associated
247 with it. In this study, we leveraged a real-world cancer genomics database and identified that
248 mutations in the POLE exonuclease domain are mutually exclusive with high aneuploidy
249 burden. We hypothesized and validated that POLE exonuclease inactivation causes synthetic

250 lethality in highly aneuploid cells, likely by inactivating essential genes with copy number losses.
251 This discovery leads to a therapeutic hypothesis that selective inhibition of POLE exonuclease
252 may benefit patients with highly-aneuploid tumors.

253 Previous reports indicate that endometrial or colorectal cancer patients with POLE-mutated
254 tumors often have favorable clinical outcomes, even without immune checkpoint inhibitors
255 [55,56]. The mechanisms underlying this prognosis are not fully understood. One hypothesis
256 suggests that these tumors may exhibit high tumor antigens, enhancing immune system
257 recognition [57]. We speculate that the near-diploidy nature of *POLE*-mutated tumors could also
258 contribute to these favorable clinical outcomes. Additionally, selective inhibition of POLE
259 exonuclease might potentiate immune checkpoint inhibitors by enhancing tumor antigens
260 through hypermutation.

261 However, inhibiting POLE's exonuclease domain carries risks. This domain plays a critical role
262 in genome duplication proofreading for both cancer and normal cells. Individuals with germline
263 mutations in POLE's exonuclease domain often develop Polymerase Proofreading-Associated
264 Polyposis (PPAP) and have a higher risk of colorectal and uterine cancers later in life [45,46].
265 Interestingly, these patients do not exhibit strong acute disease symptoms. Murine models show
266 that heterozygous inactivation of POLE exonuclease presents no discernible phenotypes
267 compared to wild-type animals, while homozygous inactivation correlates with higher cancer
268 incidence, similar to patients with heterozygous mutations [58]. The hotspot somatic mutation
269 P286R exhibits a pronounced tumorigenesis phenotype in animal models that is not observed
270 with germline mutations [59]. This aligns with yeast studies where analogous substitutions
271 induced strong mutator phenotypes exceeding those of conventional proofreading mutations
272 [60]. The introduction of a charged residue within the exonuclease site appears to sterically
273 prevent channeling of the nascent DNA strand containing an erroneous base towards the
274 exonuclease active site residues [61], instead likely facilitating continuous DNA synthesis
275 (Supplemental Figure 6A). Similarly, V411L substitution lies within the 'wedge helix' directly
276 involved in the separation of erroneous base-containing nascent DNA strand from the template
277 during the transfer from polymerase to exonuclease site (Supplemental Figure 6B) [62]. A
278 longer side chain of leucine compared to valine is likely to push forward the wedge helix,
279 creating a stronger physical barrier preventing efficient transfer of the leading strand from the
280 polymerase to the exonuclease site. This might explain the high prevalence and hypermutator
281 phenotype of the V411L mutation in cancers.

282 Although there is precedent for developing small molecule inhibitors of RNase H-like
283 exonucleases [63], it remains to be determined whether similar pharmacological interventions
284 could induce hypermutator activity in POLE and what the therapeutic implications of such an
285 approach might be. Despite the safety and druggability challenges associated with POLE
286 exonuclease, our work proposes a novel concept of introducing synthetic lethality in highly
287 aneuploid tumor cells. Aneuploidy might expose many essential genes as vulnerabilities due to
288 large-scale chromosome losses and gains. Strategies to inhibit these essential genes might
289 offer therapeutic opportunities.

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305 # Methods

306 ## Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
POLE	Abcam	Cat. #Ab226848
Flag	Cell Signaling Technology	Cat. #14793s; RRID: AB_2572291
CAS9	Cell Signal Technology	Cat. #14697; RRID: AB_2750916
Tubulin	Cell Signal Technology	Cat. #3873s; RRID: AB_1904178
α-Actin	Cell Signal Technology	Cat. #3700s; RRID: AB_2242334
IRDye® 800CW	LiCorBio	Cat. #926-32211
IRDye® 680LT	LiCorBio	Cat. #926-68022
Critical commercial assays		
BCA assay Kit	Thermo Fisher Scientific	23228
DNA extraction Kit	QIAGEN	69506
Q5 High fidelity 2x MM Kit	New England BioLabs	M0492s
Qubit™ 1X dsDNA High Sensitivity (HS) and Broad Range (BR) Assay Kit	Thermo Fisher Scientific	Q33230
PCR purification Kit	QIAGEN	28106
SE Cell Line 4D-Nucleofector® X Kit	Lonza	V4XC-1032
SF Cell Line 4D-Nucleofector® X Kit	Lonza	V4XC-2032
Lipofectamine 3000 Kit	Thermo Fisher Scientific	L3000-015
Lipofectamine RNAiMAX	Invitrogen	13778-075
Oligonucleotides		
Primer gRNA_ABE Forward: TCTTGTGGAAAGGACGACTCGA	Integrated DNA Technologies	N/A
Primer gRNA_ABE Reverse: CGTACTCACTATAGGGCGAATTG	Integrated DNA Technologies	N/A
On-TargetPlus Human POLE siRNA GCGAGGAACAGGCGAAAUA	Dharmacon	L-020132-00-0005
Recombinant DNA		
POLE WT	Genentech PARS	DNA1AI68314
POLE. P286R	Genentech PARS	DNA1AI68315
POLE. V411L	Genentech PARS	DNA1AI68316
ABE Lentiviral Vector	Genscript	N/A
Software and algorithms		
Graphpad Prism 8.0 software	GraphPad Software, Inc.	http://www.graphpad.com/scientific software/prism/
R	Version 4.4	https://www.r-project.org/
glmnet	version 4.1-8	https://cran.r-project.org/web/packages/glmnet/index.html

307

308 **Comprehensive genomic profiling in a real-world database**

309 Comprehensive genomic profiling of formalin-fixed, paraffin-embedded (FFPE) tissue sections
310 of 499,320 tumor samples during routine clinical care was performed using the FDA-approved
311 FoundationOne®CDx or the laboratory developed tests, FoundationOne® or
312 FoundationOne®Heme in a Clinical Laboratory Improvement Amendments–certified, College of
313 American Pathologists–accredited laboratory (Foundation Medicine Inc.). Hybrid capture was
314 carried out on at least 300 cancer-related genes and select introns from 28-34 genes frequently
315 rearranged in cancer. Arm level gains and losses were called in a research use only capacity
316 based on statistically lower or higher log2 CN for at least 50% of the arm, as previously
317 described [31]. Approval for this study, including a waiver of informed consent and Health
318 Insurance Portability and Accountability Act (HIPAA) waiver of authorization, was obtained from
319 the Western Institutional Review Board (protocol no. 20152817).

320 **Elastic net regression**

321 The R package glmnet version 4.1-8 (https://cran.r-project.org/web/packages/glmnet/index.html)
322 was used to fit an elastic net regression model using aneuploidy score (0-40) as a dependent
323 variable. The following features were selected as predictor variables: disease ontology terms
324 (n=487), metastasis status (n=2), chromosome sex (n=2), genetic ancestry (n=5), microsatellite
325 instability (n=2) and genetic alterations of evaluable genes (n=349). The outputs of the predictor
326 variables were converted to binary (0,1). The real-world dataset was split into a training set and
327 a test set with a ratio of 2:1. To identify the optimal model, a total number of 21 different alpha
328 values (0-1) were tested with an interval of 0.05. Mean Squared Error (MSE) was calculated for
329 each alpha value to determine the performance of each model. Alpha value with the smallest
330 MSE was chosen for the final model.

331 **Probabilistic modeling of process of inactivating essential genes**

332 Deficiency in POLE exonuclease leads to a unique mutational signature [49,50], with
333 approximately 30% of the mutations on trinucleotide motif TCT and 20% of the mutations on
334 TCG. This unique mutation spectrum allows us to estimate the probability of a mutation caused
335 by POLE exonuclease deficiency hitting an essential gene. The list of essential genes in
336 mammalian cells were extracted from a previous study [64]. First, we compute the probability of
337 a mutation caused by POLE exonuclease deficiency falling on the genomic coding region for
338 essential genes on each autosomal chromosome arm, through each trinucleotide motif:

$P_{TCT\text{-}1p}$ = probability of a mutation hitting an essential gene on chr1p through motif TCT

$$P_{TCT\text{-}1p} = M_{TCT\text{-}1p} \div N_{TCT}$$

N_{TCT} = total number of the trinucleotide motif TCT on human genome

$M_{TCT\text{-}1p}$ = total number of the trinucleotide motif TCT on coding regions for all essential genes located on chr1p

$P_{TCG\text{-}1p}$ = probability of a mutation hitting an essential gene on chr1p through motif TCG

$$P_{TCG\text{-}1p} = M_{TCG\text{-}1p} \div N_{TCG}$$

N_{TCG} = total number of the trinucleotide motif TCG on human genome

$M_{TCG\text{-}1p}$ = total number of the trinucleotide motif TCG on coding regions for all essential genes located on chr1p

Similar calculations can be done for all trinucleotide motifs.

P_{1p} = probability of a mutation hitting an essential gene on chr1p through any motif

$$P_{1p} = \Sigma\, w_i \times P_{i\text{-}1p},\ i\text{: all trinucleotide motifs;}$$

$$w_i\text{: probability of a mutation through motif } i \text{ [49]}$$

TCT and TCG are the dominant motifs [49], we therefore can simplify the equation above

$$P_{1p} \approx 0.6 \times P_{TCT\text{-}1p} + 0.4 \times P_{TCG\text{-}1p}$$

Similarly, the probabilities for all the 40 autosomal chromosome arms can be calculated as:

$$P_{1q} \approx 0.6 \times P_{TCT\text{-}1q} + 0.4 \times P_{TCG\text{-}1q}$$

...

$$P_{22q} \approx 0.6 \times P_{TCT\text{-}22q} + 0.4 \times P_{TCG\text{-}22q}$$

The values of P_{1p} - P_{22q} can be computed from the trinucleotide count table in supplementary

table 2. In order to evaluate the probability as a function of chromosome arm loss and other variables, we computed an average P_{arm} that will be used for all chromosomal arms in diploid cells.

$$P_{arm} = (P_{1p} + P_{1q} + P_{2p} + P_{2q} + \ldots + P_{22q}) \div 40 = 0.00057$$

Even though we can compute the probability of a mutation hitting essential genes for each chromosome arm, it remains unknown what portion of those mutations, when hitting an essential gene, will lead to loss of function. We therefore introduced a variable $\boldsymbol{\alpha}$, proportion of mutations on essential genes that are deleterious, to control the impact.

$$P_{arm} \times \alpha$$

Lastly, we need to determine the number of mutations (**m**) tumor cells can accumulate per cell cycle per haploid genome when POLE exonuclease is deficient. This number was estimated using the same approach as previously reported [65]. Briefly, the number of cell cycles (**cc**) required for a single POLE exonuclease mutated cancer cell to give rise to a tumor were calculated based on the average tumor volume for stage IV colorectal cancer ($45cm^3$) [66], and the average volume of a mammalian cell (1.8×10^{-6} mm^3) [65]. The variable **m** is then calculated as

$$m = \text{total number of mutations} \div cc \div 2$$

The average total number of mutations in POLE exonuclease mutated tumors is approximately 540,000 (TMB × 3,000) per genome based on the real-world data and TCGA, and the number of cell cycles (**cc**) is approximately 36, based on which we computed **m** = 7500. However, the TMB is estimated from a bulk of cancer cells, this estimation assumes that tumor cells are not evolutionally divergent, therefore this could be an overestimation of **m**. In a POLE exonuclease deficient tumor cell with **n** chromosome arm losses, the probability of inactivating at least one essential gene with copy number loss per cell cycle is:

$$\boldsymbol{Pr_{\text{cell death per cell cycle}} \approx 1 - (1 - n \times P_{arm} \times \alpha)^m}$$

As $\boldsymbol{m}$ is large and $\boldsymbol{n \times P_{arm} \times \alpha}$ is small, this equation can be simplified as

$$\boldsymbol{Pr_{\text{cell death per cell cycle}} \approx 1 - \exp(-n \times P_{arm} \times \alpha \times m)}$$

388 which follows from the Poisson approximation to the Binomial distribution, or equivalently the
389 fact that $(1 - p)^m \rightarrow \exp(-\lambda)$ when $m \rightarrow \infty, pm \rightarrow \lambda$. This formula could alternatively be derived by
390 assuming that mutations occur as a Poisson process with total rate **m** per cell cycle, of which $\boldsymbol{n}$
391 $\boldsymbol{\times P_{arm} \times \alpha}$ are lethal.

392 We note that the approximation above contains several simplifications that affect its accuracy.
393 First, there is a wide variance in the observed TMB across POLE-exonuclease deficient
394 samples, and thus in **m**. Secondly, different samples with the same **n** may have different
395 numbers of haploid bases depending on which chromosome arms are lost. Therefore, it would
396 improve accuracy to account for the specific TMB and aneuploidy pattern for any given sample.
397 Additionally, both **m** and **n** are estimated from bulk rather than single-cell data. Different cells
398 within a sample may have different values of **n**, causing overdispersion; meanwhile, the
399 estimated TMB includes mutations that are not present in all cells, which may inflate **m**, since it
400 is based on the number of mutations per cell divided by the number of cell cycles.

401 **Western blots**

402 Whole cell lysates were extracted using RIPA buffer (50 mM Tris HCl, 150 mM NaCl, 1.0% (v/v)
403 IGEPAL CA-630, 0.5% (w/v) Sodium Deoxycholate, 1.0 mM EDTA, 0.1% (w/v) SDS and 0.01%
404 (w/v) sodium azide at a pH of 7.4) supplemented with protease and phosphatase inhibitors
405 (Thermo Fisher Scientific, 78427,1861279). Protein concentrations were determined using the
406 BCA assay. Equal amounts of protein (20 µg per lane) were separated on 4-12% acrylamide
407 gels (Invitrogen, wg1403bx10) and transferred to membranes (Bio-Rad 1704159) using a
408 transfer system (trans-blot turbo, Bio-Rad). Membranes were blocked in 5% BSA (Gibco,15260-
409 037) in TBST (Tris-buffered saline with 0.1% Tween-20) for 30 minutes at room temperature,
410 followed by an overnight incubation at 4°C with primary antibodies diluted in blocking buffer at
411 1:500. After washing three times with TBST, membranes were incubated with secondary
412 antibodies for 1 hour at room temperature. Protein bands were visualized using Odyssey CLx
413 Infrared Imaging System.

414 **Inducible adenine base editor screen**

415 Tad-8e(V106W) was cloned downstream of a TRE3G doxycycline-inducible promoter in a
416 lentiviral expression vector with the EF1a promoter driving a puromycin resistance gene linked
417 to Tet-ON 3G. The vector also contains a human U6 promoter driving a single guide RNA

418 cassette into which the guide RNA library containing 89 POLE exonuclease tiling guide RNAs,
419 85 non-targeting control guide RNAs and 46 putative positive control guide RNAs were cloned.
420 All cloning was performed by Genscript. To perform the screen, 10 million cells were seeded in
421 15 cm plates and infected with the adenine base editor lentiviral library at a multiplicity of
422 infection (MOI) of 0.2. After 48 hours, the medium was replaced with a fresh growth medium
423 containing 3 μg/ml puromycin for 3–5 days. Following antibiotic selection, cells were trypsinized
424 and re-seeded at a density of 2 million cells per 10 cm dish in triplicates, with doxycycline added
425 to induce base editor expression. Cells were maintained for 8 passages, cell pellets containing 2
426 million cells were collected at the end of the experiment and before doxycycline treatment (as
427 baseline).

428 **DNA Isolation and gRNA amplification**

429 Genomic DNA was isolated from 2 million cells collected from the adenine base editor screen
430 using the QIAGEN DNA extraction kit, following the manufacturer's protocol. gRNAs were
431 amplified by PCR using the C1000 Touch Thermocycler (Bio-Rad). The PCR amplification was
432 performed using a standard three-step thermocycling protocol for 25 cycles: denaturation at
433 95 °C, primer annealing at 55 °C, and extension at 72 °C, employing the Q5 High-Fidelity 2X
434 Master Mix (New England BioLabs, M0492S). The resulting amplicons were purified using a
435 PCR purification kit (Qiagen, 28106) for downstream sequencing.

436 **gRNA amplicon library preparation and sequencing**

437 For generation of gRNA amplicon libraries, the KAPA HyperPrep Kit (Roche) that has
438 incorporated custom adapters and library amplification PCR primers from Integrated DNA
439 Technologies was used with an input of 40ng of amplicon DNA. Amplicon libraries were
440 quantified, and the average library size was determined using the D1000 ScreenTape on 4200
441 TapeStation (Agilent Technologies). Libraries were pooled and the Qubit dsDNA HS Assay Kit
442 (Thermo Fisher Scientific) was used to quantify the pool. Library pools were sequenced on
443 NextSeq 2000 (Illumina) to generate 3 million paired-end 51-base pair reads for each sample.

444 **Analyses of base editor screen**

445 Gene-level significance for differential barcode abundance was assessed using the FRY test
446 implemented in the limma R package
447 (https://bioconductor.org/packages/release/bioc/html/limma.html). Each cell line was evaluated

448 individually comparing the gRNA abundance targeting POLE exonuclease versus a set of non-
449 targeting control (NTC) gRNAs. For each gene, multiple gRNAs targeting that gene were
450 considered as a predefined set, and the FRY method was used to estimate a log2-Fold change
451 in barcode abundance with a corresponding p value.

452 **POLE knockdown and rescue experiments**

453 To establish cell lines stably expressing different *POLE* variants (WT, P286R, and V411L) in
454 SW480 cells, we transfected the cells with plasmids encoding human *POLE* variants using
455 electroporation. Briefly, 10^5 cells were suspended with 2 µg plasmids in the electroporation
456 buffer (SE and SF cell line 4D X kits, Lonza Bioscience). Electroporation was carried out using
457 optimized settings per the manufacturer's protocol (Lonza, 4D Nucleofector X unit). Post-
458 transfection, cells were plated in complete growth medium for recovery. Stable expression of
459 POLE was verified by western blot following puromycin selection for three weeks. For rescue
460 experiments, endogenous POLE expression was silenced using siRNA (5'-
461 GCGAGGAACAGGCGAAAUA). Cells were transfected with a siRNA mixture containing 20 nM
462 POLE-targeting siRNA and RNAiMAX transfection reagent (Invitrogen, Cat# 13778-075)
463 according to the manufacturer's protocol. Twenty-four hours post-transfection, 2,000 cells were
464 seeded into a 96-well plate and monitored by the Incucyte system (Essen BioScience) every 12
465 hours over a period of seven days. Incucyte's built-in software tools were employed to evaluate
466 cell confluence.

467 **Resource availability**

468 Lead contact

469 Further information or requests for data and resources should be directed to, and will be fulfilled
470 by the lead contact, Dr. Meng He (hem8@gene.com)

471 Materials availability

472 The all-in-one inducible adenine base editor system generated for this study can be requested
473 through the standard Genentech MTA process. Please reach out to the lead contact for details.

474 Data and code availability

475 Public datasets used in this study include TCGA and CCLE datasets, which are available for
476 download from cBioPortal (https://www.cbioportal.org/) and DepMap
477 (https://depmap.org/portal/), respectively. The sequencing data for the real-world database are

478 derived from clinical samples. All the relevant data supporting the findings of this study are
479 provided within the main manuscript and its supplementary files. Due to HIPAA requirements,
480 we are not authorized to share individualized patient genomic data, which contains potentially
481 identifying or sensitive patient information. Foundation Medicine is committed to collaborative
482 data analysis, with well-established, and widely utilized mechanisms by which investigators can
483 query our core genomic database of >700,000 de-identified sequenced cancers to obtain
484 aggregated datasets. For more information and mechanisms of access, please contact the
485 corresponding author(s) or the Foundation Medicine, Inc. Data Governance Council at
486 data.governance.council@foundationmedicine.com. All computational codes used in the
487 analysis are available for research purposes and can be provided upon request.

488 **Acknowledgments**

489 This study is funded by F. Hoffmann-La Roche AG

490 **Author contributions**

491 M.H. and S.G. contributed to the conception of the study. E.S., K.C., D.J. and S.M. contributed
492 to the curation of the real-world database. M.H. performed analysis. E.J.K. performed the wet
493 lab experiments. A.C. analyzed the base editor screen. K.D. designed the base editor system
494 and contributed to the interpretation of the data. S.J. performed the structural analysis of POLE.
495 R.R. and J. K. contributed to the statistical modeling. Y.W. and R. N. contributed to the
496 interpretation of the data and the writing and revision of the manuscript. M.S. and Y.L.
497 generated library from the gRNA library and performed next generation sequencing. A.D.
498 contributed to the interpretation of the data. E.J.K. and M.H. wrote the manuscript with input
499 from all authors. M.H., E.S., and S.G. provided supervision of this study.

500 **Declaration of interests**

501 M.H., S.G., E.J.K., A.C., K.D., S.J., R.R., J. K., Y.W., R. N., M.S., Y.L., and A.D. are Genentech Inc.
502 employees and may own stocks from F. Hoffmann-La Roche AG; E.S., K.C., and D.J. are
503 Foundation Medicine Inc. employees and may own stocks from F. Hoffmann-La Roche AG.

504

505

506 ## References

507 1. Gordon DJ, Resio B, Pellman D. Causes and consequences of aneuploidy in cancer.
508 Nat Rev Genet. 2012;13(3):189-203. doi:10.1038/nrg3123
509 2. Duijf PHG, Benezra R. The cancer biology of whole-chromosome instability. Oncogene.
510 2013;32(40):4727-4736. doi:10.1038/onc.2012.616
511 3. Taylor AM, Shih J, Ha G, et al. Genomic and functional approaches to understanding
512 cancer aneuploidy. Cancer Cell. 2018;33(4):676-689. doi:10.1016/j.ccell.2018.03.007
513 4. Dewhurst SM, McGranahan N, Burrell RA, Swanton C. Tolerance of whole-genome
514 doubling propagates chromosomal instability and accelerates cancer genome evolution.
515 Cancer Discov. 2014;4(2):175-185. doi:10.1158/2159-8290.CD-13-0285
516 5. Sansregret L, Vanhaesebroeck B, Swanton C. Determinants and clinical implications of
517 chromosomal instability in cancer. Nat Rev Clin Oncol. 2018;15(3):139-150.
518 doi:10.1038/nrclinonc.2017.198
519 6. Davoli T, Xu AW, Mengwasser KE, et al. Cumulative haploinsufficiency and
520 triplosensitivity drive aneuploidy patterns and shape the cancer genome. Cell.
521 2013;155(4):948-962. doi:10.1016/j.cell.2013.10.011
522 7. Vasudevan A, Schukken KM, Sausville EL, et al. Aneuploidy as a promoter and
523 suppressor of malignant growth. Nat Rev Cancer. 2021;21(2):89-103.
524 doi:10.1038/s41568-020-00321-1
525 8. Santaguida S, Amon A. Short- and long-term effects of chromosome mis-segregation
526 and aneuploidy. Nat Rev Mol Cell Biol. 2015;16(8):473-485. doi:10.1038/nrm4025
527 9. Zhu J, Tsai HJ, Gordon MR, Li R. Cellular stress associated with aneuploidy. Dev Cell.
528 2018;44(4):420-431. doi:10.1016/j.devcel.2018.02.002
529 10. Ben-David U, Amon A. Context is everything: aneuploidy in cancer. Nat Rev Genet.
530 2020;21(1):44-62. doi:10.1038/s41576-019-0171-x
531 11. Leibowitz ML, Zhang CZ, Pellman D. Chromothripsis: a new mechanism for rapid
532 karyotype evolution. Annu Rev Genet. 2015;49:183-211. doi:10.1146/annurev-genet-
533 120213-092228
534 12. McGranahan N, Swanton C. Clonal heterogeneity and tumor evolution: past, present,
535 and the future. Cell. 2017;168(4):613-628. doi:10.1016/j.cell.2017.01.018
536 13. Bakhoum SF, Cantley LC. The multifaceted role of chromosomal instability in cancer and
537 its microenvironment. Cell. 2018;174(6):1347-1360. doi:10.1016/j.cell.2018.08.027

538 14. Stingele S, Stoehr G, Peplowska K, et al. Global analysis of genome, transcriptome and
539 proteome reveals the response to aneuploidy in human cells. Mol Syst Biol. 2012;8:608.
540 doi:10.1038/msb.2012.40
541 15. Sheltzer JM, Amon A. The aneuploidy paradox: costs and benefits of an incorrect
542 karyotype. Trends Genet. 2011;27(11):446-453. doi:10.1016/j.tig.2011.07.003
543 16. Santaguida S, Vasile E, White E, Amon A. Aneuploidy-induced cellular stresses limit
544 autophagic degradation. Genes Dev. 2015;29(19):2010-2021.
545 doi:10.1101/gad.269118.115
546 17. Carter SL, Eklund AC, Kohane IS, Harris LN, Szallasi Z. A signature of chromosomal
547 instability inferred from gene expression profiles predicts clinical outcome in multiple
548 human cancers. Nat Genet. 2006;38(9):1043-1048. doi:10.1038/ng1861
549 18. Beroukhim R, Mermel CH, Porter D, et al. The landscape of somatic copy-number
550 alteration across human cancers. Nature. 2010;463(7283):899-905.
551 doi:10.1038/nature08822
552 19. Zack TI, Schumacher SE, Carter SL, et al. Pan-cancer patterns of somatic copy number
553 alteration. Nat Genet. 2013;45(10):1134-1140. doi:10.1038/ng.2760
554 20. Davoli T, Uno H, Wooten EC, Elledge SJ. Tumor aneuploidy correlates with markers of
555 immune evasion and reduced response to immunotherapy. Science.
556 2017;355(6322):eaaf8399. doi:10.1126/science.aaf8399
557 21. Ly P, Brunner SF, Shoshani O, et al. Chromosome segregation errors generate a
558 diverse spectrum of simple and complex genomic rearrangements. Nat Genet.
559 2019;51(4):705-715. doi:10.1038/s41588-019-0360-8
560 22. Richardson PG, Barlogie B, Berenson J, et al. A phase 2 study of bortezomib in
561 relapsed, refractory myeloma. N Engl J Med. 2003;348(26):2609-2617.
562 doi:10.1056/NEJMoa030288
563 23. Jhaveri K, Taldone T, Modi S, Chiosis G. Advances in the clinical development of heat
564 shock protein 90 (Hsp90) inhibitors in cancers. Biochim Biophys Acta.
565 2012;1823(3):742-755. doi:10.1016/j.bbamcr.2011.10.008
566 24. Tang YC, Williams BR, Siegel JJ, Amon A. Identification of aneuploidy-selective
567 antiproliferation compounds. Cell. 2011;144(4):499-512. doi:10.1016/j.cell.2011.01.017
568 25. Oromendia AB, Dodgson SE, Amon A. Aneuploidy causes proteotoxic stress in yeast.
569 Genes Dev. 2012;26(24):2696-2708. doi:10.1101/gad.207407.112

570 26. Dürrbaum M, Kuznetsova AY, Passerini V, et al. Unique features of the transcriptional
571 response to model aneuploidy in human cells. BMC Genomics. 2014;15:139.
572 doi:10.1186/1471-2164-15-139
573 27. Cohen-Sharir Y, McFarland JM, Abdusamad M, et al. Aneuploidy renders cancer cells
574 vulnerable to mitotic checkpoint inhibition. Nature. 2021;590(7846):486-491.
575 doi:10.1038/s41586-020-03114-6
576 28. de Wispelaere M, Du G, Donovan KA, et al. Small molecule degraders of the ATR
577 kinase checkpoint. Nat Commun. 2022;13(1):7894. doi:10.1038/s41467-022-35587-6
578 29. Jhaveri K, Miller K, Rosen L, et al. A phase I dose-escalation trial of trastuzumab and
579 alvespimycin hydrochloride (KOS-1022; 17 DMAG) in the treatment of advanced solid
580 tumors. Clin Cancer Res. 2012;18(18):5090-5098. doi:10.1158/1078-0432.CCR-12-0462
581 30. Pavlick D, Montesion M, Severson E, et al. A next-generation sequencing-based
582 karyotyping algorithm reveals the genomic structure of acute myeloid leukemia. Blood.
583 2018;132(Suppl 1):2773. doi:10.1182/blood-2018-99-116195.
584 31. Sharaf R, Pavlick DC, Frampton GM, et al. FoundationOne CDx testing accurately
585 determines whole arm 1p19q codeletion status in gliomas. Neurooncol Adv.
586 2021;3(1):vdab017. doi:10.1093/noajnl/vdab017
587 32. Brennan CW, Verhaak RGW, McKenna A, et al. The somatic genomic landscape of
588 glioblastoma. Cell. 2013;155(2):462-477. doi:10.1016/j.cell.2013.09.034
589 33. Papaemmanuil E, Rapado I, Li Y, et al. Genomic classification and prognosis in acute
590 myeloid leukemia. N Engl J Med. 2016;374(23):2209-2221.
591 doi:10.1056/NEJMoa1516192
592 34. Ben-David U, Miron K, Powell SN. Targeting vulnerabilities of aneuploid cells for cancer
593 therapy. Trends in Cancer. 2025;11(5):TBC. doi:10.1016/j.trecan.2025.05.001
594 35. Zhang Y, Akamatsu K, Henikoff S, et al. Total whole-arm chromosome losses predict
595 malignancy in human cancer. Proc Natl Acad Sci U S A. 2025;122(18):e2505385122.
596 doi:10.1073/pnas.2505385122
597 36. Pedrosa JA, Awal D, Gümüş ZH, et al. Intratumoral heterogeneity in clear cell renal cell
598 carcinoma and the impact of chromosomal alterations on metastasis. NPJ Precis Oncol.
599 2024;8(1):12. doi:10.1038/s41698-024-00499-w
600 37. Hieronymus H, Schultz N, Gopalan A, Carver BS, Chang MT, Xiao Y, et al. Copy
601 number alteration burden predicts therapy response in metastatic prostate cancer. Mol
602 Ther Oncolytics. 2020;16:234-245. doi:10.1016/j.omto.2020.02.002

603 38. Zhang CZ, Spektor A, Cornils H, et al. Chromothripsis from DNA damage in micronuclei.
604 Nature. 2015;522(7555):179-184. doi:10.1038/nature14493
605 39. Thompson SL, Compton DA. Proliferation of aneuploid human cells is limited by a p53-
606 dependent mechanism. J Cell Biol. 2010;188(3):369-381. doi:10.1083/jcb.200905057
607 40. Fukasawa K, Choi T, Kuriyama R, Rulong S, Vande Woude GF. Abnormal centrosome
608 amplification in the absence of p53. Science. 1996;271(5256):1744-1747.
609 doi:10.1126/science.271.5256.1744
610 41. Rayner E, van Gool IC, Palles C, et al. A panoply of errors: Polymerase proofreading
611 domain mutations in cancer. Nat Rev Cancer. 2016;16(2):71-81.
612 doi:10.1038/nrc.2015.12
613 42. Smerdon SJ, Petrovich RM, Goryshin IY, et al. Crystal structure of DNA polymerase ε
614 subunit POLE: Insights into proofreading and the DUF1744 domain. Structure.
615 2017;25(3):395-404. doi:10.1016/j.str.2016.12.002
616 43. Barbari SR, Shcherbakova PV. The role of DNA polymerase ε in eukaryotic genome
617 stability. Mutat Res Rev Mutat Res. 2017;771:67-72. doi:10.1016/j.mrrev.2016.09.002
618 44. Tahirov TH. Structure and function of eukaryotic DNA polymerases. Protein Cell.
619 2016;7(9):607-620. doi:10.1007/s13238-016-0305-0
620 45. Mosalem O, Coston TW, Imperial R, et al. A comprehensive analysis of POLE/POLD1
621 genomic alterations in colorectal cancer. Oncologist. 2024;29(9):e1224-e1227.
622 doi:10.1093/oncolo/oyae098
623 46. Gola M, Stefaniak P, Godlewski J, Jereczek-Fossa BA, Starzyńska A. Prospects of
624 POLD1 in Human Cancers: A Review. Cancers (Basel). 2023;15(6):1905. Published
625 2023 Mar 22. doi:10.3390/cancers15061905
626 47. Palles C, Cazier JB, Howarth KM, et al. Germline mutations affecting the proofreading
627 domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. Nat
628 Genet. 2013;45(2):136-144. doi:10.1038/ng.2503
629 48. Deng Y, Luo S, Deng C, et al. Identifying mutual exclusivity across cancer genomes:
630 computational approaches to discover genetic interaction and reveal tumor vulnerability.
631 Brief Bioinform. 2019;20(1):254-266. doi:10.1093/bib/bbx109
632 49. Alexandrov LB, Kim J, Haradhvala NJ, et al. The repertoire of mutational signatures in
633 human cancer. Nature. 2020;578(7793):94-101. doi:10.1038/s41586-020-1943-3
634 50. Campbell BB, Light N, Fabrizio D, et al. Comprehensive Analysis of Hypermutation in
635 Human Cancer. Cell. 2017;171(5):1042-1056.e10. doi:10.1016/j.cell.2017.09.048

636 51. Smith AB, Jones CD, Patel EF, et al. Mutual exclusivity of aneuploidy and microsatellite
637 instability in gastrointestinal cancers. Nat Commun. 2023;14(1):1234-1245.
638 52. Rodriguez MI, Martinez D, Garcia A. Therapeutic strategies for aneuploid and MSI
639 tumors: from synthetic lethality to combination therapies. Front Oncol. 2023;13:112233.
640 53. Lee GH, Kim YS, Park JH. Divergent evolutionary paths in early-onset colorectal cancer:
641 diploid MSI-H vs aneuploid MSS tumors. J Clin Oncol. 2022;40(15):2001-2010.
642 54. Richter MF, Zhao KT, Eton E, et al. Phage-assisted evolution of an adenine base editor
643 with improved Cas domain compatibility and activity. Nat Biotechnol. 2020;38(7):883-
644 891. doi:10.1038/s41587-020-0453-z
645 55. Yao X, Feng M, Wang W. The Clinical and Pathological Characteristics of POLE-
646 Mutated Endometrial Cancer: A Comprehensive Review. Cancer Manag Res.
647 2024;16:117-125. doi:10.2147/CMAR.S445055
648 56. Zhang Y, Li Y, Zhang Y, et al. Clinicopathological and molecular characterization of high-
649 grade endometrial carcinoma with POLE mutation: a single center study. J Transl Med.
650 2022;20(1):1-12. doi:10.1186/s12967-022-03388-7
651 57. Ma X, Riaz N, Samstein RM, et al. Functional landscapes of POLE and POLD1
652 mutations in checkpoint blockade-dependent antitumor immunity. Nat Genet.
653 2022;54(7):996-1012. doi:10.1038/s41588-022-01108-w
654 58. Albertson TM, Ogawa M, Bugni JM, et al. DNA polymerase epsilon and delta
655 proofreading suppress discrete mutator and cancer phenotypes in mice. Proc Natl Acad
656 Sci U S A. 2009;106(40):17101-17104. doi:10.1073/pnas.0907147106
657 59. Li HD, Lu C, Zhang H, et al. A PoleP286R mouse model of endometrial cancer
658 recapitulates high mutational burden and immunotherapy response. JCI Insight.
659 2020;5(14):e138829. Published 2020 Jul 23. doi:10.1172/jci.insight.138829
660 60. Kane, D. P., & Shcherbakova, P. V. (2014). A common cancer-associated DNA
661 polymerase ε mutation causes an exceptionally strong mutator phenotype, indicating
662 fidelity defects distinct from loss of proofreading. Cancer research, 74(7), 1895–1901.
663 doi.org/10.1158/0008-5472.CAN-13-2892
664 61. Parkash V, Kulkarni Y, ter Beek J, Shcherbakova PV, Kamerlin SCL, Johansson E.
665 Structural consequence of the most frequently recurring cancer-associated substitution
666 in DNA polymerase ε. Nat Commun. 2019;10(1):373. doi:10.1038/s41467-018-08114-9
667 62. Roske JJ, Yeeles JTP. Structural basis for processive daughter-strand synthesis and
668 proofreading by the human leading-strand DNA polymerase Pol ε. Nat Struct Mol Biol.
669 2024;31(12):1921–1931. doi:10.1038/s41594-024-01370-y

63. Huang KW, Hsu KC, Chu LY, Yang JM, Yuan HS, Hsiao YY. Identification of inhibitors for the DEDDh family of exonucleases and a unique inhibition mechanism by crystal structure analysis of CRN-4 bound with 2-morpholin-4-ylethanesulfonate (MES). J Med Chem. 2016;59(17):8019–8029. doi:10.1021/acs.jmedchem.6b00794
64. Hart T, Chandrashekhar M, Aregger M, et al. High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. Cell. 2015;163(6):1515-1526. doi:10.1016/j.cell.2015.11.015
65. Shlien A, Campbell BB, de Borja R, et al. Combined hereditary and somatic mutations of replication error repair genes result in rapid onset of ultra-hypermutated cancers. Nat Genet. 2015;47(3):257-262. doi:10.1038/ng.3202
66. Park JY, Kim SH, Lee SM, Lee JS, Han JK. CT volumetric measurement of colorectal cancer helps predict tumor staging and prognosis. PLoS One. 2017;12(6):e0178522. Published 2017 Jun 1. doi:10.1371/journal.pone.0178522

710 ## Figure legends

711 **Figure 1: Characterization of chromosomal arm level aneuploidy in the real-world**
712 **comprehensive genomic profiling database**

713 (A) Pie chart summarizing the real-world database by primary cancer type.

714 (B) Donut charts summarizing the real-world database by chromosome sex, histological types,
715 metastasis status and genetic ancestry.

716 (C-F) Frequencies of each chromosomal arm loss or gain in carcinoma (C), sarcoma (D), glioma
717 (E) and leukemia (F).

718 (G) Distribution of aneuploidy scores across major cancer types (In each boxplot, the center line
719 represents the median value. The lower and upper hinges correspond to the first and third
720 quartiles, respectively, while the whiskers extend up to 1.5x the interquartile range. This applies
721 to all the boxplots presented in this study).

722 **Figure 2: Mutations in POLE exonuclease domain are mutually exclusive with high**
723 **aneuploidy burden**

724 (A) Elastic net regression model identified genes with non-zero coefficient to be predictors for
725 aneuploidy scores. Top positive predictor genes are highlighted in red, while top negative
726 predictor genes are highlighted in blue.

727 (B) Distribution of aneuploidy scores across tumors with different *TP53* alterations. RE,
728 structural rearrangement; SV, short variant; CN, copy number alteration (homozygous deletion).
729 Student t-tests were used for two group comparisons.

730 (C) Frequency of *POLE* mutation across various cancer types in the real-world database.

731 (D) Representation of POLE protein domains.

732 (E) Distribution of aneuploidy scores in tumors with different *POLE* alterations across various
733 cancer types in the real-world database. mut.exo, mutations within the POLE exonuclease
734 domain (a.a.268-a.a.471); mut.other, mutations outside the POLE exonuclease domain; wt, wild
735 type. Student t-tests were used for two group comparisons (mut.exo vs. wt or mut.other vs. wt),
736 * $p<0.05$, ** $p<0.01$, *** $p<0.001$, ns $p>0.05$.
737

(F) Distribution of aneuploidy scores in tumors with different *POLE* alterations in TCGA.

(G) Distribution of aneuploidy scores in cell lines with different *POLE* alterations in CCLE.

(H) Distribution of aneuploidy scores in tumors with different *PODL1* alterations in the real-world database.

Figure 3: High mutation burden and the unique mutation spectrum caused by POLE exonuclease deficiency have a high probability of inactivating essential genes in highly aneuploid cells.

(A) Distribution of tumor mutational burden (TMB) in tumors with different *POLE* alterations in the real-world database.

(B) Aneuploidy scores and the mean TMB are negatively correlated in *POLE*-mutated tumors (blue line: a linear regression model fit by lm() function in R)

(C) Illustration of proposed mechanism of mutual exclusivity between high aneuploidy burden and POLE exonuclease mutations: mutations caused by POLE exonuclease deficiency have a high probability of inactivating essential genes on chromosome arm harboring losses in highly aneuploid cells.

(D) Most of the chromosomal arms are less likely to have co-occurrence of arm losses and genetic alterations (limited to variants with unknown significance), odds ratio < 1. Each dot represents an autosomal chromosomal arm.

(E) Probabilistic modeling of cell death per cell cycle in cells with POLE exonuclease deficiency. $Pr_{cell\ death\ per\ cell\ cycle} \approx 1 - exp\ (-\ n \times P_{arm} \times \alpha \times m)$, n = number of chromosome arm losses (mean = 6), P_{arm} = probability of a mutation (caused by POLE exonuclease deficiency) hitting an essential gene divided by total number of chromosome arms (P_{arm}=*0.00057*), α = percentage of POLE exonuclease deficiency-induced mutations on essential genes that are deleterious (unknown), m = number of mutations caused by POLE exonuclease deficiency per cell cycle per haploid genome (m=*7500* estimated for stage IV colorectal cancer), refer to methods section for more details.

Figure 4: Mutagenesis screening validated that POLE exonuclease is essential for the survival of highly aneuploid cells but not diploid cells.

766 (A) Illustration of the inducible adenine base editor screen tiling the exonuclease domain of
767 POLE.

768 (B) Cell lines used for the adenine base editor screen.

769 (C) Experimental flow of the adenine base editor screen.

770 (D) Cell growth curves post viral transduction and induction of adenine base editor.

771 (E) Western blot confirming the sustained expression of the adenine base editor system during
772 the experiment.

773 (F) Compared to non-targeting control gRNAs (NTC), gRNAs tiling the exonuclease domain of
774 POLE are significantly underrepresented in the aneuploid cell lines (EFE184, SW480, SW837),
775 but not in the diploidy cell lines (DLD1, HCT116, HEC265, LS180). P values and log2 fold
776 change were determined by comparing guide RNA counts at passage 8 to passage 0.

777 (G) Western blot confirming the siRNA mediated knockdown of endogenous POLE, and over
778 expression of ectopic POLE variants (WT, P286R and V411L) in SW480 cell line.

779 (H) Growth curves of SW480 cells under different genetic backgrounds.

780 Supplemental information

781 **Figure S1: Frequencies of each chromosomal arm loss or gain in major cancer types in**
782 **the real-world database.**

783 **Figure S2: Associations of aneuploidy scores with additional clinical features in the real-**
784 **world database.**
785 (A) Distribution of aneuploidy scores across tumors from patients of different genetic ancestries.
786 AMR, Ad Mixed American; SAS, South Asian; EAS, East Asian; EUR, European; AFR, African.
787 (B) Distribution of aneuploidy scores across tumors with different metastasis status. Student t-
788 tests were used for two group comparisons.
789 (C) Association between aneuploidy scores and patient age.
790 (D) Dot plots representing cooccurrence of chromosomal arm losses and gains.

791 **Figure S3: Association of aneuploidy scores with *TP53, AR*, *POLE* and *POLD1* genetic**
792 **alterations.**
793 (A) Evaluation of the performance of elastic net regression performance with different alpha
794 value. mse, mean squared error.

795 (B) Distribution of aneuploidy scores across tumors with different *TP53* status, separated by
796 major cancer types in the real-world database. mut, *TP53* mutated, including short variant,
797 structural rearrangement and homozygous deletion; wt, *TP53* wild type. Student t-tests were
798 used for two group comparisons (mut vs. wt), ** $p<0.01$, *** $p<0.001$.
799 (C) Summary of sample size with each genetic alteration in the real-world database. WT, wild
800 type; CN, copy number alteration (amplification); SV, short variant; RE, structural
801 rearrangement.
802 (D) Distribution of aneuploidy scores in prostate cancer with different *AR* status in the real-world
803 database. Student t-tests were used for two group comparisons.
804 (E) Distribution of aneuploidy scores across tumors with different *POLE* status in the real-world
805 database.
806 (F) Distribution of aneuploidy scores across tumors with different *POLD1* status, separated by
807 major cancer types in the real-world database. mut.exo, mutations within the POLD1
808 exonuclease domain (a.a.304-a.a.533); mut.other, mutations outside the POLD1 exonuclease
809 domain; wt, wild type. Student t-tests were used for two group comparisons (mut.exo vs. wt or
810 mut.other vs. wt), * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

811 **Figure S4: Associations of aneuploidy scores with TMB, mutational signatures and**
812 **microsatellite instability.**
813 (A) Distribution of tumor mutational burden (TMB) in tumors with different *POLE* alterations in
814 TCGA.
815 (B) Distribution of TMB in tumors with different *POLD1* alterations in the real-world database.
816 (C) Correlation of aneuploidy scores and the mean TMB in all tumors in the real-world database
817 (blue line: a linear regression model fit by lm() function in R)
818 (D) Distribution of aneuploidy scores across tumors with different mutational signatures in the
819 real-world database.
820 (E) Distribution of TMB across tumors with different mutational signatures in the real-world
821 database.
822 (F) Distribution of aneuploidy scores across tumors with different microsatellite instability status
823 in the real world database. Dotted line, median aneuploidy score for POLE exonuclease
824 mutated tumors.

825 (G) Distribution of TMB across tumors with different microsatellite instability status in the real
826 world database. Dotted line, median TMB for POLE exonuclease mutated tumors.

827 **Figure S5: The all-in-one inducible adenine base editor.**

828 (A) Western blots validating the inducible expression of Cas9n(D10A)-TadA-8e(V106W) in 293,
829 LS180 and SW837 cells.
830 (B) Log2 fold change of individual gRNAs tiling the POLE exonuclease domain in different cell
831 lines.

832 **Figure S6: Structure of POLE highlighting the exonuclease domain and hotspot**
833 **mutations P286 and V411**
834 (A) POLE captured on mismatched DNA, mismatch excision state (PDB 9F6L). Exonuclease
835 domain is colored in green, catalytic residues D275 and E277 are marked in blue and
836 represented as sticks. Template DNA strand is in plum and nascent DNA strand in orange.
837 (B) POLE captured on mismatched DNA, frayed substrate state (PDB 9F6K). Coloring scheme
838 as in (A), additionally the wedge helix is colored in salmon and V411 within the wedge helix in
839 red. Note how the wedge helix physically separates template and nascent DNA strands.

840 **Table S1: Excel file summarizing full results from the elastic net regression analysis**

841 **Table S2: Excel file summarizing tri-nucleotide motif counts on human genome and**
842 **essential genes**

843 **Table S3: Excel file summarizing gRNA counts from the adenine base editor screen.**

Figure 1

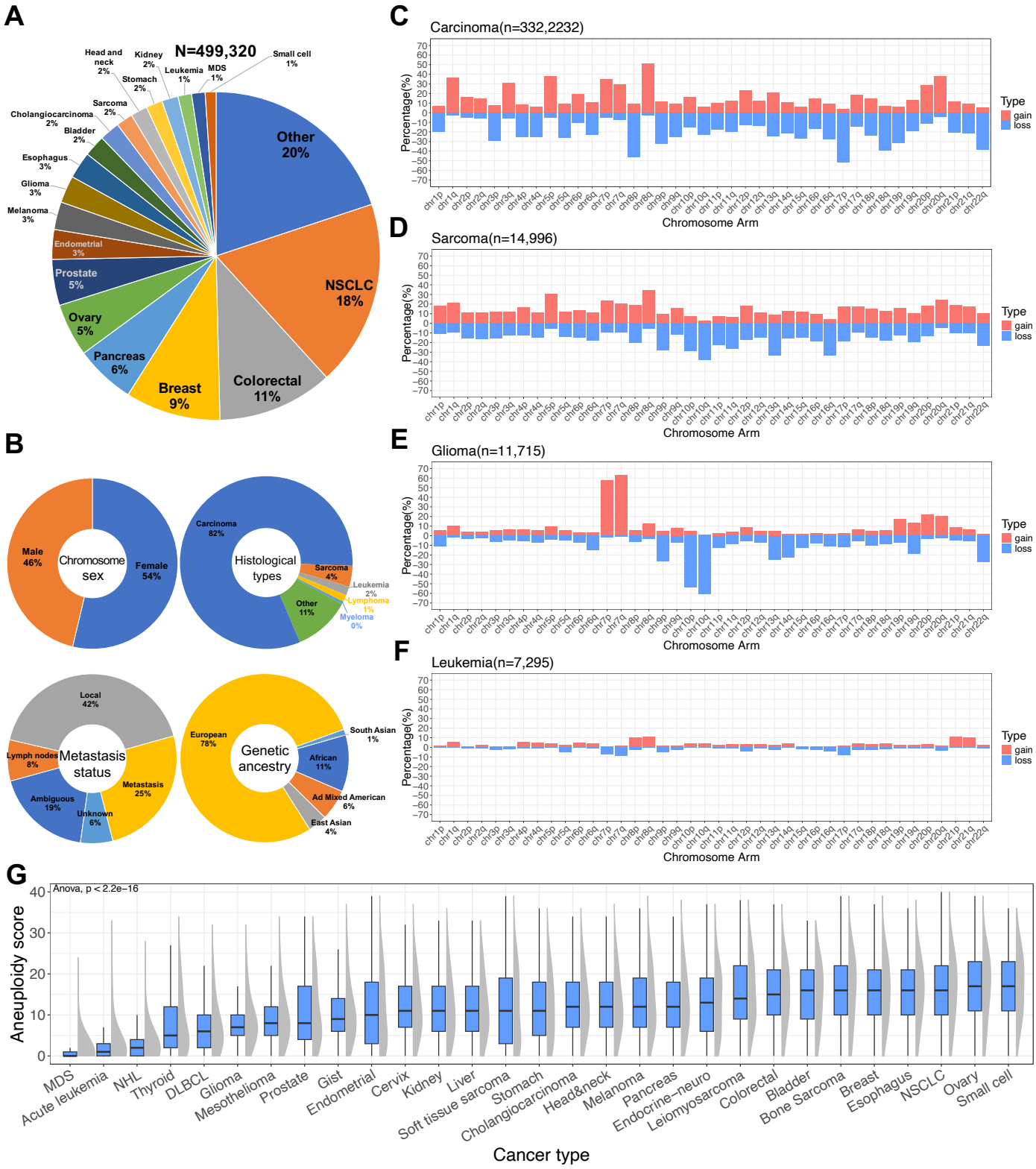

A
N=499,320
Other 20%
NSCLC 18%
Colorectal 11%
Breast 9%
Pancreas 6%
Ovary 5%
Prostate 5%
Endometrial 3%
Melanoma 3%
Glioma 3%
Esophagus 3%
Bladder 2%
Cholangiocarcinoma 2%
Sarcoma 2%
Stomach 2%
Head and neck 2%
Kidney 2%
Leukemia 1%
MDS 1%
Small cell 1%
B
Chromosome sex
Male 46%
Female 54%
Histological types
Carcinoma 82%
Sarcoma 4%
Leukemia 2%
Lymphoma 1%
Myeloma 0%
Other 11%
Metastasis status
Local 42%
Metastasis 25%
Unknown 6%
Ambiguous 19%
Lymph nodes 8%
Genetic ancestry
European 78%
South Asian 1%
African 11%
Ad Mixed American 6%
East Asian 4%
C
Carcinoma(n=332,2232)
D
Sarcoma(n=14,996)
E
Glioma(n=11,715)
F
Leukemia(n=7,295)
Percentage(%)
Chromosome Arm
Type
gain
loss
G
Anova, p < 2.2e−16
Aneuploidy score
MDS
Acute leukemia
NHL
Thyroid
DLBCL
Glioma
Mesothelioma
Prostate
Gist
Endometrial
Cervix
Kidney
Liver
Soft tissue sarcoma
Stomach
Cholangiocarcinoma
Head&neck
Melanoma
Pancreas
Endocrine−neuro
Leiomyosarcoma
Colorectal
Bladder
Bone Sarcoma
Breast
Esophagus
NSCLC
Ovary
Small cell
Cancer type

Figure 2

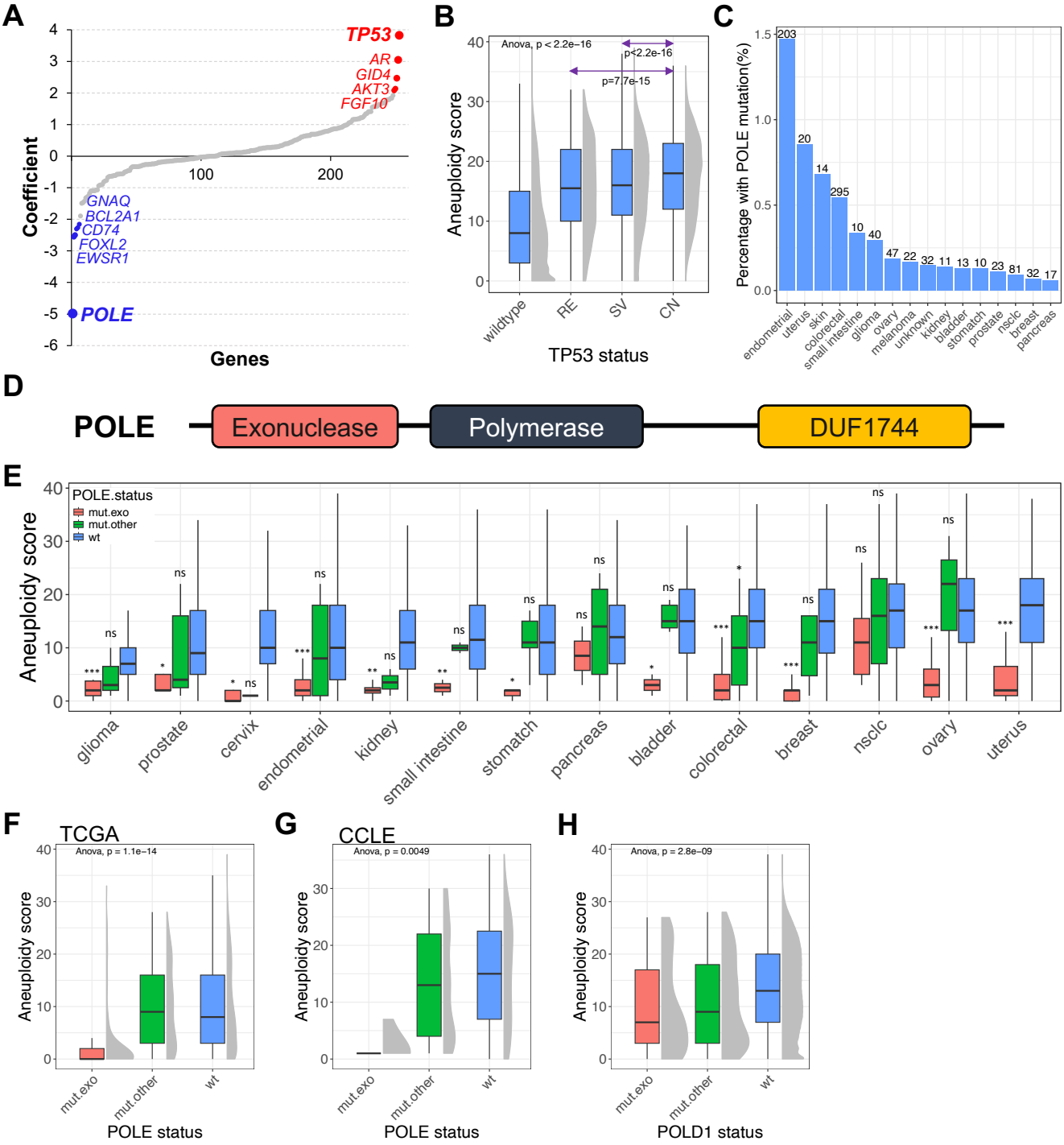

A
TP53
AR
GID4
AKT3
FGF10
GNAQ
BCL2A1
CD74
FOXL2
EWSR1
POLE
Coefficient
Genes
B
Anova, p < 2.2e−16
p<2.2e-16
p=7.7e-15
Aneuploidy score
wildtype
RE
SV
CN
TP53 status
C
Percentage with POLE mutation(%)
203
20
14
295
10
40
47
22
32
11
13
10
23
81
32
17
endometrial
uterus
skin
colorectal
small intestine
glioma
ovary
melanoma
unknown
kidney
bladder
stomach
prostate
nsclc
breast
pancreas
D
POLE
Exonuclease
Polymerase
DUF1744
E
POLE.status
mut.exo
mut.other
wt
Aneuploidy score
glioma
prostate
cervix
endometrial
kidney
small intestine
stomach
pancreas
bladder
colorectal
breast
nsclc
ovary
uterus
F
TCGA
Anova, p = 1.1e−14
Aneuploidy score
mut.exo
mut.other
wt
POLE status
G
CCLE
Anova, p = 0.0049
Aneuploidy score
mut.exo
mut.other
wt
POLE status
H
Anova, p = 2.8e−09
Aneuploidy score
mut.exo
mut.other
wt
POLD1 status

Figure 3

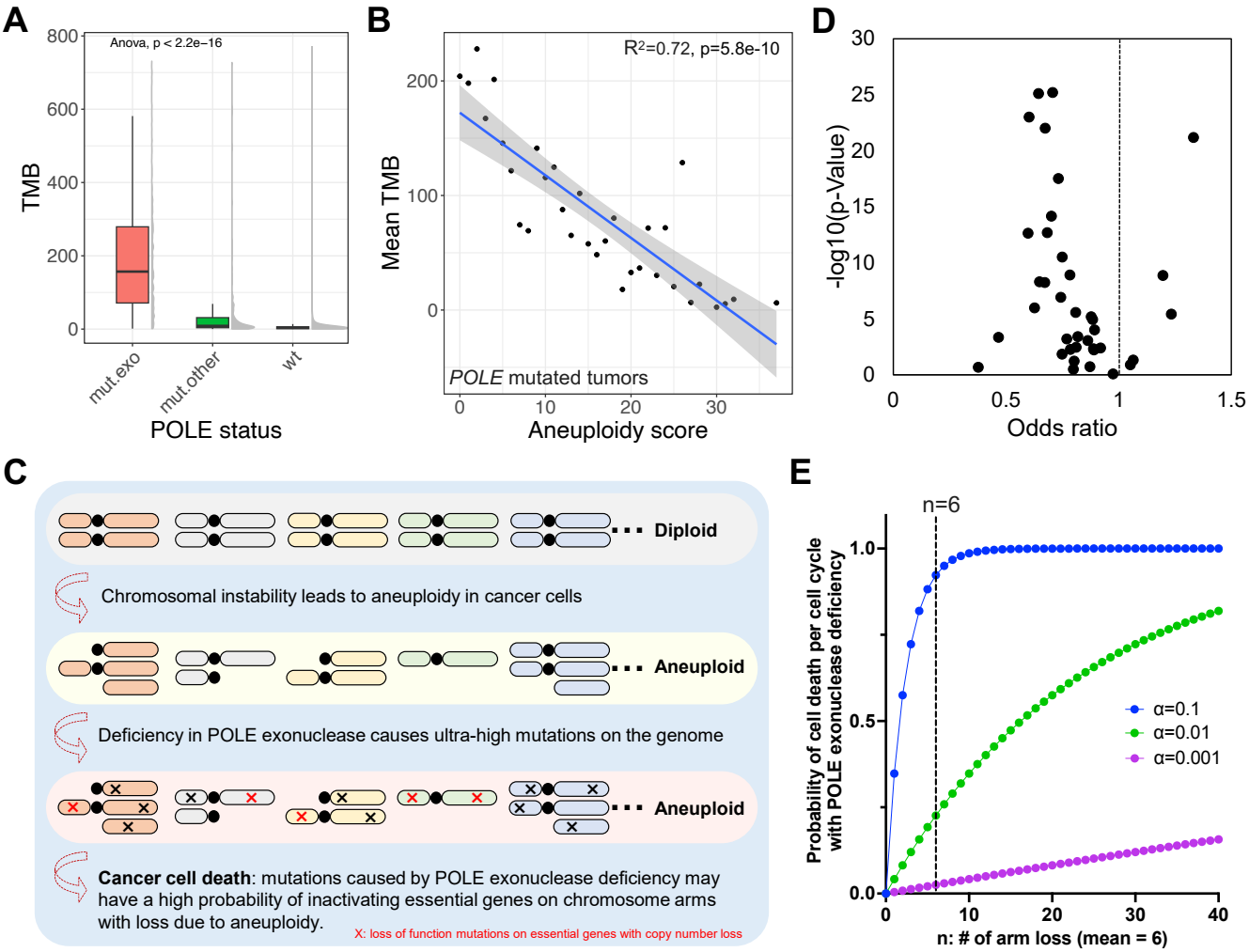
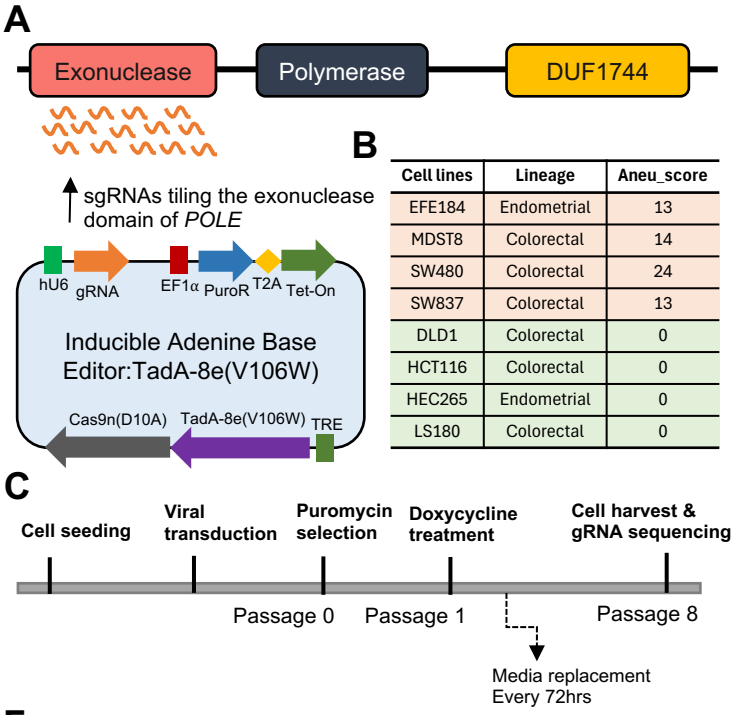
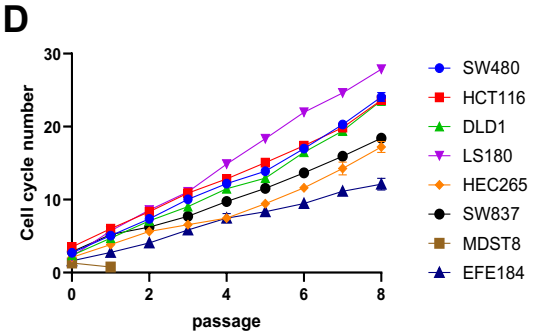

A
Anova, p < 2.2e−16
TMB
800
600
400
200
0
mut.exo
mut.other
wt
POLE status
B
R²=0.72, p=5.8e-10
Mean TMB
200
100
0
POLE mutated tumors
0
10
20
30
Aneuploidy score
D
-log10(p-Value)
30
25
20
15
10
5
0
0
0.5
1
1.5
Odds ratio
C
Diploid
Chromosomal instability leads to aneuploidy in cancer cells
Aneuploid
Deficiency in POLE exonuclease causes ultra-high mutations on the genome
Aneuploid
Cancer cell death: mutations caused by POLE exonuclease deficiency may have a high probability of inactivating essential genes on chromosome arms with loss due to aneuploidy.
X: loss of function mutations on essential genes with copy number loss
E
n=6
Probability of cell death per cell cycle with POLE exonuclease deficiency
1.0
0.5
0.0
α=0.1
α=0.01
α=0.001
0
10
20
30
40
n: # of arm loss (mean = 6)

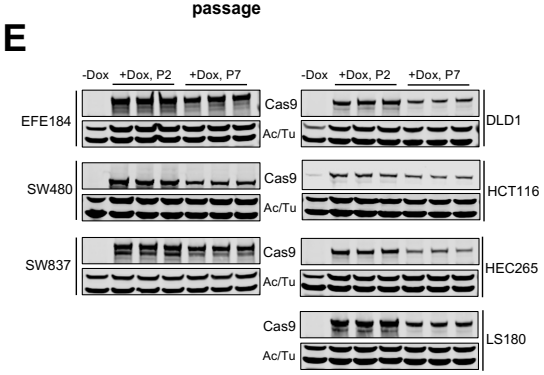
Figure 4

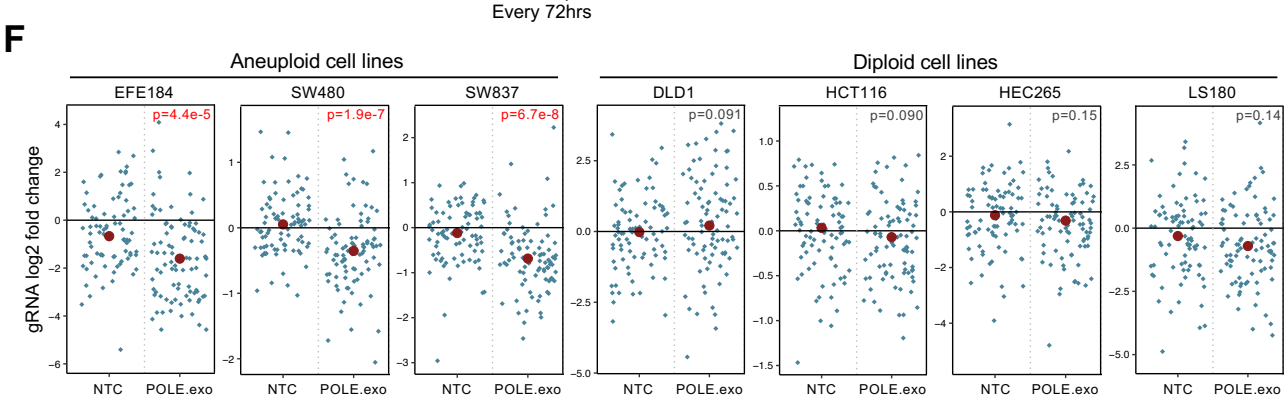


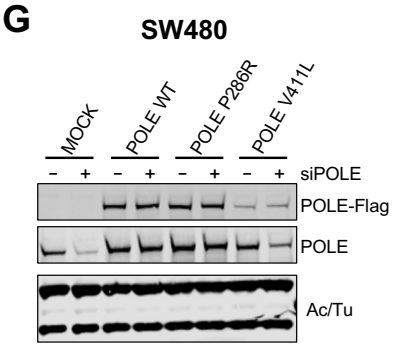
B

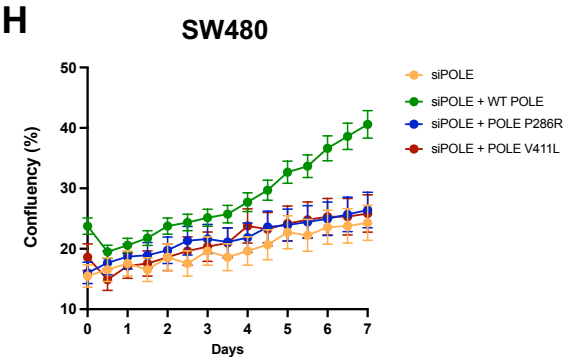
Cell lines	Lineage	Aneu_score
EFE184	Endometrial	13
MDST8	Colorectal	14
SW480	Colorectal	24
SW837	Colorectal	13
DLD1	Colorectal	0
HCT116	Colorectal	0
HEC265	Endometrial	0
LS180	Colorectal	0



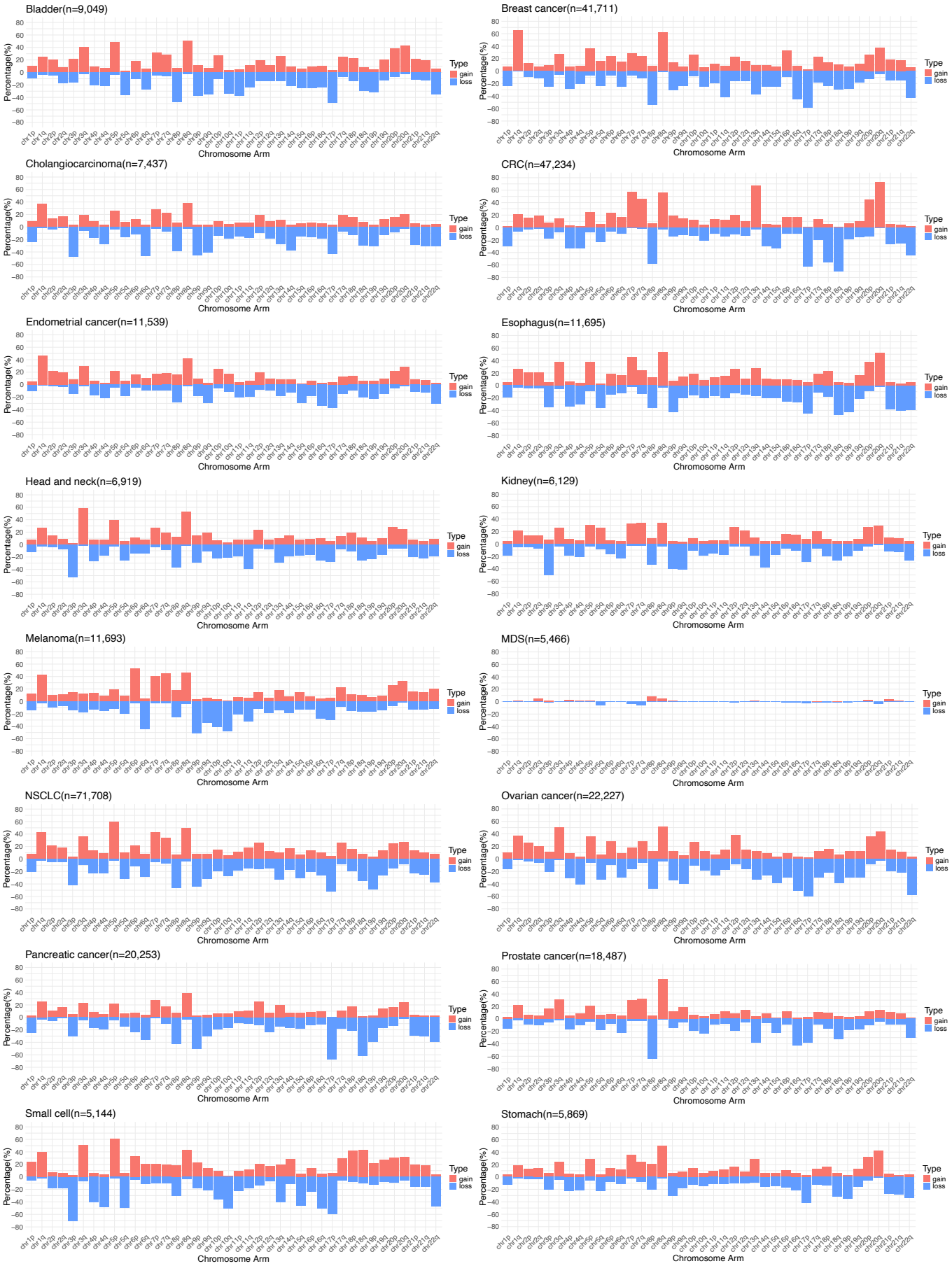




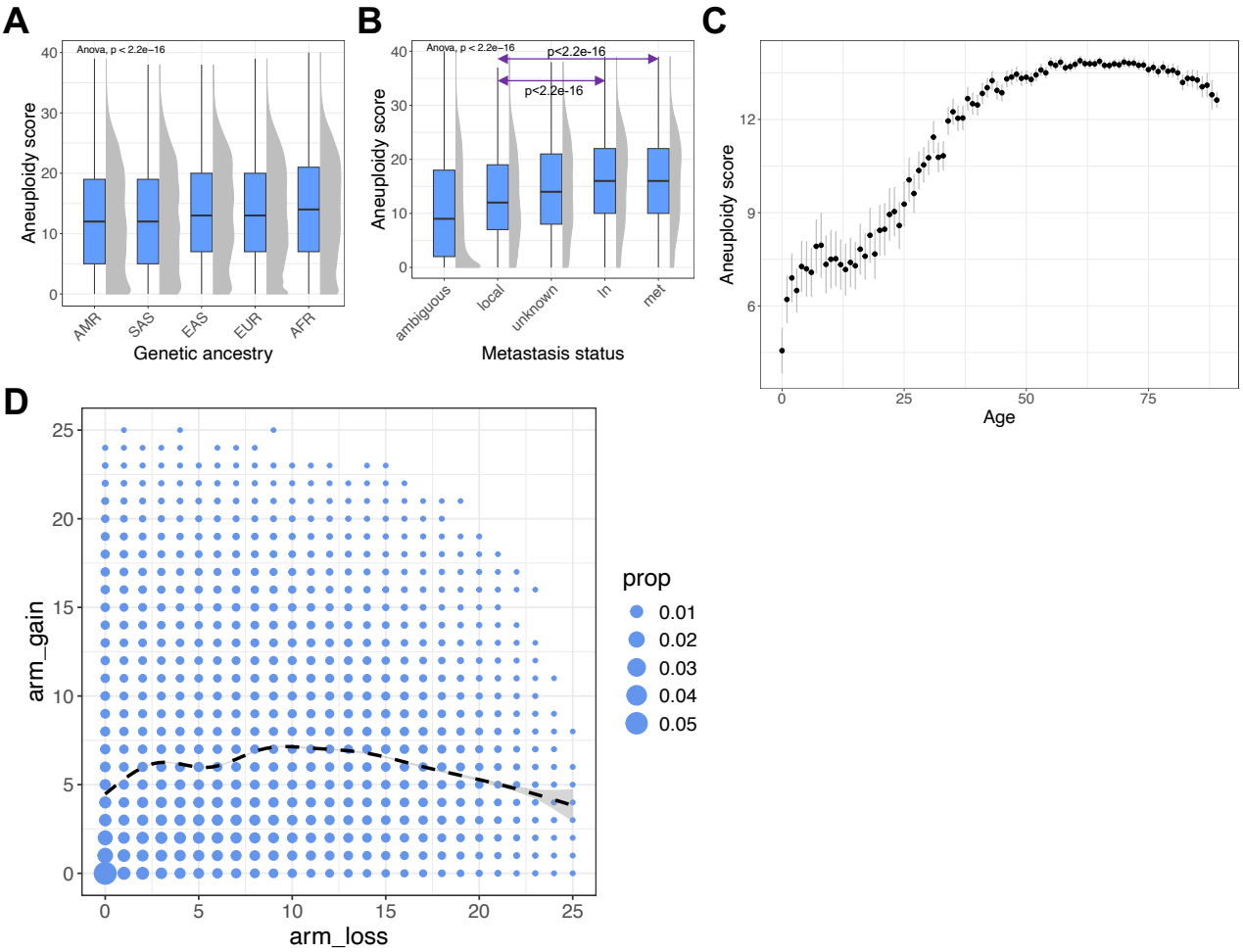




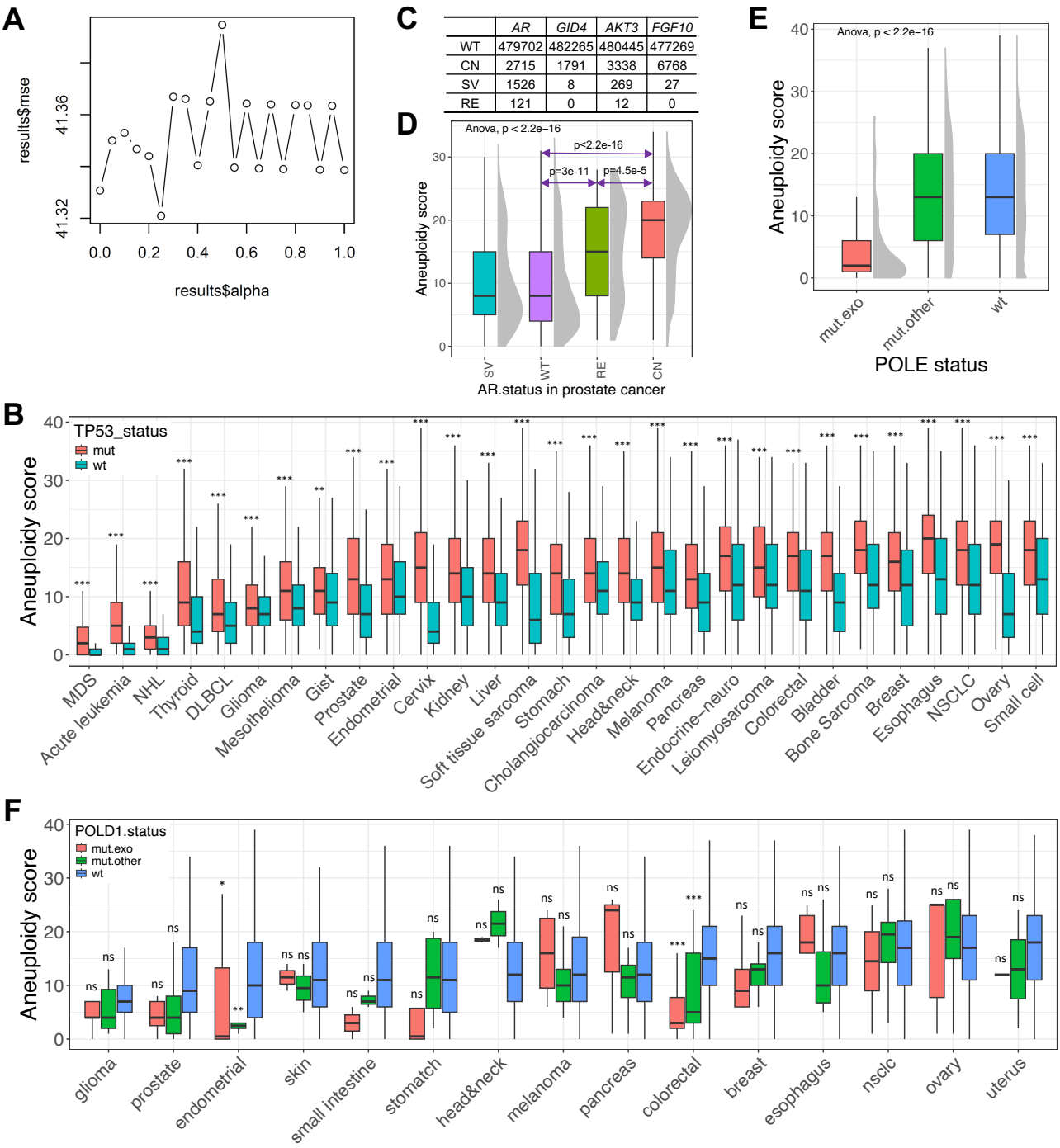
Supplemental figure 1



Supplemental figure 2

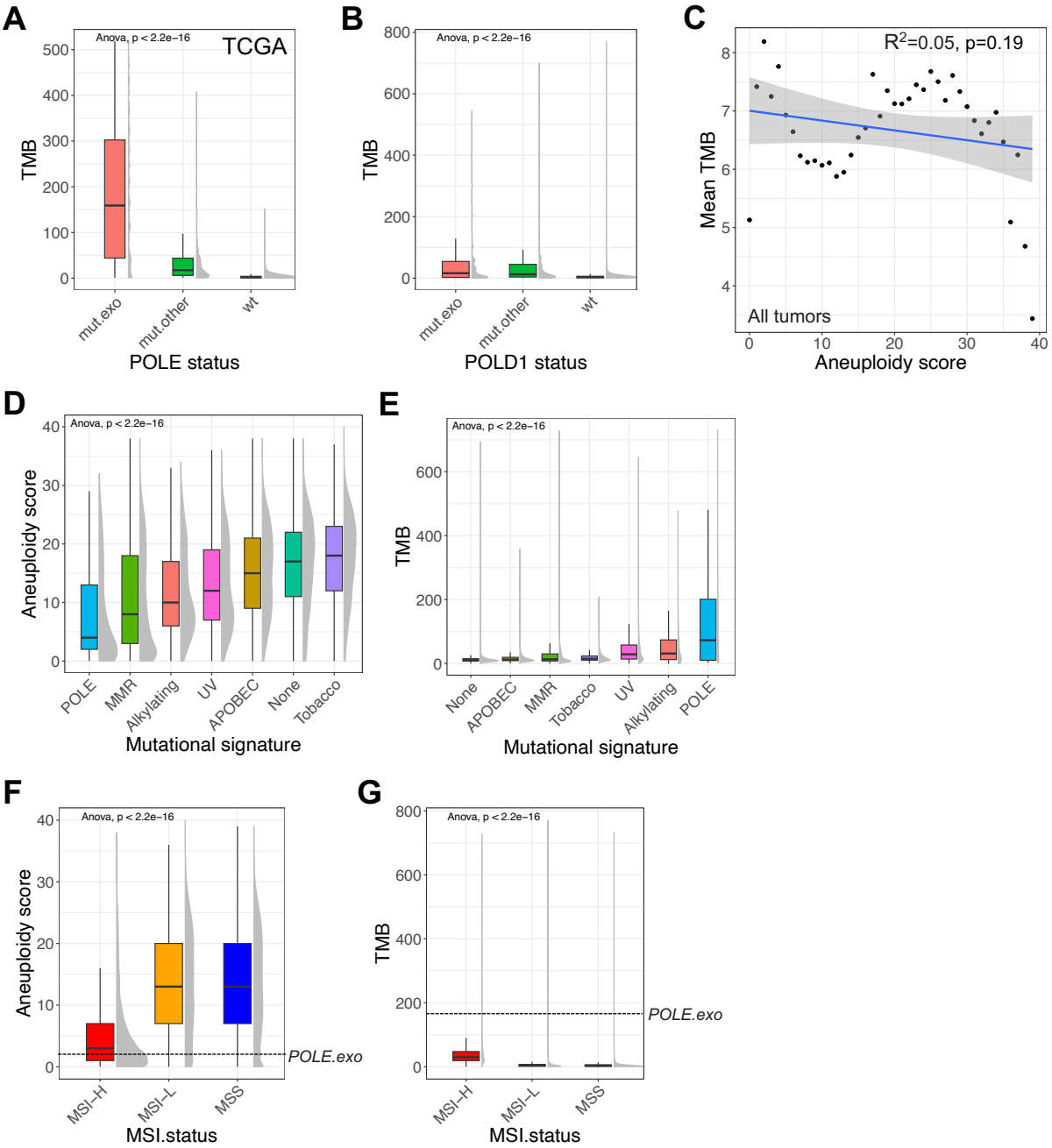


Supplemental figure 3



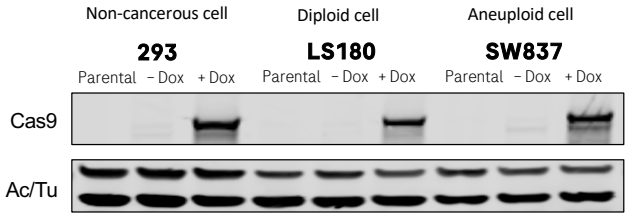
	AR	*GID4*	*AKT3*	*FGF10*
WT	479702	482265	480445	477269
CN	2715	1791	3338	6768
SV	1526	8	269	27
RE	121	0	12	0

Supplemental figure 4

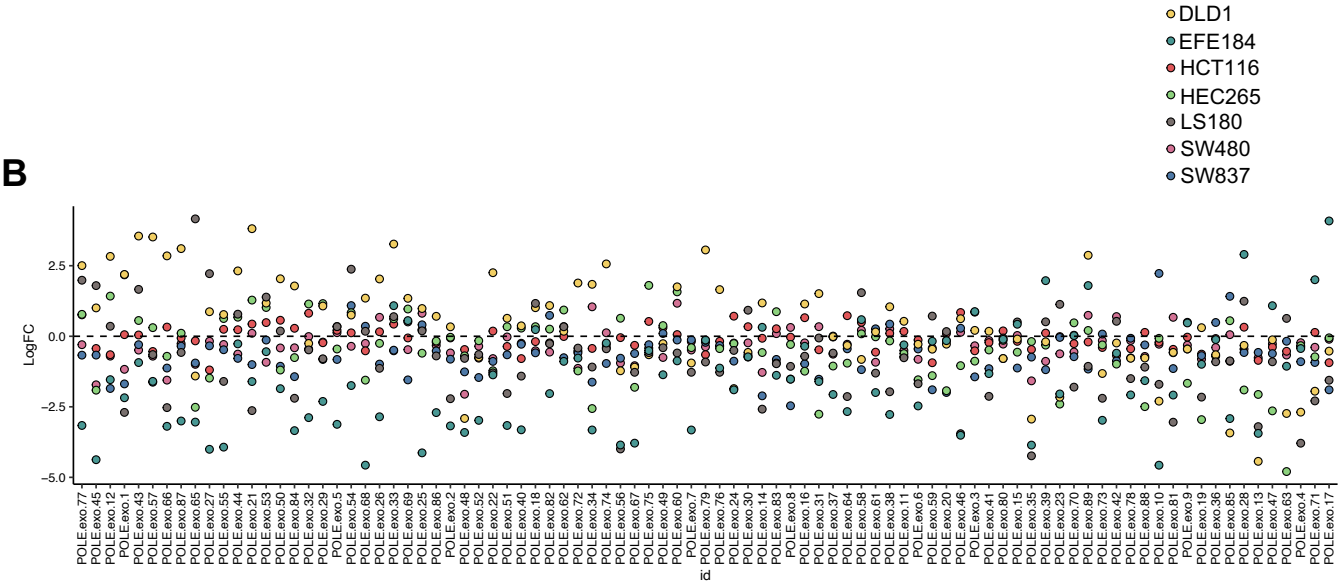
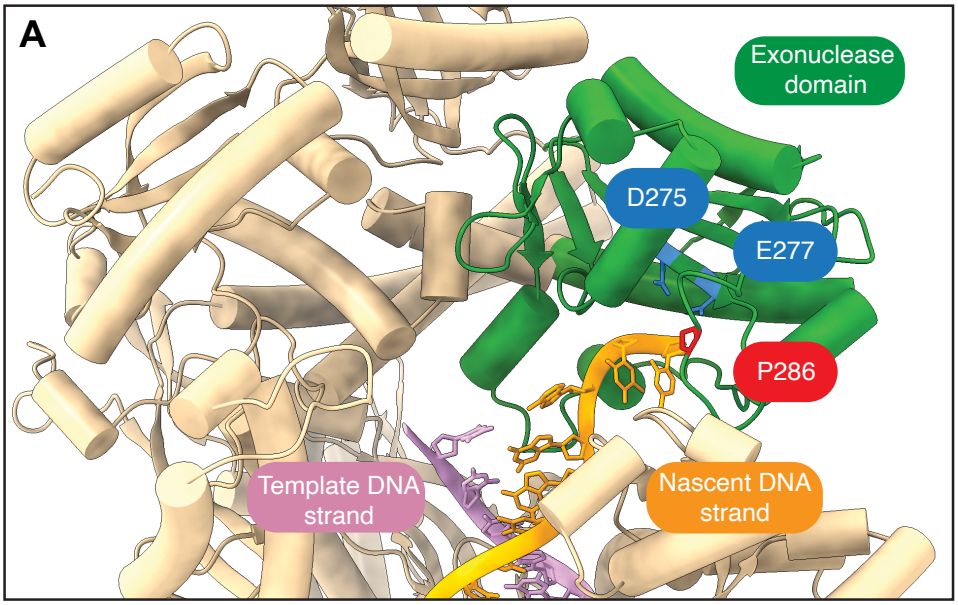
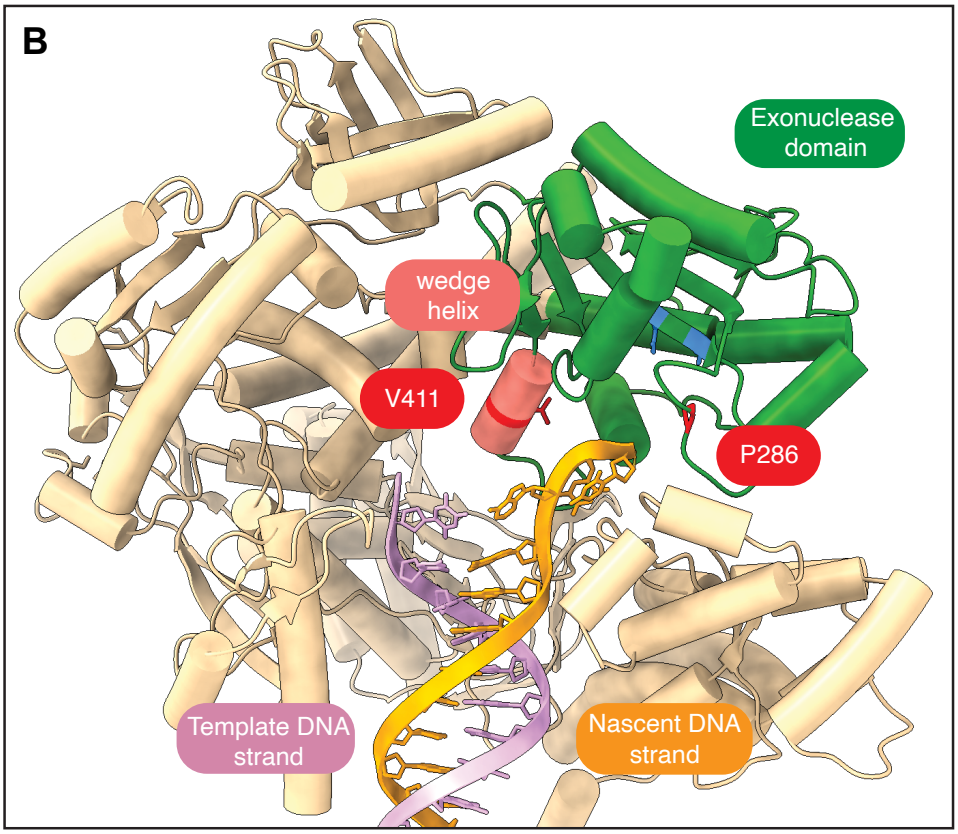


Supplemental figure 5

A



B



Supplemental figure 6