

# SCIENTIFIC REPORTS



OPEN

## Molecular profiling and sequential somatic mutation shift in hypermutator tumours harbouring POLE mutations

Keiichi Hatakeyama<sup>1</sup>, Keiichi Ohshima<sup>1</sup>, Takeshi Nagashima<sup>2,3</sup>, Shumpei Ohnami<sup>2</sup>, Sumiko Ohnami<sup>2</sup>, Masakuni Serizawa<sup>4</sup>, Yuji Shimoda<sup>2,3</sup>, Koji Maruyama<sup>5</sup>, Yasuto Akiyama<sup>6</sup>, Kenichi Urakami<sup>2</sup>, Masatoshi Kusuha<sup>7,4</sup>, Tohru Mochizuki<sup>1</sup> & Ken Yamaguchi<sup>8</sup>

Defective DNA polymerase  $\epsilon$  (POLE) proofreading leads to extensive somatic mutations that exhibit biased mutational properties; however, the characteristics of *POLE*-mutated tumours remain unclear. In the present study, we describe a molecular profile using whole exome sequencing based on the transition of somatic mutations in 10 *POLE*-mutated solid tumours that were obtained from 2,042 Japanese patients. The bias of accumulated variations in these mutants was quantified to follow a pattern of somatic mutations, thereby classifying the sequential mutation shift into three periods. During the period prior to occurrence of the aberrant *POLE*, bare accumulation of mutations in cancer-related genes was observed, whereas *PTEN* was highly mutated in conjunction with or subsequent to the event, suggesting that *POLE* and *PTEN* mutations were responsible for the development of *POLE*-mutated tumours. Furthermore, homologous recombination was restored following the occurrence of *PTEN* mutations. Our strategy for estimation of the footprint of somatic mutations may provide new insight towards the understanding of mutation-driven tumourigenesis.

Large-scale genomic studies have revealed the rare occurrence of a diversity of mutation frequency and somatic hypermutation in specific tumours, which are termed ‘hypermutators’<sup>1–3</sup>. This mutator effect in colorectal and endometrial cancers is occasionally accompanied by mutations in the exonuclease domain of DNA polymerase epsilon (*POLE*)<sup>2–4</sup>. The *POLE* gene is responsible for the replication of a leading DNA strand<sup>5–7</sup>; hence, the defective proofreading consequent to its mutation is considered to generate extensive somatic mutations that constitute a distinct mutation spectrum<sup>8</sup>. In particular, p.P286R and p.V411L substitutions lead to loss of *POLE* proofreading activity<sup>8,9</sup>. Furthermore, mutations in the exonuclease domain of *POLE* are observed in microsatellite stable (MSS) tumours that are characterised mainly by a biased prevalence of TCT > TAT and TCG > TTG mutation patterns<sup>10</sup>. Although the characteristics of *POLE* mutation are well understood, that of its associated tumours remains unclear, in large part because in previous reports, tumours harbouring this mutant have been specifically eliminated from the interpretation of genomic analysis as an anomalous case<sup>2,3</sup>.

Nearly three decades ago, Vogelstein *et al.* proposed that sequential mutations in specific oncogenes and tumour suppressor genes (TSGs) lead to tumourigenesis<sup>11</sup> based on the two-hit hypothesis formulated by Knudson<sup>12</sup>. As such mutations are observed in many tumours, this concept is still widely accepted. However, it is difficult to directly trace the sequential shift of somatic mutations in a solid tumour that occurs in a patient.

<sup>1</sup>Medical Genetics Division, Shizuoka Cancer Center Research Institute, Sunto-gun, Shizuoka, 411-8777, Japan.

<sup>2</sup>Cancer Diagnostics Research Division, Shizuoka Cancer Center Research Institute, Sunto-gun, Shizuoka, 411-8777, Japan. <sup>3</sup>SRL Inc., Shinjuku-ku, Tokyo, 163-0409, Japan. <sup>4</sup>Drug Discovery and Development Division, Shizuoka Cancer Center Research Institute, Sunto-gun, Shizuoka, 411-8777, Japan. <sup>5</sup>Experimental Animal Facility, Shizuoka Cancer Center Research Institute, Sunto-gun, Shizuoka, 411-8777, Japan. <sup>6</sup>Immunotherapy Division, Shizuoka Cancer Center Research Institute, Sunto-gun, Shizuoka, 411-8777, Japan. <sup>7</sup>Regional Resource Division, Shizuoka Cancer Center Research Institute, Sunto-gun, Shizuoka, 411-8777, Japan. <sup>8</sup>Shizuoka Cancer Center, Sunto-gun, Shizuoka, 411-8777, Japan. Correspondence and requests for materials should be addressed to K.H. (email: [k.hatakeyama@scchr.jp](mailto:k.hatakeyama@scchr.jp))

<sup>9</sup>Shizuoka Cancer Center, Sunto-gun, Shizuoka, 411-8777, Japan. <sup>10</sup>Shizuoka Cancer Center, Sunto-gun, Shizuoka, 411-8777, Japan. <sup>11</sup>Shizuoka Cancer Center, Sunto-gun, Shizuoka, 411-8777, Japan. <sup>12</sup>Shizuoka Cancer Center, Sunto-gun, Shizuoka, 411-8777, Japan.

Accordingly, in the majority of large-scale genomic studies, the mutation footprint is predicted only from an endpoint of base substitutions in genes determined after lesion resection.

The molecular profile in a normal cell is dramatically changed upon the occurrence of mutation in a specific gene as an early event of tumorigenesis. In cultured human intestinal stem cells, artificial mutation of *APC*, *TP53* (also known as P53), and/or *KRAS* by CRISPR/Cas9 technology is sufficient to serve as a hallmark of tumour progression<sup>13</sup>. Hence, estimation of the period when key gene mutations arise and investigation of cancer-related genes that mutated prior to or concomitant with this event are of considerable importance to the understanding of mutation-driven tumorigenesis.

In the current study, to reveal the molecular profile and sequential mutation shift of hypermutators harbouring a *POLE* mutant, we investigated the mutation pattern and gene expression based on biased base substitutions. *POLE*-mutated samples were first extracted from more than 2,000 Japanese patients with cancer followed by confirmation of whether these tumours possessed characteristics similar to those reported in previous studies. We next revealed that the susceptibility to effects of the *POLE* mutant differed according to gene based on the pattern of base substitutions. The sequential mutation shift of genes that accumulated in tumours with *POLE* mutants were then classified into three periods. *POLE* and phosphatase and tensin homolog (*PTEN*) mutations were therefore predicted to influence early tumorigenesis. Furthermore, defective homologous recombination (HR) was restored in comparison with that in common hypermutator tumours, and somatic mutations were found to accumulate in pathways associated with hypoxia-inducible factor, implying their role in maintaining HR repair activity. Overall, our analysis using the biased base substitution paradigm allows an accelerated understanding of mutation-driven tumorigenesis.

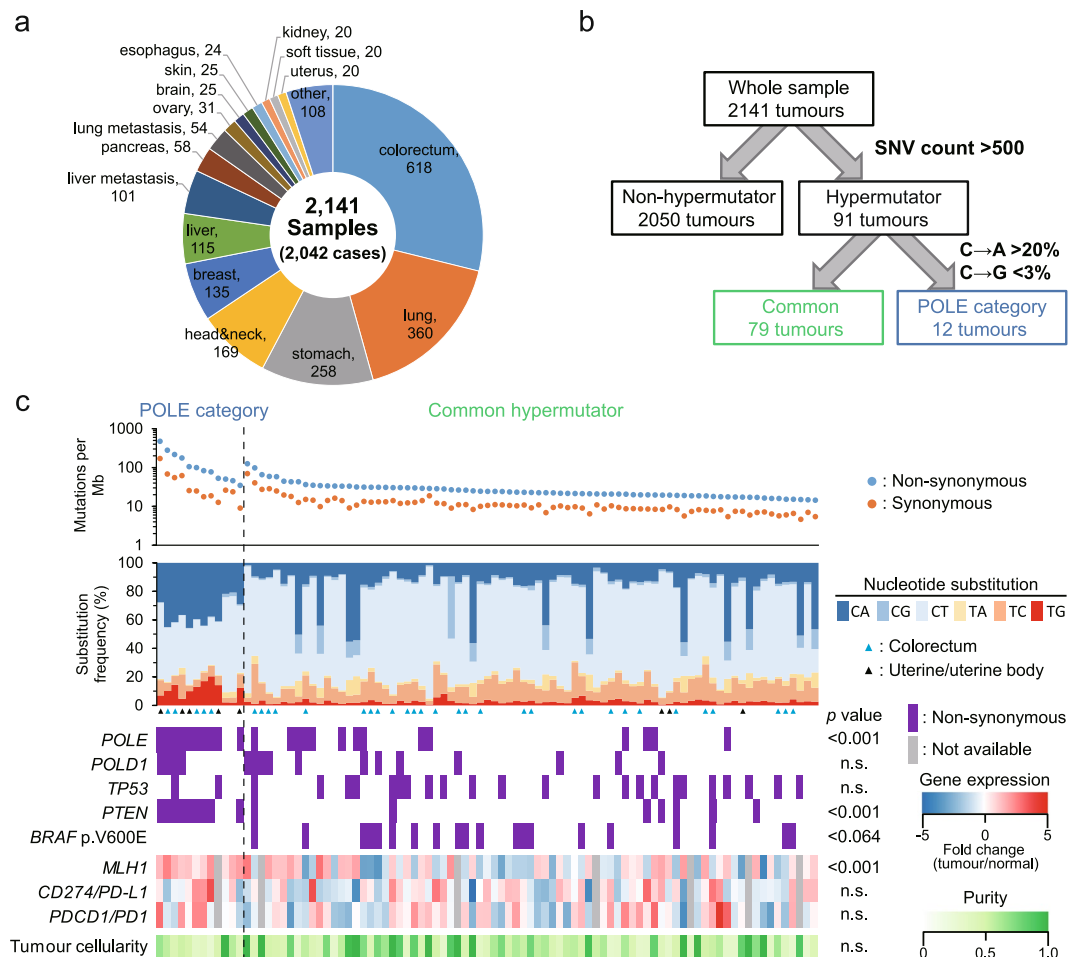
## Results

**Molecular subtypes of hypermutators.** We performed whole exome sequencing (WES) using 2,141 solid tumour samples derived from 2,042 patients with cancer. All samples collected from our single hospital were comprised of multiple tissues, among which colorectal, lung, and stomach cancers occupied 58% of the whole content (Fig. 1a). The tumours exceeding 500 counts of non-synonymous single nucleotide variation (SNV) were grouped as hypermutators (Fig. 1b). A total of 91 samples (4.3%) belonged to this group, yielding a frequency that was similar to that in a previous large-scale analysis<sup>1</sup>. The hypermutators were then classified into two subtypes according to nucleotide substitution frequency and pattern<sup>2</sup>. The tumours that met the criteria were defined as *POLE* category in this study, whereas the remaining hypermutators were categorised as common.

**Landscape of *POLE* category and hypermutators.** To further clarify the differences between the two subtypes, gene expression profiling (GEP) and an estimation of tumour cellularity were performed along with mutation analysis (Fig. 1c). The number of mutations (per Mb) in *POLE*-category tumours was significantly higher than in common hypermutators ( $p = 1.14 \times 10^{-2}$  by Welch's *t*-test). Additionally, patients classified with *POLE* category tumours were significantly younger than those of with hypermutator tumours (Supplementary Fig. S1). In *POLE* category, tumours derived from the colorectum and uterus were dominant except in two samples of soft tissue that were dissected from the same patient with a diagnosis of melanoma. These samples without *POLE* mutations were checked for 30 mutational signatures<sup>1</sup> using deconstructSigs<sup>14</sup>, indicating the presence of predominantly ultraviolet-induced Signature.7 (data not shown). To characterise *POLE* mutant tumours, the samples without *POLE* mutation were removed from consideration in this study. Consistent with a previous report<sup>2</sup>, accumulation of *PTEN* mutation and up-regulation of *MLH1* were observed in *POLE*-category tumours (Fig. 1c). No mutation of *BRAF* p.V600E was found; additionally, no differences of gene expression on immune checkpoint molecules (PD-L1 and PD1) or tumour cellularity were observed.

**Propensity of mutation in *POLE*-category tumours.** Although a common hypermutator is frequently characterised as having microsatellite instability (MSI), a tumour harbouring *POLE* mutations was described as belonging to MSS cancer type<sup>4</sup>. MSI testing was therefore performed in *POLE*-category tumours, indicating that all samples were MSS (Table 1). The exonuclease domain of *POLE* and/or *POLD1* was mutated in all samples of this category except those derived from soft tissue. The *POLE* p.V411L or p.P286R substitutions are known to generate a relatively large excess of C > A/T transversion in the specific three-letter base motifs (TCT and TCG) owing to defects in DNA proofreading<sup>8,10</sup>. A similar propensity was observed in the *POLE*-category group (Fig. 2a) and these were especially accumulated in tumours harbouring *POLE* p.P286R and p.V411L (Fig. 2b). Furthermore, T > G substitution tended to preferentially occur with neighbour thymines (i.e., TTT). These results indicated that *POLE* p.P286R and p.V411L generated a biased substitution depending on the surrounding sequence.

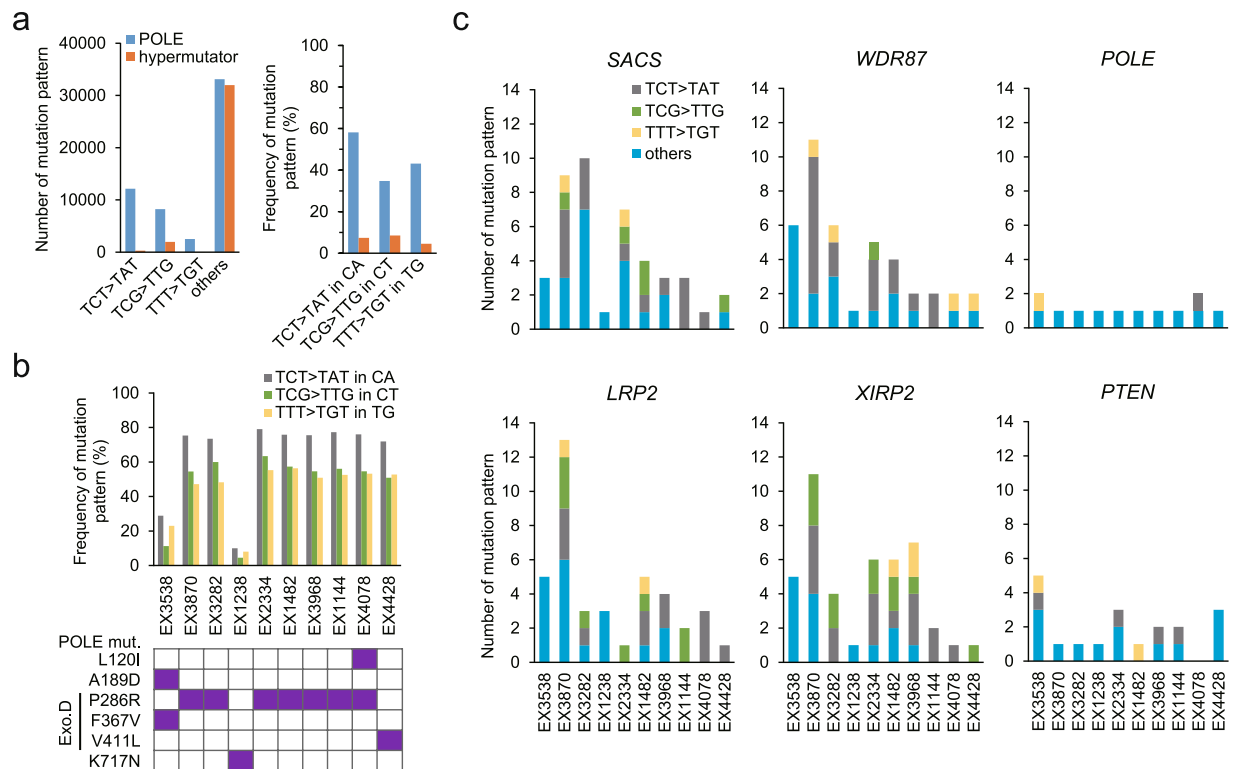
**Identification of variation-susceptible genes.** To identify genes susceptible to *POLE* mutant effects, we counted the mutation patterns in individual tumours. The biased mutation patterns TCT > TAT, TCG > TTG, and TTT > TGT were highly accumulated in *SACS*, *LRP2*, *WDR87*, and *XIRP2* genes, whereas no accumulation of these patterns was observed in *POLE* and *PTEN* (Fig. 2c). In these genes, the biased mutation patterns were detected in all *POLE* category tumours with *POLE* p.P286R or p.V411L. To remove any concern of inherent bias derived from gene sequence, all patterns of three-letter bases in the exons of each gene were tallied and the occurrence frequency of mutation patterns was then estimated. Although these genes contained low frequencies of TCG sequence, a cytosine in this motif was highly mutated compared to other sites except in *POLE* and *PTEN* (Supplementary Fig. S2a). Furthermore, over 80% of TCG and TCT motifs evidenced the biased transversion; i.e., propensity mutation (Supplementary Fig. S2b). These results indicated that propensity mutation via *POLE* p.P286R or p.V411L was less affected by the appearance frequency of three-letter base sequences. Conversely, in a public repository (COSMIC, <http://cancer.sanger.ac.uk/cosmic>), the mutation patterns tended to exhibit a similar



**Figure 1.** Sample classification and mutation spectra. **(a)** Distribution of tumour types included in this analysis. The ‘other’ group contains multiple tumour types that comprise less than 20 samples. **(b)** All samples were stratified into three groups according to the nucleotide substitution frequency and pattern. SNV, single nucleotide variation. **(c)** Mutation frequencies (vertical axis, top panel) plotted for POLE-category and common hypermutator (horizontal axis) tumours. Nucleotide substitutions are shown in the middle panel. Tumours of the colorectum and uterine/uterine body are indicated by red and black arrowheads, respectively. Mutation pattern, expression profile, and tumour cellularity are represented in the bottom panel. Gene expression is analysed and coloured based on fold change (tumour/normal). In cases where tumour type created difficulty with respect to dissection of the normal sample, a grey-filled rectangle (not available) is shown in the profile. The *p* values of mutation and expression/cellularity are calculated using Fisher’s exact test and Welch’s *t*-test, respectively. n.s., not significant.

Substitution		No. of SNVs	MSI	Tissue
POLE	POLD1			
p.A189D, p.F367V	<b>p.R244H</b>	16632	MSS	uterus
p.K717N	<b>p.E318G</b>	6155	MSS	uterus
<b>p.P286R</b>	—	3686	MSS	uterine body
p.L120I, <b>p.P286R</b>	—	1846	MSS	uterus
<b>p.V411L</b>	—	1212	MSS	uterus
<b>p.P286R</b>	p.A706T	9782	MSS	colon (transverse)
<b>p.P286R</b>	<b>p.D402N</b>	7593	MSS	colon (ascending)
<b>p.P286R</b>	—	3491	MSS	colon (rectum)
<b>p.P286R</b>	—	2903	MSS	colon (rectum)
<b>p.P286R</b>	—	2699	MSS	colon (rectum)

**Table 1.** POLE mutations in POLE-category tumours. The substitutions located in the exonuclease domain in POLE/POLD1 are represented in bold text. Microsatellite instability (MSI) was checked using PCR of five microsatellite markers (*BAT25*, *BAT26*, *NR21*, *NR24*, and *MONO27*) by an outsourcing company. SNV, single nucleotide variation (non-synonymous); MSS, microsatellite stable.

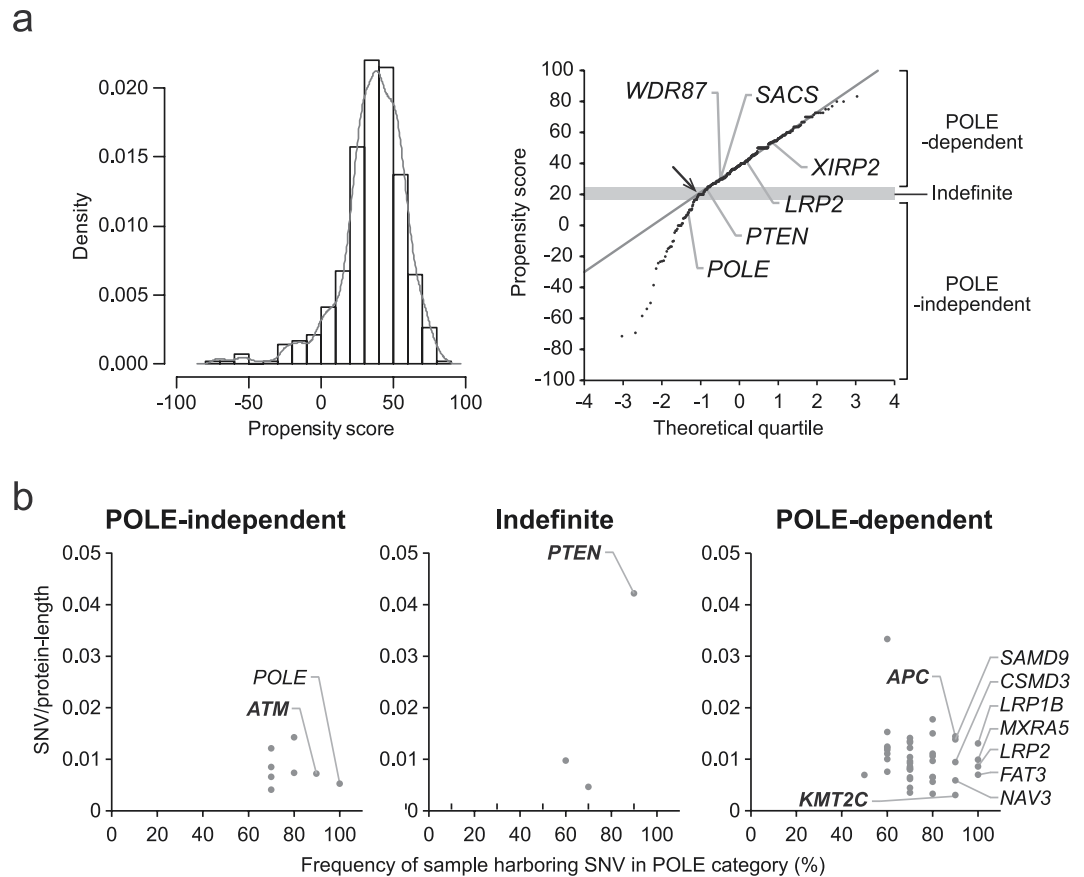


**Figure 2.** Propensity of mutation patterns in POLE-category tumours. **(a)** Number and frequency of mutation patterns in POLE-category and common hypermutator tumours. **(b)** Individual frequency of mutation patterns and amino acid substitution on POLE. The mutation is shown as purple-filled cells. Exo D, exonuclease domain. **(c)** Distribution of mutation patterns in *SACS*, *WDR87*, *LRP2*, and *XIRP2* genes that were specifically enriched for propensity mutations in the POLE category. *POLE* and *PTEN* are shown as known mutation-enriched genes in this category.

propensity (Supplementary Fig. S3a and S3b). In this study, *SACS*, *LRP2*, *WDR87*, and *XIRP2* were therefore evaluated as highly susceptible genes to POLE mutant status.

**Gene classification by POLE-category specific mutations.** Based on the mutation pattern, we found four genes that were highly mutated by the POLE mutant. To further identify other variation-susceptible genes, propensity score (PS) was determined and calculated using the frequency of mutation patterns in individual genes within POLE-category and common hypermutator tumours (see Methods, equations (1) and (2)). The PS based on the pattern frequency was independent of length of coding region. In particular, low PS indicates that a gene was mutated independently of a POLE mutant, whereas high PS specifies that the propensity mutation was accumulated dependently of the aberrant *POLE*. The workflow of classification using PS is shown in Supplementary Fig. S4. Briefly, candidates harbouring POLE-category specific mutations were first extracted prior to the calculation of PS to filter out extraneous mutated genes (listed in Supplementary Table S1). Then, low-mutation-number genes (SNV < 10) in the POLE-category group were removed, resulting in 414 genes isolated. The contribution of propensity sequence (TCT and TCG) appearance frequency in the isolated genes to PS was negligible (Supplementary Fig. S6a). Variant allele frequency (VAF) of propensity mutation patterns (TCT > TAT and TCG > TTG) was additionally calculated in these genes (Supplementary Fig. S5). Next, inflection point (IP) was derived from the distribution and quantile-quantile (Q-Q) plot of PS (Fig. 3a) based on segmented regression (standard linear model; break-point, 1; see Methods). In order to construct a confidence interval (CI) of IP, we employed a bootstrap method. Finally, genes in the CI of IP were classified in the indefinite period; the remaining low PS (<lower-CI) genes were determined as POLE-independent, and the high PS ( $\geq$ upper-CI) genes were considered POLE-dependent. We expected that the four genes (*SACS*, *LRP2*, *WDR87*, and *XIRP2*) that were detected in all POLE category tumours with *POLE* p.P286R or p.V411L would exhibit high PS if this score reflected an accumulation of biased mutation patterns. Consistent with this supposition, these genes were grouped in the POLE-dependent category. This result suggested that the classification into three periods using PS was satisfactory to isolate genes susceptible to POLE mutant activity.

A common hypermutator is frequently characterised as having MSI. To confirm the influence of this status on PS, the mutation pattern in other MSI high/low samples was checked as well as in common hypermutators derived from colon and uterus tumours. Tumours with MSI high were extracted based on 30 mutational signatures<sup>1</sup> assessed using deconstructSigs (Signature.6 (MSI) > 0.5)<sup>14</sup>. Accumulation of the POLE-specific mutation patterns (TCT > TAT, TCG > TTG, and TTT > TGT) was detected at less than 10% in all groups (Supplementary Fig. S6b). Moreover, PS is composed only of FMPs (propensity mutations/total mutations in each group, see



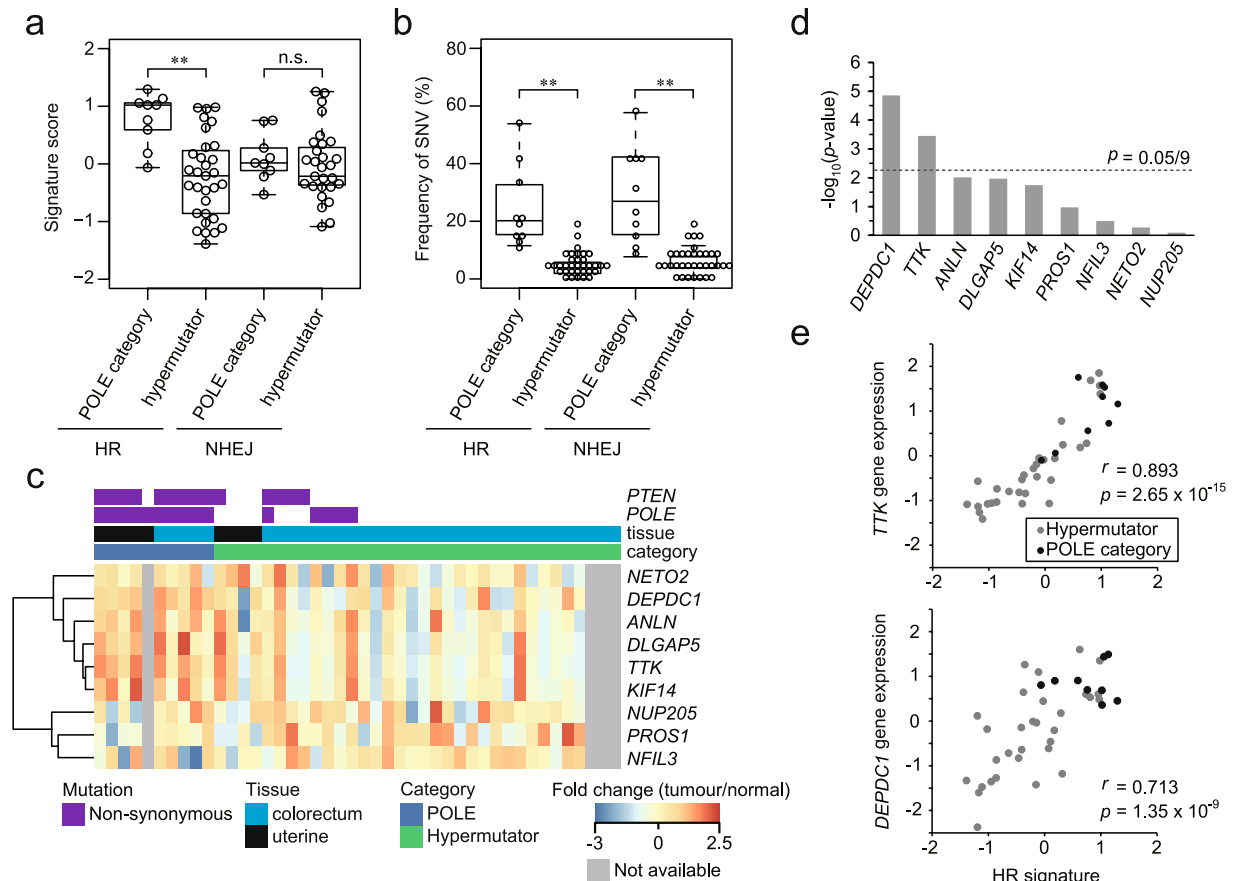
**Figure 3.** Classification of POLE-category-specific mutations into POLE-independent and -dependent mutations. **(a)** Frequency density (left) and quantile-quantile (Q-Q) plot (right) of propensity score (PS) in the gene harbouring POLE-category-specific mutations. Gene symbols correspond to Fig. 2c. Inflection point (IP) in the Q-Q plot is indicated by an arrow and is defined as the threshold of PS to separate genes possessing POLE-independent and -dependent mutations. The genes in the grey rectangle (confidence interval of IP) are classified in the group of indefinite mutation as a border for buffering. **(b)** Comparison of the mutation rates for protein coding regions between cancer-related genes in the classified groups. Gene symbols are represented when more than 80% of mutations could be classified as POLE category. Tumour suppressor genes are shown with a bold symbol.

Methods). We performed correlation analysis of all FMPs including MSI-high/low groups, demonstrating that no significant correlation was observed between FMPs (Supplementary Fig. S6c). These results indicated that the tumour MSI status had no influence on PS.

**Discovery of variation-susceptible tumour suppressor genes.** Using PS based on the propensity of POLE mutant effects, POLE-category specific genes were classified according to the three periods. To estimate the impact of mutations generated by the POLE mutant on tumourigenesis, we analysed the substitution frequency of cancer-related genes. This gene set was curated in-house using multiple public databases (Supplementary Table S1). Additionally, TSGs and oncogenes were defined based on Vogelstein's list<sup>15</sup>. No mutations of TSGs and oncogenes were observed in the POLE-independent period, whereas *PTEN*, *ATM*, *APC*, and *KMT2C* were mutated in both the indefinite and POLE-dependent periods (Fig. 3b). Notably, the non-synonymous mutation rate per protein-length of *PTEN* was higher than that of other TSGs.

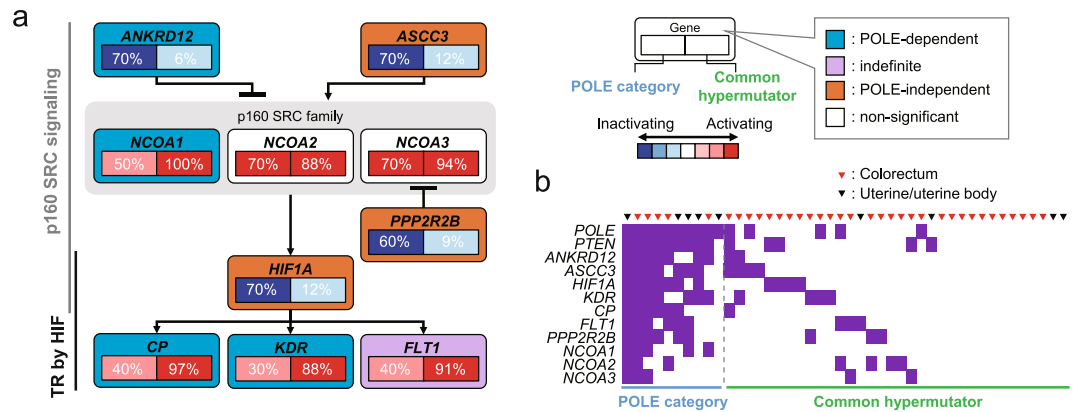
**Validation of propensity mutations using public data.** To validate the variation susceptibility of POLE-category specific genes, classification using PS was performed in the International Cancer Genome Consortium (ICGC) data set. The PS distribution that was similar to our study was classified into three periods to visualise sequential mutation shift (Supplementary Fig. S7a). The mutation accumulation in five genes (*POLE*, *PTEN*, *CSMD3*, *APC*, and *NAV3*) was consistent with our data set. In POLE-category tumours derived from ICGC data, *POLE* and *PTEN* accumulated POLE-independent mutations (Supplementary Fig. S7b). This result suggested that *POLE* and *PTEN* were almost unsusceptible to defective POLE function.

**Estimation of HR activity.** Aberrant *PTEN* often leads to dysfunction of HR<sup>16-18</sup>. The HR activity in POLE category tumours was first evaluated based on expression signatures. HR and non-homologous end joining (NHEJ) signatures were composed of five (*BRCA1*, *RAD51*, *BRCA2*, *RAD54L*, and *RAD52*) and four (*LIG4*,



**Figure 4.** Expression profiles of homologous recombination (HR)-related genes in POLE-category tumours. Gene signature analysis (**a**) and mutation frequency (**b**) of HR and non-homologous end joining (NHEJ) genes between POLE-category and common hypermutator tumours. Frequency of SNV represents accumulation of mutations found in the related gene sets (52 and 27 genes in HR and NHEJ, respectively)<sup>20,21</sup>. All SNVs were counted in each sample. n.s., not significant;  $**p < 0.01$ . (**c**) Heat map of genes that were up-regulated in the reversal of HR deficiency. The column represents each sample in POLE category and common hypermutators derived from colon and uterus. In cases where tumour type created difficulty with respect to dissection of the normal sample, a grey-filled rectangle (not available) is shown in the profile. (**d**) Bar chart representing the significant genes that are compared between POLE-category and common hypermutator tumours in the heat map (**c**). To counteract the problem of multiple comparisons in the expression profile,  $p$ -value was adjusted with Bonferroni correction. (**e**) Correlation between HR signature and significantly up-regulated genes in the bar chart (**d**).

*PRKDC*, *XRCC4*, and *XRCC6*) genes that up-regulated activation of these functions, respectively<sup>19</sup>. The HR signature in POLE mutants was high in comparison with that in common hypermutator tumours ( $p = 6.06 \times 10^{-5}$ , Fig. 4a), whereas mutations of HR-/NHEJ-related genes<sup>20,21</sup> were accumulated in POLE-category tumours (Fig. 4b), implying that HR activity in tumours harbouring defective POLE were elevated despite mutation accumulation in HR genes. To confirm the influence of MSI status, we compared these signatures and the frequency of SNV between MSI high/low samples in common hypermutators. MSI high were extracted using deconstruct-Sigs<sup>14</sup> (Signature.6 (MSI)  $> 0.5$ ). No significant differences were observed between MSI high and MSI low subsets among common hypermutators (Supplementary Fig. S8a,b). Additionally, no significant impact of tissue distribution was found in POLE category and common hypermutator tumours (Supplementary Fig. S8c,d). These results indicated that MSI status and tissue distribution had no influence on HR-/NHEJ-signatures and mutations in common hypermutators. As *PTEN*-deficient tumours with mutations of HR-related genes are considered as a reversal of HR deficiency<sup>22</sup>, we therefore next investigated nine genes up-regulated in the process of HR repair restoration (in the previous report<sup>22</sup>,  $\geq 0$  score in  $\geq 75\%$  samples). Six genes (67%) tended to be highly expressed in POLE-category tumours, with *DEPDC1* and *TTK* genes in particular exhibiting significant differences in expression (Fig. 4c,d). Additionally, expression of *TTK*, which can increase HR repair<sup>22</sup>, was strongly positively correlated with HR signature (Fig. 4e). To further confirm whether *PTEN* mutated tumours restored HR repair deficiency, we investigated the relationship between *PTEN* mutation and HR-related genes (HR signature, *TTK* and *DEPDC1*) in colon and uterine tumours including non-hypermutators. These genes were found to be up-regulated in *PTEN* mutated tumours (Supplementary Fig. S9). These results suggested that POLE-category tumours with *PTEN* mutations restored HR repair deficiency through *TTK* overexpression.



**Figure 5.** Pathway alterations in POLE-category tumours. **(a)** The p160 steroid receptor co-activator (SRC) signalling and transcriptional regulation (TR) by hypoxia inducible factor (HIF) pathways are altered through accumulation of somatic mutations. Alteration frequencies are expressed as a percentage of all cases. The genes that were significantly mutated between our study and public repository are represented. **(b)** Mutation patterns in the p160 SRC signalling and TR by HIF pathways.

**Pathway alteration in POLE-category tumours.** To further characterise POLE-category tumours, we investigated pathways accumulating POLE-category-specific mutations between our data and ICGC data sets. Mutation accumulation in genes associated with transcriptional regulation (TR) by hypoxia inducible factor (HIF) and p160 steroid receptor co-activator (SRC) signalling pathways were observed in both data sets (Supplementary Fig. S10). In the two pathways, we identified eight genes that accumulated POLE-category-specific mutations (Fig. 5a and b). The pathways were linked via HIF1A, which is categorised as POLE-independent, and the frequency of mutation in the eight genes was similar to that in the ICGC data set (Supplementary Fig. S11). These results suggested that the two pathways linked via HIF1A were compromised by excess accumulation of POLE category mutations.

## Discussion

In tumours exhibiting a hypermutator phenotype, the presence of a subtype that possesses different characteristics has been revealed by a team of The Cancer Genome Atlas Network<sup>2,3</sup>. This subtype harbours POLE mutations and is predominantly found in colorectal and uterine tumours. We also identified hypermutators harbouring such mutations from the same tissue types that had been collected from more than 2,000 Japanese patients with cancer. These mutators were designated as POLE category in the current study, and were considered according to the generation of biased substitutions such as C > A high and C > G low<sup>8</sup>. In addition, a high frequency of C > A with accompanying C > G has been observed in the common hypermutator tumour type derived from the lung, indicating a signature of tobacco exposure<sup>23</sup>. Although the characteristics of the POLE category were consistent with those of previous reports, our analysis revealed that this category was characterised by not only the staggering number of mutations but also by early onset colorectal or uterine tumour in comparison with the common hypermutator tumours. A similar finding was also observed in non-Japanese patients<sup>24</sup>. Furthermore, Shlien *et al.* recently reported that a patient with an inherited biallelic mismatch repair deficiency developed early onset brain tumours harbouring somatic POLE mutations<sup>25</sup>. Therefore, POLE-category status in the current study may be influenced by germline mutational status. Germline mutation should be also checked and validated in further studies.

Sporadic MSI usually arises from epigenetic silencing of the *MLH1* promoter via a global increase in CpG island methylation<sup>26,27</sup>, which has been correlated to down-regulation of its gene<sup>28</sup>. We checked MSI in POLE-category tumours using *MLH1* expression in GEP and by PCR of microsatellite markers, demonstrating loss of *MLH1* silencing and MSI negative status, respectively. Furthermore, although typical MSI is associated with somatic *BRAF* p.V600E status<sup>29</sup>, this mutation was not found in the POLE category. Previously, it has been demonstrated that an endometrial or colorectal cancer harbouring POLE mutations without MSI induces immune-suppression through inhibition of the PD1/PD-L1 axis<sup>30,31</sup>. Therefore, although MSI testing has attracted attention as a biomarker for PD-1 blockade<sup>32</sup>, an assessment of somatic mutation frequency and POLE mutation is instead recommended at least for patients with colorectal or uterine cancer.

We identified four genes (*SACS*, *LRP2*, *WDR87*, and *XIRP2*) as being highly susceptible to the POLE mutant in our analysis. Although several biased mutations were observed in these genes, a diverse profile pattern appeared. For example, TCT > TAT was found to be accumulated in *WDR87* in both this study and in a public repository. This propensity implies that susceptibility to POLE mutant status depended on the individual gene. To evaluate this susceptibility variation, we scored mutation propensity. The propensity mutations in the common hypermutator tumours rarely occurred without the aberrant *POLE*. Thus, propensity mutations may occur independently of POLE deficiency in POLE category tumours. To bypass this possibility, the frequency of these were thus subtracted as background. Our analysis identified not only two known three-letter base motifs (TCT and TCG) that accumulate massive quantities of mutation but also highly-mutated pattern (TTT) in the POLE category. Furthermore, T > G substitution was increased in comparison with the common hypermutator tumours, over 40% of which had occurred in the core of the TTT motif. However, T > G substitution is known to be associated

with an imbalance of the dNTP pool<sup>33,34</sup>. Therefore, to determine the genes susceptible to POLE mutant status, mutation patterns of TCT > TAT and TCG > TTG except for TTT > TGT were utilised in the calculation of PS as the propensity mutation.

The propensity mutation patterns (TCT > TAT and TCG > TTG) were continually generated during defective POLE, implying inclusion of new (later-occurring) mutations. In these mutation patterns in 414 genes that were enriched in POLE-category, VAF score was decreased in comparison with other patterns. This finding may indicate that propensity mutation patterns include abundant mutations occurring subsequently.

We determined PS as a variation-susceptible indicator using the frequency of mutation pattern. This score was expected to represent a normal distribution when almost mutations randomly occurred in specific sequences (TCT and TCG) owing to defective POLE. However, the score of each gene that was enriched in the POLE category does not follow a normal distribution; additionally, IP was found on the Q-Q plot. This suggested that the subpopulation of low PS was less than the number of IP-enriched genes that had mutated independently of the POLE mutant. To clearly separate these genes, the indefinite period (CI of IP) was established as a border for buffering. Consequently, 414 genes that were enriched in the POLE category were classified into three periods. It is expected that the genes that mutated after the occurrence of the aberrant *POLE* would be concentrated in the POLE-dependent period ( $\geq$  upper-CI), whereas the POLE-independent ( $<$  lower-CI) period was dominated by genes that mutated prior to or contemporary with this event, although this period also contained a certain number of genes unsusceptible to defective proofreading by POLE. Our classification was thus considered a reflection of the sequential shift of mutation profile resulting from the propensity of the POLE mutant. Accordingly, the *POLE* gene itself was classified in the POLE-independent period. Notably, a high mutation rate per protein-length of *PTEN* was observed in the indefinite period and the POLE-independent period using our and ICGC data sets, respectively, inferring that this gene was preferentially mutated during the early generation of the *POLE* mutation. Other cancer-related genes including *TSG* were also enriched in the POLE-dependent period. Therefore, we concluded that, based on the two-hit hypothesis, the somatic mutation of *POLE* and *PTEN* served as a primary trigger for tumorigenesis in POLE-category tumours.

Recently, characterisation of tumours has focused on the relationship between driver mutation and mutational signature. Poulos *et al.* revealed that mutations of methylated CpG dinucleotides were accumulated as potential driver events in POLE-mutant colorectal cancer<sup>35</sup>. In a pre-print repository (bioRxiv), Temko *et al.* reported that mutational signatures were affected by specific mutations in *APOBEC* and DNA mismatch repair genes<sup>36</sup>. Tumour-specific mutation pattern is considered to occur under conditions of defective DNA repair/replication-related genes. In our analysis, biased mutations in POLE mutants were scored based on the characteristic of visualisation of mutation pattern transition, resulting in the identification of *PTEN* mutation as potential tumorigenesis event that coincided roughly with the occurrence of defective *POLE*. However, significant accumulation of APC mutations that was detected in previous studies<sup>35,36</sup> was not detected in our analysis. Quantification of mutation pattern in tumours with deficient DNA repair activity may constitute a helpful tool to estimate the procession of tumorigenesis.

*POLE* or *POLD1* mutation with the accompanying defect of DNA proofreading leads to reduction of fidelity in DNA replication, and a high level of somatic mutations in double-strand break repair (DSBR) genes (including HR and NHEJ) are also observed in these tissues<sup>37</sup>. However, it is unclear whether such mutants reduce activity of DSBR towards the excess accumulation of mutations. Our analysis revealed that HR repair activity based on gene expression was increased in POLE-category tumours despite the occurrence of *PTEN* mutations as an early event in tumorigenesis. As HR-related gene mutation is considered a compensatory event following *PTEN*-deficiency<sup>22</sup>, these findings imply that HR repair activity was sequentially shifted together with mutation accumulation after the occurrence of aberrant *PTEN*. Although POLE category tumours predominantly indicated the presence of Signature.10 based on deconstructSigs<sup>14</sup>, Signature.3 representing double-strand breaks was rare in these tumours (data not shown), suggesting that HR function may be gradually restored after occurrence of *PTEN* mutation. In particular, *PTEN* mutations initially compromise HR function, and then an increase of mutated DSBR genes that are generated by aberrant POLE may lead to a reversal of HR deficiency. To our knowledge, this study is the first report of a tumour exome profile that could be used to infer a footprint of somatic mutations.

Our classification showed that the sequential shift of mutation profile in genes had occurred in POLE-category tumours owing to the propensity of the POLE mutant effect. Similarly, to estimate the transition of mutation profile in pathways, this classification was applied to pathway analysis. *ASCC3* and *HIF1A* were categorised as POLE-independent, whereas the remaining POLE-category specific genes were susceptible to POLE mutant effects. Fms related tyrosine kinase 1 (*FLT1*) and kinase insert domain receptor (*KDR*), which are transcriptionally regulated by HIF1A, are known to function as VEGF receptors, affecting HIF1A activity through an autocrine loop<sup>38,39</sup>. In addition, ceruloplasmin (CP) stabilises HIF1A in a self-stimulating feedback loop<sup>40,41</sup>. Notably, HIF1A activity appears to be suppressed in POLE-category tumours. Activation of HIF1A under hypoxia leads to a decrease of HR function<sup>19,42</sup>; therefore, POLE-category tumours may compromise HIF1A activity via somatic mutations in itself and related genes to maintain the function of HR repair.

This study of the molecular and genomic mutational basis of hypermutator tumours harbouring POLE mutations allowed us to describe mutation footprints that could be classified according to POLE-dependent/-independent and indefinite periods. Notably, these periods illustrated the sequential shift of mutation that occurred via the defective proofreading of POLE in a clinical sample, identifying candidate genes as primary triggers of tumorigenesis or the suggested restoration of HR function resulting from mutation accumulation. The classification method developed through this study could be applied to tumours harbouring somatic mutations that were biased by defects of DNA repair.

In conclusion, the present study identified hypermutator tumours harbouring POLE mutation from among over 2,000 Japanese patients with cancer, and showed the similarity of observed molecular profile to that of



previous studies. Although the aberrant *POLE* led to a mutation bias in the tumour, variation-insusceptible genes were also identified. Based on these characteristics, the transition of mutation in a clinical sample was classified into three periods. Furthermore, we revealed that the somatic mutation of *POLE* and *PTEN* represented a primary trigger for tumourigenesis in tumours with defective *POLE* proofreading ability. This is, to our knowledge, the first description of the two-hit hypothesis estimated from the classification of sequential mutation shift using a large-scale genomic study. A tracing of somatic mutation propensity may therefore provide new insight into the understanding of mutation-driven tumourigenesis.

## Methods

**Patients and specimens.** Informed consent was obtained from all patients and the Institutional Review Board of Shizuoka Cancer Center approved all aspects of this study (Authorization Number: 25–33). All experiments using clinical samples were performed in accordance with the approved guidelines.

Each tumour and its surrounding normal tissue ( $\geq 0.1$  g) were dissected from surgical specimens immediately after resection of the lesion. The tumour sample was visually assessed by a clinical pathologist in our hospital when tumour content was  $\geq 50\%$ . In addition, peripheral blood was collected as a control for WES.

**DNA and RNA isolation.** For DNA analysis, tumour and normal tissues were immediately frozen in liquid nitrogen prior to DNA extraction. DNA was extracted from tissue and peripheral blood samples using a QIAamp DNA blood Mini Kit (Qiagen, Venlo, The Netherlands). Purified DNA was quantified using a NanoDrop and Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA).

For RNA analysis, samples were immersed in RNeasy lysis solution (Thermo Fisher Scientific), then stored overnight at 4 °C prior to RNA extraction. Total RNA was extracted from approximately 10 mg hashed tissue sample using the miRNeasy Mini Kit (Qiagen) according to manufacturer's instruction. Initially, QIAzol reagent was applied to the tissue sample and then shaken with a 5-mm zirconia bead using a TissueLyser II (Qiagen) for 10 min at room temperature. The extracted RNA sample was quantified using a NanoDrop and its quality was checked using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). In our analysis, samples with RNA integrity number (RIN)<sup>43</sup>  $< 6.0$  were discarded.

**Next-generation sequencing.** The exome library for WES was constructed using an Ion Torrent AmpliSeq RDY Exome Kit (Thermo Fisher Scientific) in accordance with manufacturer recommended protocol<sup>44,45</sup>. For the construction, 100 ng DNA was used in the target amplification under the following conditions: 99 °C for 2 min, followed by 10 cycles at 95 °C for 15 s and 60 °C for 16 min, and a final hold at 10 °C. The amplicons were ligated with Ion Torrent Proton adapters (Thermo Fisher Scientific) at 22 °C for 30 min followed by 72 °C for 10 min, and the library was purified using Agencourt Ampure XT beads (Beckman Coulter). This exome library supplied 292,903 amplicons covering 57.7 Mb of the human genome comprising 34.8 Mb exons of 18,835 genes registered in RefSeq. The constructed library was quantified using quantitative PCR, and DNA was sequenced using a semiconductor DNA sequencer (Ion Torrent Proton Sequencer, Thermo Fisher Scientific) according to manufacturer's instruction.

**Sequencing workflow for identification of somatic mutations.** The semiconductor DNA sequencer-derived binary raw data were converted, using Torrent Suite software (ver. 4.4, Thermo Fisher Scientific), into sequence reads that were mapped to the reference human genome (UCSC hg19). At this step, sequence data derived from tumour and blood samples were individually analysed, and mapping results were saved as BAM files. Two BAM files were uploaded to the Ion Reporter system and analysed concurrently using AmpliSeq exome tumour-normal pair workflow (ver. 4.4, Thermo Fisher Scientific) with a Custom Hotspot file that specifies somatic and pathogenic mutations registered in COSMIC and ClinVar, respectively. The list of identified mutations was processed by in-house scripts to remove false positive calls including sequencer-derived error. Mutations fulfilling at least one of the following criteria were discarded as false positive: (1) quality score  $< 60$ , (2) depth of coverage  $< 20$ , (3) variant read observed in one strand only, (4) clipped sequence length  $< 100$  (avg\_clipped\_length  $< 100$ ), (5) variant located on either sequence end (avg\_pos\_as\_fraction  $< 0.05$ ), or (6) mutation matches to in-house false positive list. Parameters specified in criteria (4) and (5) were calculated by bam-readcount with option “-q 1” (ver. 0.8.0) (<https://github.com/genome/bam-readcount>). Our analysis focused on nonsynonymous SNVs located in an exon or splice site, and the mutation frequency on genome was then evaluated using 34.8 Mb exons. Arbitrary somatic mutations were manually inspected using the Integrative Genomics Viewer<sup>46</sup> to avoid sequencer and amplicon-derived errors. Briefly, frequently mutated sites and variations that were barely detected in both tumour and blood were extracted from sequencing data. These somatic mutation candidates containing multiple nucleotide variation ( $\sim 1,000$  sites) were validated by Sanger sequencing to exclude false positive variation. The mutations that were not detected in Sanger sequencing were included in the in-house false positive list. To predict the effects of mutations, SnpEff<sup>47</sup> and RefSeq were adopted as the source of curated and annotated sequences. Somatic mutations on the exome were annotated using the databases as follows: COSMIC<sup>48</sup>, ClinVar<sup>49</sup>, dbSNP<sup>50</sup>, UniProt<sup>51</sup>, and DrugBank<sup>52</sup>.

**Microarray analysis.** Purified total RNA for GEP was amplified and fluorescent-labelled using a One-Color Low Input Quick Amp Labeling Kit (Agilent Technologies) according to manufacturer's instruction. Hybridisation and scanning were performed as in a previous report<sup>53,54</sup>. Cy3-labelled cRNAs were hybridised to a SurePrint G3 Human Gene Expression 8 × 60 K v2 Microarray (Agilent Technologies), which has 50,599 probes capable of detecting 29,833 genes registered in the Entrez Gene Database. Fluorescent signal after hybridisation and washing was scanned using a DNA Microarray Scanner (Agilent Technologies), and then assessed by Agilent Feature Extraction software.

**Gene expression profiling (GEP).** To establish a correspondence between somatic mutation and gene expression, probes on the microarray were selected according to the reference human genome (UCSC hg19). Raw signal intensity derived from the scanned image was filtered by Agilent Flag Values to maintain reliability of microarray data, and then log-transformed and normalised to the 75th percentile. To compare gene expression between samples, z-score of the target gene was calculated from fold change (tumour vs normal in the same patient). In cases where tumour type created difficulty with respect to dissection of the normal sample, this expression profile was excluded from the analysis. These data were prepared and output using GeneSpring GX software (Agilent Technologies) and a Subio Platform. HR and NHEJ signatures were composed of multiple genes that showed up-regulated activation of these functions in a previous report<sup>19</sup>, and the signature score was calculated based on gene signature analysis using average of gene expression<sup>28</sup>. Microarray analysis was performed in accordance with MIAME guidelines<sup>55</sup>.

**Calculation of PS for classification.** To discriminate whether a somatically mutated gene was susceptible to POLE mutant effects, a PS was employed as a variation-susceptible indicator. The PS was composed of frequencies of specific mutation patterns in POLE-category and common hypermutator tumours (classification described in Fig. 1b), and was calculated using the following formula:

$$PS = FMP_{POLE \text{ category}} - FMP_{Common \text{ hypermutator}} \quad (1)$$

where FMP is the frequency of a mutation pattern in an individual gene. Then,

$$FMP(\%) = propensity \text{ mutation} / total \text{ mutation} \times 100, \quad (2)$$

where propensity mutation is the sum of TCT > TAT and TCG > TTG substitutions in a gene sequence. The above propensity mutation was exclusive of TTT > TGT. Total mutation is the number of all substitutions in each group (POLE category or common hypermutator).

An IP was determined using the Pearson correlation coefficient in the Q-Q plot of PS, and CI of IP (see Statistical analysis) was set as an indefinite region to clearly discriminate between POLE mutant-susceptible and unsusceptible genes.

**Pathway analysis.** To estimate characteristic pathways, knowledge database analysis was performed based on sequential somatic mutation shift. Here, the transition of specific mutation pattern derived by defective protein (such as mutated POLE) was represented as sequential (somatic) mutation shift. To visualise the above shift, mutated genes in POLE category and common hypermutator were classified to three groups using PS (see Results section). For pathway analysis, our WES data and public data set of ICGC (<https://dcc.icgc.org/>) were prepared. Colon adenocarcinoma and uterine endometrium carcinoma data set obtained from ICGC (COAD-US and UCEC-US) were grouped into the POLE category and common hypermutator. POLE-category-specific mutations were then extracted, and their PS was calculated from this data set. The pathway analyses were performed using KeyMolnet software (ver.6.0.16.329, purchased from KM Data, <http://www.immd.co.jp/en/index.html>), which contains a stand-alone knowledge database. Upon interrelationship search of mutated genes in POLE category, the database was used with the following parameters: network termination, molecules; maximum number of path, 1; source, core and secondary; usage, positive/negative expression and direct relation. These analyses using knowledge database were scored by an internal index (H score), the elevated value of which indicated a high-relationship pathway.

**Statistical analysis.** A significance of association of the mutations between the two groups was analysed using Fisher's exact test. Microarray-derived gene expression data were normalised, and a significant difference in expression including the signature of gene set was calculated by Welch's *t*-test. For mutation frequency and tumour cellularity as determined using PurBayes<sup>56</sup>, the assumptions of normality and the equality of two variances were tested by the Shapiro-Wilk test and *F*-test, respectively. Welch's *t*-test was performed in the assumed normal distribution. Against comparison of samples assuming non-normal distribution, a Mann-Whitney-Wilcoxon or Brunner-Munzel test (also known as the generalised Wilcoxon test) was performed depending on the assumption of the *F*-test. To identify genes harbouring POLE-category-specific mutations, a Fisher's exact test and the Benjamini-Hochberg procedure ( $q < 0.05$ ) were carried out. Here, *P*-values < 0.05 were considered as significant. The break-point of segmented regression was estimated by Davies' test<sup>57,58</sup> (using R package 'segmented'). In segmented regression, bootstrapping was employed to construct 99.999% CI, which was designated along with its underestimation by this method.

**Data availability.** The WES data referenced during the study are available in a public repository that is accessible through the COSMIC and ICGC (<https://dcc.icgc.org/>) websites. The identifiers of individual samples are listed in Supplementary Table S2. The gene expression data including the signatures are described in Supplementary Table S3. The authors declare that all the other data supporting the findings of this study are available within the article and its supplementary information files and from the corresponding author upon reasonable request. The somatic mutation data of POLE-mutated samples and hypermutators from exome sequencing are available in the National Bioscience Database Center (NBDC) and Japanese Genotype-phenotype Archive (JGA) databases under the accession number hum0127 and JGAS0000000130, respectively.

## References

- Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421, <https://doi.org/10.1038/nature12477> (2013).
- Cancer Genome Atlas Research, N. *et al.* Integrated genomic characterization of endometrial carcinoma. *Nature* **497**, 67–73, <https://doi.org/10.1038/nature12113> (2013).
- Cancer Genome Atlas, N. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337, <https://doi.org/10.1038/nature11252> (2012).

4. Shinbrot, E. *et al.* Exonuclease mutations in DNA polymerase epsilon reveal replication strand specific mutation patterns and human origins of replication. *Genome Res* **24**, 1740–1750, <https://doi.org/10.1101/gr.174789.114> (2014).
5. Pursell, Z. F., Isoz, I., Lundstrom, E. B., Johansson, E. & Kunkel, T. A. Yeast DNA polymerase epsilon participates in leading-strand DNA replication. *Science* **317**, 127–130, <https://doi.org/10.1126/science.1144067> (2007).
6. Nick McElhinny, S. A., Gordenin, D. A., Stith, C. M., Burgers, P. M. & Kunkel, T. A. Division of labor at the eukaryotic replication fork. *Mol Cell* **30**, 137–144, <https://doi.org/10.1016/j.molcel.2008.02.022> (2008).
7. Larrea, A. A. *et al.* Genome-wide model for the normal eukaryotic DNA replication fork. *Proc Natl Acad Sci USA* **107**, 17674–17679, <https://doi.org/10.1073/pnas.1010178107> (2010).
8. Church, D. N. *et al.* DNA polymerase epsilon and delta exonuclease domain mutations in endometrial cancer. *Hum Mol Genet* **22**, 2820–2828, <https://doi.org/10.1093/hmg/ddt131> (2013).
9. Kane, D. P. & Shcherbakova, P. V. A common cancer-associated DNA polymerase epsilon mutation causes an exceptionally strong mutator phenotype, indicating fidelity defects distinct from loss of proofreading. *Cancer Res* **74**, 1895–1901, <https://doi.org/10.1158/0008-5472.CAN-13-2892> (2014).
10. Palles, C. *et al.* Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nat Genet* **45**, 136–144, <https://doi.org/10.1038/ng.2503> (2013).
11. Vogelstein, B. *et al.* Genetic alterations during colorectal-tumor development. *N Engl J Med* **319**, 525–532, <https://doi.org/10.1056/NEJM198809013190901> (1988).
12. Knudson, A. G. Jr. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci USA* **68**, 820–823 (1971).
13. Drost, J. *et al.* Sequential cancer mutations in cultured human intestinal stem cells. *Nature* **521**, 43–47, <https://doi.org/10.1038/nature14415> (2015).
14. Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B. S. & Swanton, C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol* **17**, 31, <https://doi.org/10.1186/s13059-016-0893-4> (2016).
15. Vogelstein, B. *et al.* Cancer genome landscapes. *Science* **339**, 1546–1558, <https://doi.org/10.1126/science.1235122> (2013).
16. Shen, W. H. *et al.* Essential role for nuclear PTEN in maintaining chromosomal integrity. *Cell* **128**, 157–170, <https://doi.org/10.1016/j.cell.2006.11.042> (2007).
17. Mendes-Pereira, A. M. *et al.* Synthetic lethal targeting of PTEN mutant cells with PARP inhibitors. *EMBO Mol Med* **1**, 315–322, <https://doi.org/10.1002/emmm.200900041> (2009).
18. Dedes, K. J. *et al.* PTEN deficiency in endometrioid endometrial adenocarcinomas predicts sensitivity to PARP inhibitors. *Sci Transl Med* **2**, 53ra75, <https://doi.org/10.1126/scitranslmed.3001538> (2010).
19. Meng, A. X. *et al.* Hypoxia down-regulates DNA double strand break repair gene expression in prostate cancer cells. *Radiother Oncol* **76**, 168–176, <https://doi.org/10.1016/j.radonc.2005.06.025> (2005).
20. Pastwa, E. & Blasiak, J. Non-homologous DNA end joining. *Acta Biochim Pol* **50**, 891–908, 035004891 (2003).
21. Daley, J. M., Kwon, Y., Niu, H. & Sung, P. Investigations of homologous recombination pathways and their regulation. *Yale J Biol Med* **86**, 453–461 (2013).
22. Peng, G. *et al.* Genome-wide transcriptome profiling of homologous recombination DNA repair. *Nat Commun* **5**, 3361, <https://doi.org/10.1038/ncomms4361> (2014).
23. Pleasance, E. D. *et al.* A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* **463**, 184–190, <https://doi.org/10.1038/nature08629> (2010).
24. Domingo, E. *et al.* Somatic POLE proofreading domain mutation, immune response, and prognosis in colorectal cancer: a retrospective, pooled biomarker study. *Lancet Gastroenterol Hepatol* **1**, 207–216, [https://doi.org/10.1016/S2468-1253\(16\)30014-0](https://doi.org/10.1016/S2468-1253(16)30014-0) (2016).
25. Shlien, A. *et al.* Combined hereditary and somatic mutations of replication error repair genes result in rapid onset of ultra-hypermuted cancers. *Nat Genet* **47**, 257–262, <https://doi.org/10.1038/ng.3202> (2015).
26. Kane, M. F. *et al.* Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines. *Cancer Res* **57**, 808–811 (1997).
27. Veigl, M. L. *et al.* Biallelic inactivation of hMLH1 by epigenetic gene silencing, a novel mechanism causing human MSI cancers. *Proc Natl Acad Sci USA* **95**, 8698–8702 (1998).
28. Cristescu, R. *et al.* Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. *Nat Med* **21**, 449–456, <https://doi.org/10.1038/nm.3850> (2015).
29. Funkhouser, W. K. *et al.* Relevance, pathogenesis, and testing algorithm for mismatch repair-defective colorectal carcinomas: a report of the association for molecular pathology. *J Mol Diagn* **14**, 91–103, <https://doi.org/10.1016/j.jmoldx.2011.11.001> (2012).
30. Ahn, S. M. *et al.* The somatic POLE P286R mutation defines a unique subclass of colorectal cancer featuring hypermutation, representing a potential genomic biomarker for immunotherapy. *Oncotarget* **7**, 68638–68649, <https://doi.org/10.18632/oncotarget.11862> (2016).
31. Mehnert, J. M. *et al.* Immune activation and response to pembrolizumab in POLE-mutant endometrial cancer. *J Clin Invest* **126**, 2334–2340, <https://doi.org/10.1172/JCI84940> (2016).
32. Dudley, J. C., Lin, M. T., Le, D. T. & Eshleman, J. R. Microsatellite Instability as a Biomarker for PD-1 Blockade. *Clin Cancer Res* **22**, 813–820, <https://doi.org/10.1158/1078-0432.CCR-15-1678> (2016).
33. Meuth, M. The molecular basis of mutations induced by deoxyribonucleoside triphosphate pool imbalances in mammalian cells. *Exp Cell Res* **181**, 305–316 (1989).
34. Kunz, B. A. *et al.* International Commission for Protection Against Environmental Mutagens and Carcinogens. Deoxyribonucleoside triphosphate levels: a critical factor in the maintenance of genetic stability. *Mutat Res* **318**, 1–64 (1994).
35. Poulos, R. C., Olivier, J. & Wong, J. W. H. The interaction between cytosine methylation and processes of DNA replication and repair shape the mutational landscape of cancer genomes. *Nucleic Acids Res* **45**, 7786–7795, <https://doi.org/10.1093/nar/gkx463> (2017).
36. Daniel, T., Tomlinson, I., Severini, S., Schuster-Boeckler, B. & Graham, T. The effects of mutational processes and selection on driver mutations across cancer types. *bioRxiv*, <https://doi.org/10.1101/149096> (2017).
37. Chae, Y. K. *et al.* Genomic landscape of DNA repair genes in cancer. *Oncotarget* **7**, 23312–23321, <https://doi.org/10.18632/oncotarget.8196> (2016).
38. Das, B. *et al.* A hypoxia-driven vascular endothelial growth factor/Flt1 autocrine loop interacts with hypoxia-inducible factor-1alpha through mitogen-activated protein kinase/extracellular signal-regulated kinase 1/2 pathway in neuroblastoma. *Cancer Res* **65**, 7267–7275, <https://doi.org/10.1158/0008-5472.CAN-04-4575> (2005).
39. Giatromanolaki, A. *et al.* Phosphorylated KDR expression in endometrial cancer cells relates to HIF1alpha/VEGF pathway and unfavourable prognosis. *Mod Pathol* **19**, 701–707, <https://doi.org/10.1038/modpathol.3800579> (2006).
40. Martin, F. *et al.* Copper-dependent activation of hypoxia-inducible factor (HIF)-1: implications for ceruloplasmin regulation. *Blood* **105**, 4613–4619, <https://doi.org/10.1182/blood-2004-10-3980> (2005).
41. Mukhopadhyay, C. K., Mazumder, B. & Fox, P. L. Role of hypoxia-inducible factor-1 in transcriptional activation of ceruloplasmin by iron deficiency. *J Biol Chem* **275**, 21048–21054, <https://doi.org/10.1074/jbc.M000636200> (2000).
42. Bindra, R. S. *et al.* Down-regulation of Rad51 and decreased homologous recombination in hypoxic cancer cells. *Mol Cell Biol* **24**, 8504–8518, <https://doi.org/10.1128/MCB.24.19.8504-8518.2004> (2004).

43. Schroeder, A. *et al.* The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol Biol* **7**, 3, <https://doi.org/10.1186/1471-2199-7-3> (2006).
44. Shimoda, Y. *et al.* Integrated next-generation sequencing analysis of whole exome and 409 cancer-related genes. *Biomed Res* **37**, 367–379, <https://doi.org/10.2220/biomedres.37.367> (2016).
45. Nagashima, T. *et al.* Optimizing an ion semiconductor sequencing data analysis method to identify somatic mutations in the genomes of cancer cells in clinical tissue samples. *Biomed Res* **37**, 359–366, <https://doi.org/10.2220/biomedres.37.359> (2016).
46. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24–26, <https://doi.org/10.1038/nbt.1754> (2011).
47. 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)* **6**, 80–92, <https://doi.org/10.4161/fly.19695> (2012).
48. Forbes, S. A. *et al.* COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res* **43**, D805–811, <https://doi.org/10.1093/nar/gku1075> (2015).
49. Landrum, M. J. *et al.* ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* **42**, D980–985, <https://doi.org/10.1093/nar/gkt1113> (2014).
50. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* **29**, 308–311 (2001).
51. UniProt, C. UniProt: a hub for protein information. *Nucleic Acids Res* **43**, D204–212, <https://doi.org/10.1093/nar/gku989> (2015).
52. Wishart, D. S. *et al.* DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res* **34**, D668–672, <https://doi.org/10.1093/nar/gkj067> (2006).
53. Hatakeyama, K. *et al.* Identification of a novel protein isoform derived from cancer-related splicing variants using combined analysis of transcriptome and proteome. *Proteomics* **11**, 2275–2282, <https://doi.org/10.1002/pmic.201100016> (2011).
54. Ohshima, K. *et al.* Integrated analysis of gene expression and copy number identified potential cancer driver genes with amplification-dependent overexpression in 1,454 solid tumors. *Sci Rep* **7**, 641, <https://doi.org/10.1038/s41598-017-00219-3> (2017).
55. Brazma, A. *et al.* Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet* **29**, 365–371, <https://doi.org/10.1038/ng1201-365> (2001).
56. Larson, N. B. & Fridley, B. L. PurBayes: estimating tumor cellularity and subclonality in next-generation sequencing data. *Bioinformatics* **29**, 1888–1889, <https://doi.org/10.1093/bioinformatics/btt293> (2013).
57. Davies, R. B. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **74**, 33–43, <https://doi.org/10.1093/biomet/74.1.33> (1987).
58. Davies, R. B. Hypothesis testing when a nuisance parameter is present only under the alternative: Linear model case. *Biometrika* **89**, 484–489, <https://doi.org/10.1093/biomet/89.2.484> (2002).

## Acknowledgements

We thank the members of the Shizuoka Cancer Center Hospital and Research Institute for their support and suggestions. This work was supported by the Shizuoka Prefectural Government, Japan.

## Author Contributions

K.Y. and K.H. designed the study. K.O., T.N., Y.S., K.U., and K.H. performed the experiments. K.H. analysed the data. T.N. and K.H. performed statistical analyses. K.H. wrote the paper, with contributions from Sh.O., Su.O., M.S., Y.S., K.M., Y.A., M.K., and T.M.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-26967-4>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018