

## Research Article

# Abnormal Expressions of DNA Glycosylase Genes NEIL1, NEIL2, and NEIL3 Are Associated with Somatic Mutation Loads in Human Cancer

Kazuya Shinmura,<sup>1</sup> Hisami Kato,<sup>1</sup> Yuichi Kawanishi,<sup>2</sup>  
Hisaki Igarashi,<sup>1</sup> Masanori Goto,<sup>3</sup> Hong Tao,<sup>1</sup> Yusuke Inoue,<sup>1</sup> Satoki Nakamura,<sup>1</sup>  
Kiyoshi Misawa,<sup>4</sup> Hiroyuki Mineta,<sup>4</sup> and Haruhiko Sugimura<sup>1</sup>

<sup>1</sup>Department of Tumor Pathology, Hamamatsu University School of Medicine, Hamamatsu 431-3192, Japan

<sup>2</sup>Research Equipment Center, Hamamatsu University School of Medicine, Hamamatsu 431-3192, Japan

<sup>3</sup>Division of Carcinogenesis and Prevention, National Cancer Center Research Institute, Tokyo 104-0045, Japan

<sup>4</sup>Department of Otolaryngology/Head and Neck Surgery, Hamamatsu University School of Medicine, Hamamatsu 431-3192, Japan

Correspondence should be addressed to Kazuya Shinmura; [kzshinmu@hama-med.ac.jp](mailto:kzshinmu@hama-med.ac.jp)

Received 15 November 2015; Revised 23 January 2016; Accepted 31 January 2016

Academic Editor: Kum Kum Khanna

Copyright © 2016 Kazuya Shinmura et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The effects of abnormalities in the DNA glycosylases NEIL1, NEIL2, and NEIL3 on human cancer have not been fully elucidated. In this paper, we found that the median somatic total mutation loads and the median somatic single nucleotide mutation loads exhibited significant inverse correlations with the median NEIL1 and NEIL2 expression levels and a significant positive correlation with the median NEIL3 expression level using data for 13 cancer types from the Cancer Genome Atlas (TCGA) database. A subset of the cancer types exhibited reduced NEIL1 and NEIL2 expressions and elevated NEIL3 expression, and such abnormal expressions of NEIL1, NEIL2, and NEIL3 were also significantly associated with the mutation loads in cancer. As a mechanism underlying the reduced expression of NEIL1 in cancer, the epigenetic silencing of *NEIL1* through promoter hypermethylation was found. Finally, we investigated the reason why an elevated NEIL3 expression level was associated with an increased number of somatic mutations in cancer and found that NEIL3 expression was positively correlated with the expression of APOBEC3B, a potent inducer of mutations, in diverse cancers. These results suggested that the abnormal expressions of NEIL1, NEIL2, and NEIL3 are involved in cancer through their association with the somatic mutation load.

## 1. Introduction

*NEIL1* (ENSG00000140398, OMIM #608844), *NEIL2* (ENSG00000154328, OMIM #608933), and *NEIL3* (ENSG00000109674, OMIM #608934) are structural human homologues of *Escherichia coli* (*E. coli*) *Nei* and *Fpg*, the genes encoding a DNA glycosylase that initiates the base excision repair (BER) process. These three homologues also have actual functional activities as DNA glycosylases [1–5], although the modes of strand incision differ; NEIL1 and NEIL2 have strong  $\beta$ ,  $\delta$  elimination activities, but NEIL3 has only a weak  $\beta$  elimination activity [6, 7]. Regarding substrate specificity, the three DNA glycosylases have broad and overlapping specificities for modified bases. The preferred

substrates for all of them are spiroiminodihydantoin and guanidinohydantoin, which are highly mutagenic DNA lesions, but various other DNA lesions are also recognized by some of them. For example, 8-hydroxyguanine, which is also a mutagenic base lesion, is a substrate for NEIL1 and NEIL2, but not for NEIL3 [7, 8]. Because of these DNA glycosylase activities, NEIL1, NEIL2, and NEIL3 also have the ability to regulate the mutation frequency in cells. Deficiencies of NEIL1 or NEIL2 in mammalian cells reportedly lead to an elevated mutation frequency [9–11], and the overproduction of mouse NEIL3 in an *E. coli fpg nei mutY* strain reduced the spontaneous mutation frequency [12]. These findings indicate that NEIL1, NEIL2, and NEIL3 have the ability to suppress mutations in cells. Therefore, NEIL1, NEIL2, and

NEIL3 are important enzymes to maintain the stability of genomic DNA by preventing mutations.

Recent advances in high-throughput sequencing technology have enabled associations between specific gene abnormalities and the somatic mutation load to be investigated in human cancer. Such investigations have revealed that the inactivation of mismatch repair genes, inactivating mutations of *BRCA1* (ENSG0000012048), *BRCA2* (ENSG00000139618), *POLE* (ENSG00000177084), and *POLK* (ENSG00000122008), or the overexpression of *APOBEC3B* (ENSG00000179750) causes mutagenesis in cancer [13–17]. At present, however, a definitive relationship between the status of DNA glycosylases, including NEIL1, NEIL2, and NEIL3, and the extent of somatic mutations in genomic DNA has not been demonstrated. Since NEIL1, NEIL2, and NEIL3 are involved in the repair of mutagenic bases and are capable of suppressing mutations, we investigated the relationship between the expression levels of NEIL1, NEIL2, and NEIL3 and the somatic mutation load using whole-exome sequencing data derived from the Cancer Genome Atlas (TCGA) database. We found, for the first time, that the abnormal expressions of NEIL1, NEIL2, and NEIL3 are associated with somatic mutation loads in diverse cancers.

## 2. Materials and Methods

**2.1. Collection of Publicly Available Data on Somatic Mutations, mRNA Expression, and DNA Methylation.** mRNA expression, somatic mutation, and the DNA methylation data of 13 cancer types [bladder urothelial carcinoma (TCGA ID: BLCA), breast invasive carcinoma (BRCA), colon adenocarcinoma (COAD), head and neck squamous cell carcinoma (HNSC), kidney chromophobe renal cell carcinoma (KICH), kidney renal clear cell carcinoma (KIRC), kidney renal papillary cell carcinoma (KIRP), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), prostate adenocarcinoma (PRAD), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), and thyroid carcinoma (THCA)] were collected from the TCGA data portal (<https://tcga-data.nci.nih.gov/tcga/>) in April 2014. The number of cases used in this study is summarized in Supplementary Table S1 in Supplementary Material available online at <http://dx.doi.org/10.1155/2016/1546392>. The expression data were obtained as processed RNA-sequence (RNA-seq) data in the form of RNA-seq by Expectation Maximization (RSEM) [18], excluding the RNA-seq data of STAD, which was obtained in the form of Reads Per Kilobase of Exon Model (RPKM) per million mapped reads [19]. The somatic mutation data were obtained using whole-exome sequencing and are shown in the form of a mutation annotation format (MAF) file. The DNA methylation data obtained using the HumanMethylation450 platform (Illumina Inc., CA, USA) were shown as the  $\beta$  value (ratio of the methylated probe intensity and the overall intensity). Whether the expressions of NEIL1, NEIL2, and NEIL3 are epigenetically silenced by promoter hypermethylation was determined based on the following 4 criteria, according to a previous report [13] with some modifications: (1) a mean DNA methylation  $\beta$  value

at the CpG site near the transcription start site in normal tissue  $< 0.4$ ; (2) a difference in the  $\beta$  value between the 90th percentile of  $\beta$  value in tumor tissue and the mean in normal tissue  $> 0.1$ ; (3) a fold expression change between the mean in normal tissue and the mean of the 10% of tumor tissue with the highest  $\beta$  value  $> 1.5$ ; (4) a Spearman rank correlation value between DNA methylation and gene expression  $< -0.25$ .

**2.2. 5-Aza-Deoxycytidine (5-aza-dC) Treatment.** The gastric cancer cell lines MKN45 and MKN74, which were obtained from the Human Science Research Resource Bank (Osaka, Japan), were treated with  $2 \mu\text{M}$  of 5-aza-dC (Sigma-Aldrich, St. Louis, MO, USA) for 48 h, as described previously [20].

**2.3. Quantitative Reverse-Transcription- (QRT-) Polymerase Chain Reaction (PCR).** Total RNA was extracted using an RNeasy Plus Mini Kit (Qiagen, Valencia, CA, USA) and was converted to cDNA using a SuperScript First-Strand Synthesis System for RT-PCR (Invitrogen, Carlsbad, CA, USA). Real-time QRT-PCR was performed using cDNA, a set of primers, a QuantiTect SYBR Green PCR kit (Qiagen), and a LightCycler instrument (Roche, Palo Alto, CA, USA). The PCR primers were as follows: 5'-AAG TCA GGT TCT TCC GCC AC-3' and 5'-CGG TAG GCA CTG CTC TCA AAG-3' for the NEIL1 transcript (transcript variant 2: NM\_024608), 5'-GCA GAA TAA CTG TGT GCC GCT-3' and 5'-ACC CTG CTA GAT GTC CAA CTG ATT-3' for the NEIL3 transcript, and 5'-GCT CAG ACA CCA TGG GGA AG-3' and 5'-TGT AGT TGA GGT CAA TGA AGG GG-3' for the GAPDH (ENSG0000011640) transcript. The relative amounts of NEIL1 or NEIL3 transcript were normalized to the amount of the GAPDH transcript.

**2.4. Immunohistochemical Analysis.** Paraffin embedded blocks of head and neck squamous cell carcinoma (HNSCC) cancer tissue and corresponding normal tissue from a total of 77 sporadic cases of primary HNSCC were obtained from Hamamatsu University Hospital (Japan). The mean age of the patients was 67.2 years (standard deviation: 9.4 years), and the sample included 69 men and 8 women. The sections were boiled at  $96^\circ\text{C}$  for 40 min in TE solution (pH 9.0) for antigen retrieval and incubated for 5 min in a 3% hydrogen peroxide solution to block endogenous peroxidase activity. Then, the sections were incubated with an anti-NEIL1 polyclonal antibody (Sigma-Aldrich) followed by an amino acid polymer conjugated with goat anti-rabbit IgG and HRP (Histofine Simple Stain MAX PO, Nichirei, Tokyo, Japan). The antigen-antibody complex was visualized with 3,3'-diaminobenzidine tetrahydrochloride and was counterstained with hematoxylin. The intensity values of the tumor cells were determined using a 3-point scale according to the color of the cells after NEIL1 immunostaining: 0: blue; 1: light brown; 2: brown. The percentage of cells with each intensity value was then multiplied by the intensity value, to obtain an immunohistochemical score of 0–200. The use of HNSCC tissues was approved by the Institutional Review Board of Hamamatsu University School of Medicine.

**2.5. Statistical Analysis.** The statistical analysis was performed using a Mann-Whitney  $U$  test, Spearman rank correlation test, or Wilcoxon matched pairs test. Overall survival curves were constructed using the Kaplan-Meier method, and the differences in the curves were evaluated using the log-rank test. The hazard ratio (HR) and its 95% confidence interval (CI) were calculated using the Cox proportional hazard model in both univariate and multivariate analyses. JMP version 9.0 software (SAS Institute, Cary, NC, USA) was used for all the statistical analyses.  $P$  values less than 0.05 were considered statistically significant.

### 3. Results

**3.1. Correlations between the Expression Levels of NEIL1, NEIL2, and NEIL3 and the Extent of Somatic Mutation in Human Cancer.** To determine whether the cancer mutation load is correlated with the expression levels of NEIL1, NEIL2, and NEIL3 in human cancer, mRNA expression data and somatic mutation data for 13 cancer types were obtained from the TCGA database. Regarding the somatic mutation data, along with the total mutation loads, SNP-type mutations, corresponding to single nucleotide exchange including synonymous and nonsynonymous mutations and not including insertion-type and deletion-type mutations, were also calculated to investigate the effects of NEIL1, NEIL2, and NEIL3 on such mutation types. Then, the median mutation loads for each cancer type and the median NEIL1, NEIL2, and NEIL3 expression values normalized to the expression value of the constitutive housekeeping gene YWHAZ (ENSG00000164924) [21] were analyzed to identify correlations. As expected, the median total mutation load and the median SNP-type mutation load showed a strong inverse correlation with the median NEIL1 expression level ( $\rho = -0.6382$ ,  $P = 0.0189$  and  $\rho = -0.6429$ ,  $P = 0.0178$ , resp.) (Figure 1(a)). In addition, the median total mutation load and the median SNP-type mutation load showed a strong inverse correlation with the median NEIL2 expression level ( $\rho = -0.6713$ ,  $P = 0.0120$  and  $\rho = -0.6758$ ,  $P = 0.0112$ , resp.). On the other hand, the median total mutation load and the median SNP-type mutation load showed a strong positive correlation with the median NEIL3 expression level ( $\rho = 0.6630$ ,  $P = 0.0135$  and  $\rho = 0.6593$ ,  $P = 0.0142$ , resp.). Similar to above, a significant correlation was also observed when another housekeeping gene, PSMB2 (ENSG00000126067) [22, 23], was used (Supplementary Figure S1); YWHAZ and PSMB2 were used because their expression levels were usually correlated with the expressions of several other housekeeping genes in various organ tissues (Supplementary Figure S2). These results suggest that the expression levels of NEIL1, NEIL2, and NEIL3 are differentially correlated with the extent of somatic mutation in human cancer.

**3.2. Expression Statuses of NEIL1, NEIL2, and NEIL3 and Their Associations with the Extent of Somatic Mutation in Each Cancer Type.** Next, we attempted to investigate the expression statuses of NEIL1, NEIL2, and NEIL3 in each cancer type and to determine whether their abnormal expressions

were associated with the mutation load of each cancer. The levels of NEIL1 and NEIL2 mRNA expression in tumor tissue, compared with normal tissue, were significantly reduced in 6 of the 13 (46.2%) cancer types and 4 of the 13 (30.8%) cancer types, respectively (Supplementary Table S2, Supplementary Figure S3). On the other hand, the level of NEIL3 mRNA expression was significantly increased in tumor tissue, compared with normal tissue, in all 13 cancer types (100%). When the 0.5-fold, 0.5-fold, and 2.5-fold values of the median expression value in noncancerous tissue samples of each organ were used as cut-off values to dichotomize the NEIL1, NEIL2, and NEIL3 expression values in the cancer cases, respectively, cancers with reduced NEIL1 expression, reduced NEIL2 expression, and elevated NEIL3 expression were detected in 31.4%, 9.0%, and 79.4% of all cancers, respectively (Supplementary Table S3). These results suggested that a subset of human cancers exhibited reduced NEIL1 and NEIL2 expressions and an elevated NEIL3 expression.

We next investigated whether the abnormal expressions of NEIL1, NEIL2, and NEIL3 were associated with the mutation load in each cancer type. The total mutation loads were significantly higher in the group of cancers with the lower NEIL1 and NEIL2 expression levels in 4 of the 13 (30.8%) cancer types and 2 of the 13 (15.4%) cancer types, respectively (Figure 1(b), Supplementary Figure S4, and Supplementary Tables S4 and S5). In addition, the total mutation loads were significantly higher in the group of cancers with the higher NEIL3 expression levels in 7 of the 13 (53.8%) cancer types (Table 1, Figure 1(b), and Supplementary Figure S4). These results suggested that the abnormal expressions of NEIL1, NEIL2, and NEIL3 are associated with the mutation load in cancer.

**3.3. Epigenetic Silencing of NEIL1 Expression in Human Cancer.** To identify the mechanism underlying the reduction in NEIL1 and NEIL2 expression in cancer, we investigated whether these genes were epigenetically silenced in cancer using DNA methylation data from the TCGA database. Nine [breast invasive carcinoma, colon adenocarcinoma, HNSCC, clear cell renal cell carcinoma (RCC), papillary RCC, lung adenocarcinoma, lung squamous cell carcinoma, rectal adenocarcinoma, and stomach adenocarcinoma] of the 13 (69.2%) cancer types satisfied the 4 criteria for epigenetic silencing described in Section 2 for the NEIL1 gene, whereas none of the cancer types satisfied the criteria for the NEIL2 or NEIL3 gene (Table 2, Figures 2(a) and 2(b), Supplementary Figure S5, and Supplementary Table S6). Together with a previous finding that the region around the transcription start site of the NEIL1 gene exhibits promoter activity [24], these results suggest that these cancer types exhibit epigenetic silencing of the NEIL1 via promoter hypermethylation. To confirm the possibility of NEIL1 epigenetic silencing, we treated two gastric cancer cell lines (MKN45 and MKN74) with the cytosine methylation inhibitor 5-aza-dC and measured the levels of NEIL1 expression using QRT-PCR. The expression level of NEIL1, but not of NEIL3, was increased in both cell lines by the 5-aza-dC treatment, strengthening the notion of the epigenetic silencing of NEIL1 expression

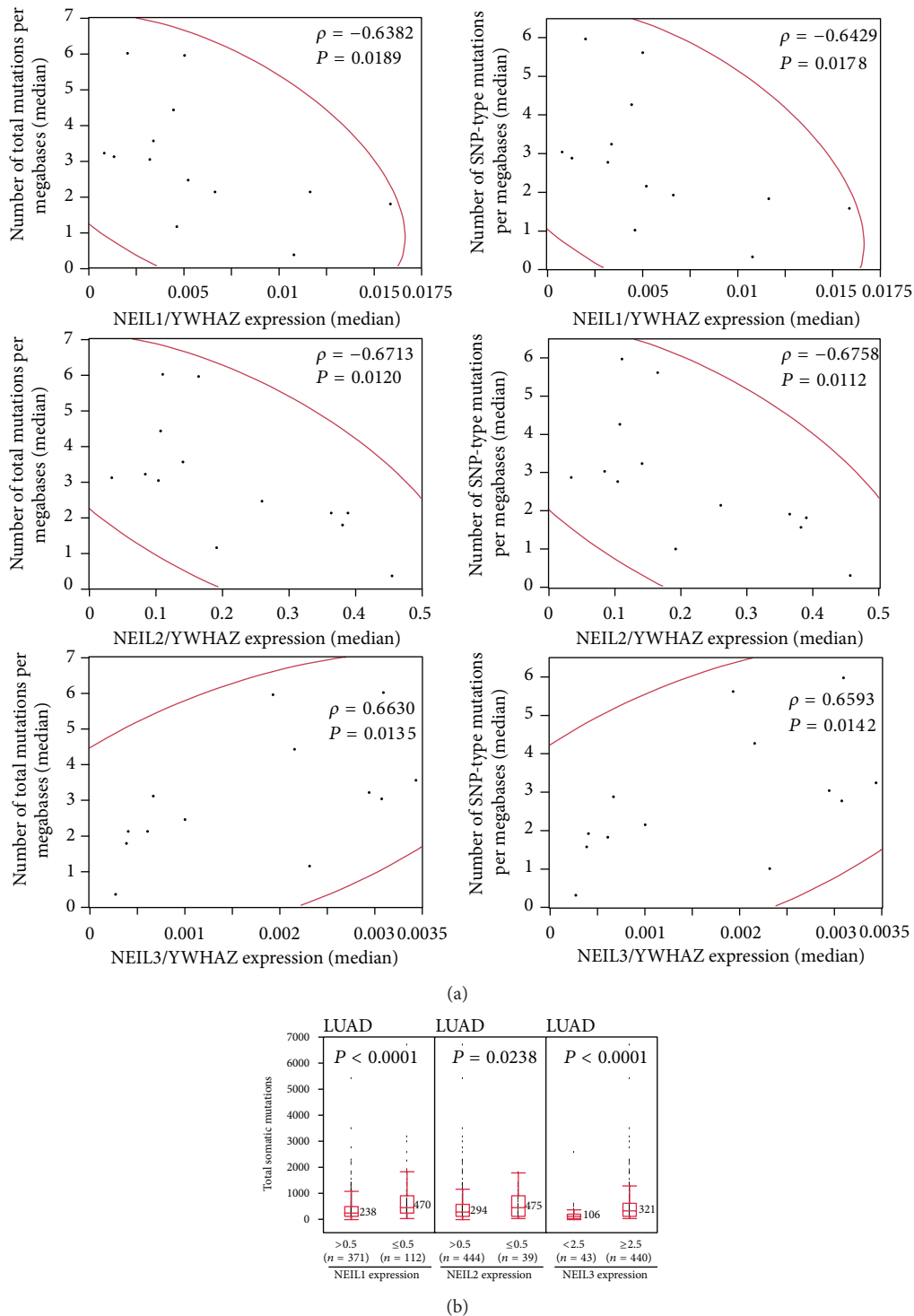


FIGURE 1: Associations between the expression levels of NEIL1, NEIL2, and NEIL3 and the somatic mutation load in human cancer. (a) Scatter plots of the median NEIL1, NEIL2, and NEIL3 expression levels and the median mutation loads in 13 cancer types, based on data from the TCGA database. The expression data for each gene was divided by that for the YWHAZ housekeeping gene. The median number of total mutations per Mb (left panels) or the median number of SNP-type mutations per Mb (right panels) was analyzed, and the Spearman rank correlation coefficient ( $\rho$ ) and  $P$  values were provided. In the analysis, the prevalence of somatic mutations in exomes was calculated based on the identified mutations in the captured region. A bivariate normal ellipse ( $P = 0.95$ ) was observed. (b) Comparison of the total somatic mutation loads between the group showing abnormal NEIL1, NEIL2, and NEIL3 expressions and the other group in lung adenocarcinoma ( $n = 483$ ), as performed using a box-plot analysis of the data from the TCGA database. Values that were 0.5-fold the median NEIL1 expression value, 0.5-fold the median NEIL2 expression value, and 2.5-fold the median NEIL3 expression value in noncancerous lung tissue were used as the cut-off values to dichotomize the cancer cases. The  $P$  values (Mann-Whitney  $U$  test) and median mutation values are shown.

TABLE 1: Associations between elevated NEIL3 expression levels and increased numbers of somatic mutations in human cancers.

Organ	TCGA ID	Number of cases	Mann-Whitney $U$ test		$P^c$ ("increase" or "decrease" in mutation number)	Spearman rank correlation	
			Grouped by NEIL3 expression level (<2.5/ $\geq$ 2.5 or <10/ $\geq$ 10 <sup>a</sup> )	Number of cases		Rho	$P^d$
			Median mutation number per sample <sup>b</sup>				
Urinary bladder	BLCA	129	122/ <b>236</b>	8/121	0.0341 ( <b>increase</b> )	0.1593	0.0714
Breast	BRCA	977	30/ <b>45</b>	62/915	<0.0001 ( <b>increase</b> )	0.3006	< <b>0.0001</b>
Colon	COAD	209	127/ <b>146.5</b>	83/126	0.0942	0.1121	0.1061
Head and neck	HNSC	489	151/ <b>167.5</b>	203/286	0.0477 ( <b>increase</b> )	0.0019	0.9673
Kidney	KICH	66	81.5/ <b>92</b>	28/38	0.3705	0.2637	<b>0.0324</b>
Kidney	KIRC	212	<b>387</b> /86.5	12/200	0.0770	0.1179	0.0867
Kidney	KIRP	168	<b>92</b> /87.5	20/148	0.7409	0.0117	0.8800
Lung	LUAD	483	106/ <b>321</b>	43/440	<0.0001 ( <b>increase</b> )	0.3287	< <b>0.0001</b>
Lung	LUSC	179	196.5/ <b>316</b>	8/171	0.0133 ( <b>increase</b> )	0.2112	<b>0.0045</b>
Prostate	PRAD	258	50/ <b>60.5</b>	76/182	0.0012 ( <b>increase</b> )	0.3030	< <b>0.0001</b>
Rectum	READ	81	113.5/ <b>123</b>	32/49	0.2707	0.2900	<b>0.0086</b>
Stomach	STAD	224	92/ <b>187</b>	76/148	<0.0001 ( <b>increase</b> )	0.5105	< <b>0.0001</b>
Thyroid gland	THCA	404	9/ <b>10</b>	220/184	0.1612	0.0690	0.1664

<sup>a</sup>A value 2.5-fold the median NEIL3 expression value in noncancerous tissue samples of each organ was used as the cut-off value to dichotomize the cancer cases. In the LUSC cases, a value 10-fold the median NEIL3 expression value in noncancerous lung tissue samples was used.

<sup>b</sup>Higher numbers of median somatic mutation per sample are shown in bold face.

<sup>c</sup>A Mann-Whitney  $U$  test was used to perform the statistical analysis. If the  $P$  value was less than 0.05, indicating a significant change, a significant "increase" or "decrease" in the number of somatic mutations per sample was shown.

<sup>d</sup>If significant (less than 0.05), the  $P$  value was shown in bold face.

(Figure 2(c)). We next compared the level of NEIL1 protein expression between cancerous tissues and corresponding noncancerous epithelial tissues using an immunohistochemical analysis of 77 primary HNSCCs. The NEIL1 protein expression level was significantly lower in the cancerous tissues than in the noncancerous tissues ( $P < 0.0001$  by Wilcoxon matched pairs test) (Figure 2(d), Supplementary Table S7). This result suggests that the level of NEIL1 protein expression is reduced in a subset of primary HNSCCs, possibly supporting the idea that the NEIL1 expression level was reduced in cancer because of promoter hypermethylation. Finally, we investigated the impact of the reduction in NEIL1 expression in cancer on the overall survival of the patients. A Kaplan-Meier analysis showed that a reduction in NEIL1 expression was associated with a poorer outcome in patients with breast invasive carcinoma ( $P = 0.0025$ , log-rank test) (Figure 2(e)) but not in patients with the other 12 cancer types. Moreover, a multivariate analysis using the Cox proportional hazard model showed that a reduction in NEIL1 expression was associated with a significantly elevated risk of a poor survival outcome among patients with breast invasive carcinoma (HR: 2.194; 95% CI: 1.417–3.394;  $P = 0.0005$ ) (Supplementary Table S8). These results suggest that a reduction in NEIL1 expression is an independent predictor of a poor survival outcome among patients with breast invasive carcinoma.

**3.4. Cooccurrence of Elevated NEIL3 and APOBEC3B Expressions in Human Cancer.** Since NEIL1, NEIL2, and NEIL3

have been experimentally shown to have the ability to suppress mutations in human cells and/or in bacterial cells [9–12], the finding that the reductions in NEIL1 and NEIL2 expression were associated with the increase in the number of somatic mutations in cancer seems reasonable. However, the association between the elevation in NEIL3 expression and the increased number of somatic mutations in cancer is surprising. To clarify the reason for this association, we investigated the relationship between the expressions of NEIL3 and APOBEC3B, a known inducer of mutations [15, 16]. The APOBEC3B expression level was significantly higher in the group of cancers with a high NEIL3 expression level than in the group of cancers with a low NEIL3 expression level in 10 of the 13 (76.9%) cancer types (Table 3, Supplementary Figure S6). Moreover, a significant positive correlation was found between the NEIL3 and APOBEC3B expression levels in 10 (76.9%) cancer types (Table 3, Supplementary Figure S6). These results suggested that the expressions of NEIL3 and APOBEC3B were positively correlated in human cancer. We suspect that this correlation may explain why the elevation in NEIL3 expression was associated with an increased number of somatic mutations in cancer.

## 4. Discussion

Using data for 13 cancer types from the TCGA database, we revealed that the median somatic total and SNP-type mutation loads exhibited significant inverse correlations with the median NEIL1 and NEIL2 expression levels and a significant

TABLE 2: Epigenetic silencing of NEIL1 expression in human cancer.

Organ (TCGA ID)	cg number (CpG site ID)	Number of cases (N/T)	DNA methylation level		Percentage of the hypermethylated tumors <sup>a</sup>	Gene expression level		DNA methylation level and gene expression level		
			Difference in the $\beta$ value between the 90th percentile of $\beta$ value in tumor tissue and the mean in normal tissue	Difference in the $\beta$ value between the 90th percentile of $\beta$ value in normal tissue and the mean of the 10% of tumor tissue with the highest $\beta$ value		Number of cases (N/T)	Fold expression change between the mean in normal tissue and the mean of the 10% of tumor tissue with the highest $\beta$ value	Number of cases (N/T)	Rho	$P$
Breast (BRCA)	12978308	96/745	0.109	0.109	7.2%	2.039	69/723	69/723	-0.3729	<0.0001
Colon (COAD)	12978308	38/301	0.291	0.291	29.9%	2.087	19/250	19/250	-0.5241	<0.0001
Head and neck (HNSC)	12978308	50/529	0.323	0.323	57.7%	2.354	20/498	20/498	-0.3699	<0.0001
Kidney (KIRC)	00836571	160/325	0.110	0.110	7.1%	2.407	24/303	24/303	-0.4755	<0.0001
Kidney (KIRP)	12978308	45/226	0.351	0.351	21.2%	2.276	23/182	23/182	-0.6219	<0.0001
Lung (LUAD)	12978308	32/465	0.109	0.109	64.3%	2.249	21/422	21/422	-0.3695	<0.0001
Lung (LUSC)	12978308	42/359	0.254	0.254	27.0%	1.624	8/358	8/358	-0.2799	<0.0001
Rectum (READ)	12978308	7/99	0.145	0.145	11.1%	1.998	2/90	2/90	-0.3699	0.0003
Stomach (STAD)	12978308	2/325	0.290	0.290	40.3%	2.208 <sup>b</sup>	0/231	0/231	-0.5465	<0.0001

Epigenetic silencing of the *NEIL1* gene was determined according to the following four criteria: (1) a mean  $\beta$  value in normal tissue < 0.4; (2) a difference in the  $\beta$  value between the 90th percentile of  $\beta$  value in tumor tissue and the mean in normal tissue > 0.1; (3) a fold expression change between the mean in normal tissue and the mean of the 10% of tumor tissue with the highest  $\beta$  value > 1.5; (4) a Spearman rank correlation value between DNA methylation and gene expression < -0.25. This table includes only the cancers that fulfilled these four criteria.

<sup>a</sup>Percentage of cancers with the following  $\beta$  value: the mean  $\beta$  value in normal tissue + more than 0.15.

<sup>b</sup>Since there is no normal tissue with DNA methylation data in STAD, the mean expression value of 33 gastric normal tissues used in Supplementary Table S2 was utilized for the calculation.

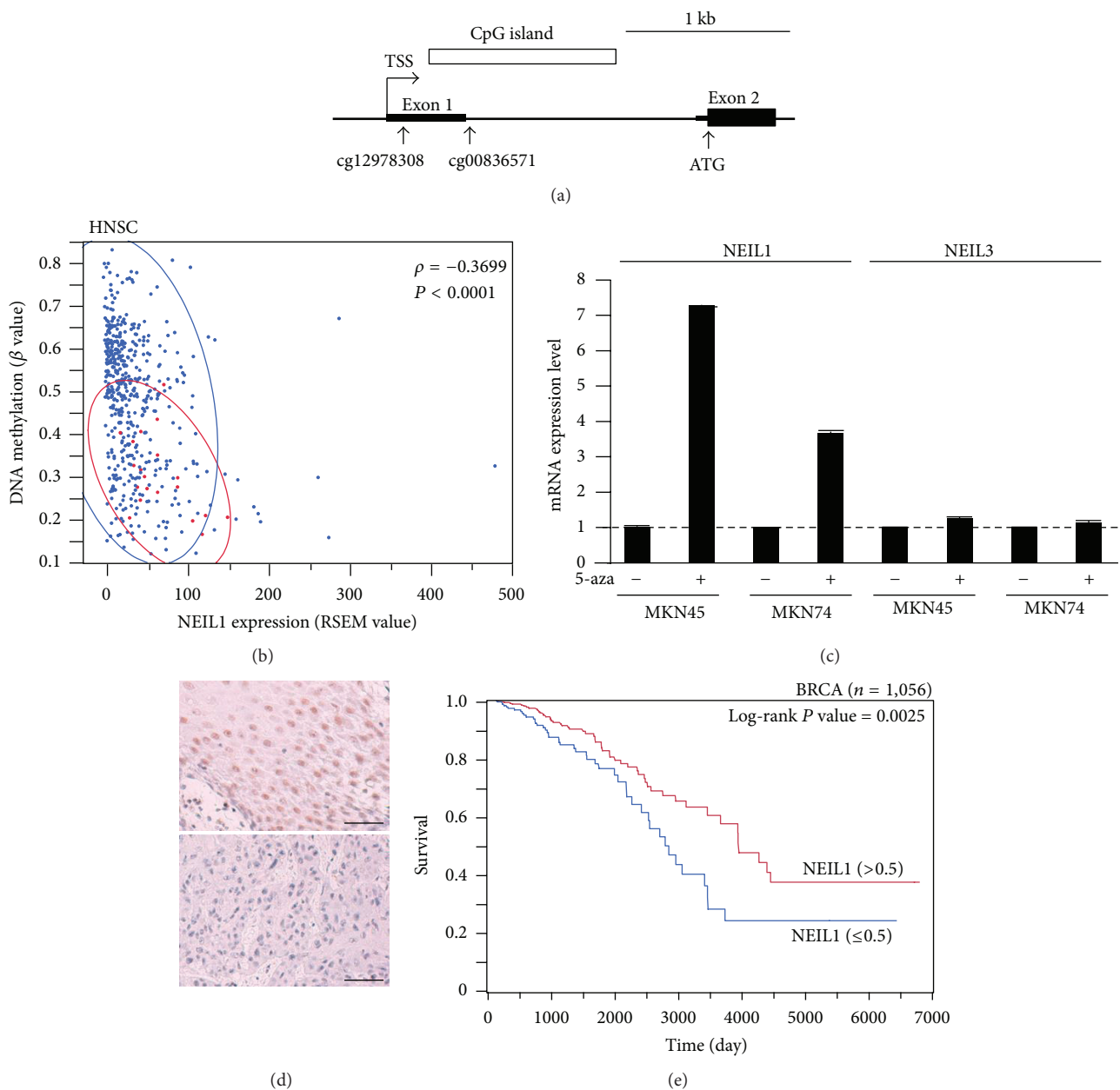


FIGURE 2: Epigenetic silencing of *NEIL1* expression in human cancer. (a) Map of the DNA methylation probes near the transcription start sites (TSSs) of the *NEIL1* gene. The vertical arrows mark the position of the DNA methylation probes (CpG sites) or the translation initiation site (ATG). The thicker section in the exon region indicates the coding sequence. (b) Representative result showing the inverse correlation between DNA methylation at the *NEIL1* CpG site and *NEIL1* expression in cancer. A scatter plot analysis was performed for DNA methylation at the cg12978308 probe site and the *NEIL1* mRNA expression level in HNSCC using data from the TCGA database. The Spearman rank correlation coefficient ( $\rho$ ) and  $P$  values were provided. A bivariate normal ellipse ( $P = 0.95$ ) was observed for normal tissue samples (red) and cancerous tissue samples (blue). (c) Effects of 5-aza-dC on the *NEIL1* and *NEIL3* expression levels in gastric cancer cell lines. The cell lines were treated with 5-aza-dC, and the *NEIL1* and *NEIL3* expression levels were measured using a real-time QRT-PCR analysis. The amounts of *NEIL1* or *NEIL3* transcripts normalized to the amount of GAPDH transcript are shown in the graph. The average expression levels in untreated cells were set at 1.0. Values are the mean  $\pm$  standard error of three independent experiments. (d) Downregulation of *NEIL1* protein expression in primary HNSCC. Representative results for *NEIL1* expression in noncancerous head and neck epithelium (upper panel) and HNSCC (lower panel) are shown. Scale bar = 50  $\mu$ m. (e) Impact of reduced *NEIL1* expression on overall survival in primary breast cancer patients. The survival curves for breast cancer patients ( $n = 1,056$ ) were based on data from the TCGA database and were generated using the Kaplan-Meier method. The patients were divided into two groups using a cut-off value of 0.5-fold the median *NEIL1* expression value in noncancerous breast tissue. Log-rank:  $P = 0.0025$ .

TABLE 3: Associations between NEIL3 and APOBEC3B expression levels in human cancer.

Organ	TCGA ID	Number of cases	Grouped by NEIL3 expression level (<2.5/≥2.5 <sup>a</sup> )		Mann-Whitney <i>U</i> test	Spearman rank correlation	
			Median APOBEC3B expression <sup>b</sup>	Number of cases	<i>P</i> <sup>c</sup> (“increase” or “decrease” in APOBEC3B expression)	Rho	<i>P</i> <sup>d</sup>
Urinary bladder	BLCA	241	146/ <b>382</b>	18/223	0.0034 ( <b>increase</b> )	0.0697	0.2811
Breast	BRCA	1056	45/ <b>175</b>	76/980	<0.0001 ( <b>increase</b> )	0.5215	< <b>0.0001</b>
Colon	COAD	260	169/ <b>231</b>	108/152	0.0191 ( <b>increase</b> )	0.1569	<b>0.0113</b>
Head and neck	HNSC	498	<b>501</b> /466	209/289	0.2122	0.2326	< <b>0.0001</b>
Kidney	KICH	66	132/ <b>149</b>	28/38	0.5463	0.1312	0.2935
Kidney	KIRC	519	26.9/ <b>65.1</b>	31/488	<0.0001 ( <b>increase</b> )	0.5777	< <b>0.0001</b>
Kidney	KIRP	198	33.1/ <b>55</b>	22/176	0.0017 ( <b>increase</b> )	0.4466	< <b>0.0001</b>
Lung	LUAD	490	60.4/ <b>159</b>	43/447	<0.0001 ( <b>increase</b> )	0.2759	< <b>0.0001</b>
Lung	LUSC	490	195/ <b>496</b>	5/485	0.0187 ( <b>increase</b> )	0.1302	<b>0.0039</b>
Prostate	PRAD	333	19/ <b>32.9</b>	97/236	<0.0001 ( <b>increase</b> )	0.5876	< <b>0.0001</b>
Rectum	READ	92	172/ <b>247</b>	36/56	0.0985	0.1559	0.1379
Stomach	STAD	238	2.75/ <b>3.92</b>	79/159	0.0007 ( <b>increase</b> )	0.1848	<b>0.0042</b>
Thyroid gland	THCA	508	37.2/ <b>66.5</b>	276/232	<0.0001 ( <b>increase</b> )	0.4242	< <b>0.0001</b>

<sup>a</sup> A value 2.5-fold the median NEIL3 expression value in noncancerous tissue samples of each organ was used as the cut-off value to dichotomize the cancer cases.

<sup>b</sup> Higher numbers of median APOBEC3B expression values are shown in bold face. RPKM value was used to show expression level in stomach cancer; on the other hand RSEM value was used in the other organs' cancers.

<sup>c</sup> A Mann-Whitney *U* test was used to perform the statistical analysis. If the *P* value was less than 0.05, indicating a significant change, a significant “increase” or “decrease” in the APOBEC3B expression was shown.

<sup>d</sup> If significant (less than 0.05), the *P* value was shown in bold face.

positive correlation with the median NEIL3 expression level. We also showed that a subset of human cancers exhibited reduced NEIL1 and NEIL2 expression levels and an elevated NEIL3 expression level, and these abnormal expressions of NEIL1, NEIL2, and NEIL3 were associated with the mutation load in cancer. We then showed that the reduced NEIL1 expression level observed in various cancers was due to epigenetic silencing by promoter hypermethylation and that such reduction was an independent predictor of a poor outcome among patients with breast invasive carcinoma. Finally, NEIL3 expression was shown to be correlated with the expression of APOBEC3B, a potent inducer of mutations, possibly explaining why an increased NEIL3 expression level was associated with the somatic mutation load in cancer. Thus, our results suggest that the abnormal regulation of NEIL1, NEIL2, and NEIL3 expression is involved in the development of cancer via an increase in the prevalence of somatic mutations, providing a new and important link between abnormalities in the DNA glycosylases NEIL1, NEIL2, and NEIL3 and human cancer.

Using a TCGA-based analysis, associations between abnormal NEIL1, NEIL2, or NEIL3 expressions and the somatic mutation load were apparently demonstrated in various cancer types for the first time. The association

between reductions in NEIL1 and NEIL2 expressions and the increased number of somatic mutations in cancer is understandable, but the association between an elevation in NEIL3 expression and an increased number of somatic mutations in cancer seems surprising at first glance, since NEIL1, NEIL2, and NEIL3 all have the ability to suppress mutations [9–12]. The upregulation of NEIL3 expression in diverse cancer types is consistent with the results of a previous report by Hildrestrand et al. [25], but the effect of such upregulation on cancer has not yet been determined. Our demonstration of a correlation between NEIL3 expression and APOBEC3B expression may explain why an increase in NEIL3 expression is associated with the somatic mutation load, since APOBEC3B is involved in mutagenesis in multiple distinct human cancers [15, 16]. Although the precise mechanism was not investigated in the present study, we speculated that since the effect of APOBEC3B on the increase in mutations may be greater than the effect of NEIL3 on a decrease in mutations through its DNA glycosylase activity, the coelevated expressions of NEIL3 and APOBEC3B may lead to the observed increase in the number of somatic mutations in cancer. Alternatively, NEIL3 might be involved in APOBEC3B-induced mutagenesis. Further investigation of such issues is needed.



In this study, we found 9 cancer types that showed epigenetic silencing of the *NEIL1* gene via promoter hypermethylation using data from the TCGA database. Among them, the epigenetic silencing of *NEIL1* expression in HNSCC, lung adenocarcinoma, lung squamous cell carcinoma, colon adenocarcinoma, and rectal adenocarcinoma was consistent with the findings of previous reports [24, 26, 27], whereas the findings in the remaining 4 cancer types, that is, breast invasive carcinoma, clear cell RCC, papillary RCC, and stomach adenocarcinoma, were novel findings. Although further experiments, such as 5-aza-dC treatment and a *NEIL1* protein expression analysis for each of the latter 4 cancer types, are needed to determine the epigenetic silencing of the *NEIL1* gene via promoter hypermethylation in these cancer types, we suspect that the epigenetic silencing of the *NEIL1* gene via promoter hypermethylation might be the chief mechanism underlying the downregulation of *NEIL1* expression in diverse human cancers. Interestingly, in breast invasive carcinoma, which is one of the cancers that shows the epigenetic silencing of *NEIL1*, a reduction in *NEIL1* expression was shown to be an independent predictor of a poor survival outcome. This novel finding may be useful for the management of breast cancer patients, and if this marker is used in conjunction with other prognosis markers, such as the hormone receptor status [28], the management of breast cancer patients could be further improved. Regarding this point, in our preliminary analysis using data from the TCGA database, combinations of the *NEIL1* mRNA expression level and either the hormone receptor status or the *HER2* status were shown to be excellent prognostic markers (Supplementary Figure S7). Since a reduction in *NEIL1* expression was associated with an increased somatic mutation level and mutations in cancer-associated genes can lead to the exaggeration of the malignant potential, such as an increase in the proliferation rate, this kind of phenotypic change might explain the difference in survival outcomes between patients with and those without a reduction in *NEIL1* expression.

So far, several forms of germline nonsynonymous *NEIL1* or *NEIL2* mutations have been experimentally demonstrated to actually have reduced or absent repair activity [10, 11, 29, 30]. Human cells containing such *NEIL1* or *NEIL2* mutations are considered to have a reduced capacity to repair mutagenic bases; thus, similar to cancers with a reduced *NEIL1* or *NEIL2* expression levels, a higher incidence of mutation is likely to occur in the cells, leading to cancer susceptibility. This scenario is compatible with a previous paper reporting a germline *NEIL2* variant that is a marker for risk and the progression of squamous cell carcinomas of the oral cavity and oropharynx [31] and that is selectively found in familial colorectal cancer patients, but not in healthy controls [32]. Future genome-wide analyses of cancers derived from individuals with germline *NEIL1* or *NEIL2* mutations should clarify the role of *NEIL1* and *NEIL2* in the prevention of mutations.

In this study, the 0.5-fold, 0.5-fold, and 2.5-fold values of the median expression value in noncancerous tissue samples of each organ were used as cut-off values to dichotomize the *NEIL1*, *NEIL2*, and *NEIL3* expression values in the cancer cases, respectively. If an expression level is downregulated or

upregulated in a disease, a fold-change value of 0.5 and 2.5, respectively, has been used to dichotomize disease cases in previous reports [33, 34]; therefore, these values were used in our analysis.

In conclusion, our study indicates that the abnormal expressions of *NEIL1*, *NEIL2*, and *NEIL3* are likely to be involved in mutagenesis in human cancer. Since little is known about gene abnormalities identified by whole-exome sequencing data that induce mutations in cancer, our findings regarding these novel mutagenic factors should contribute to our general understanding of human cancer.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

This work was supported in part by a Grant-in-Aid from the MHLW (21-1), the JSPS (25460476, 24659161), the MEXT (221S0001), the Takeda Science Foundation, and the Smoking Research Foundation.

## References

- [1] T. K. Hazra, T. Izumi, I. Boldogh et al., "Identification and characterization of a human DNA glycosylase for repair of modified bases in oxidatively damaged DNA," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 6, pp. 3523–3528, 2002.
- [2] T. K. Hazra, Y. W. Kow, Z. Hatahet et al., "Identification and characterization of a novel human DNA glycosylase for repair of cytosine-derived lesions," *The Journal of Biological Chemistry*, vol. 277, no. 34, pp. 30417–30420, 2002.
- [3] I. Morland, V. Rolseth, L. Luna, T. Rognes, M. Bjørås, and E. Seeberg, "Human DNA glycosylases of the bacterial Fpg/MutM superfamily: an alternative pathway for the repair of 8-oxoguanine and other oxidation products in DNA," *Nucleic Acids Research*, vol. 30, no. 22, pp. 4926–4936, 2002.
- [4] K. Shinmura, H. Tao, M. Goto et al., "Inactivating mutations of the human base excision repair gene *NEIL1* in gastric cancer," *Carcinogenesis*, vol. 25, no. 12, pp. 2311–2317, 2004.
- [5] A. Prakash, S. Doublé, and S. S. Wallace, "The Fpg/Nei family of DNA glycosylases: substrates, structures, and search for damage," *Progress in Molecular Biology and Translational Science*, vol. 110, pp. 71–91, 2012.
- [6] I. R. Grin and D. O. Zharkov, "Eukaryotic endonuclease VIII-Like proteins: new components of the base excision DNA repair system," *Biochemistry*, vol. 76, no. 1, pp. 80–93, 2011.
- [7] S. Z. Krokeide, J. K. Laerdahl, M. Salah et al., "Human *NEIL3* is mainly a monofunctional DNA glycosylase removing spiroimidiohydantoin and guanidinohydantoin," *DNA Repair*, vol. 12, no. 12, pp. 1159–1164, 2013.
- [8] M. Liu, S. Doublé, and S. S. Wallace, "Neil3, the final frontier for the DNA glycosylases that recognize oxidative damage," *Mutation Research*, vol. 743–744, pp. 4–11, 2013.
- [9] A. K. Maiti, I. Boldogh, H. Spratt, S. Mitra, and T. K. Hazra, "Mutator phenotype of mammalian cells due to deficiency

- of NEIL1 DNA glycosylase, an oxidized base-specific repair enzyme,” *DNA Repair*, vol. 7, no. 8, pp. 1213–1220, 2008.
- [10] S. Dey, A. K. Maiti, M. L. Hegde et al., “Increased risk of lung cancer associated with a functionally impaired polymorphic variant of the human DNA glycosylase NEIL2,” *DNA Repair*, vol. 11, no. 6, pp. 570–578, 2012.
- [11] K. Shinmura, H. Kato, Y. Kawanishi et al., “NEIL1 p.Gln282Stop variant is predominantly localized in the cytoplasm and exhibits reduced activity in suppressing mutations,” *Gene*, vol. 571, no. 1, pp. 33–42, 2015.
- [12] M. Liu, V. Bandaru, J. P. Bond et al., “The mouse ortholog of NEIL3 is a functional DNA glycosylase in vitro and in vivo,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 11, pp. 4925–4930, 2010.
- [13] Cancer Genome Atlas Network, “Comprehensive molecular characterization of human colon and rectal cancer,” *Nature*, vol. 487, no. 7407, pp. 330–337, 2012.
- [14] L. B. Alexandrov, S. Nik-Zainal, D. C. Wedge et al., “Signatures of mutational processes in human cancer,” *Nature*, vol. 500, no. 7463, pp. 415–421, 2013.
- [15] S. A. Roberts, M. S. Lawrence, L. J. Klimczak et al., “An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers,” *Nature Genetics*, vol. 45, no. 9, pp. 970–976, 2013.
- [16] M. B. Burns, N. A. Temiz, and R. S. Harris, “Evidence for APOBEC3B mutagenesis in multiple human cancers,” *Nature Genetics*, vol. 45, no. 9, pp. 977–983, 2013.
- [17] S. Yadav, S. Mukhopadhyay, M. Anbalagan, and N. Makridakis, “Somatic mutations in catalytic core of POLK reported in prostate cancer alter translesion DNA synthesis,” *Human Mutation*, vol. 36, no. 9, pp. 873–880, 2015.
- [18] B. Li and C. N. Dewey, “RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome,” *BMC Bioinformatics*, vol. 12, article 323, 2011.
- [19] A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer, and B. Wold, “Mapping and quantifying mammalian transcriptomes by RNA-Seq,” *Nature Methods*, vol. 5, no. 7, pp. 621–628, 2008.
- [20] K. Shinmura, M. Goto, M. Suzuki et al., “Reduced expression of MUTYH with suppressive activity against mutations caused by 8-hydroxyguanine is a novel predictor of a poor prognosis in human gastric cancer,” *The Journal of Pathology*, vol. 225, no. 3, pp. 414–423, 2011.
- [21] J. Vandesompele, K. De Preter, F. Pattyn et al., “Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes,” *Genome Biology*, vol. 3, no. 7, 2002.
- [22] E. Eisenberg and E. Y. Levanon, “Human housekeeping genes are compact,” *Trends in Genetics*, vol. 19, no. 7, pp. 362–365, 2003.
- [23] E. Kriegova, A. Arakelyan, R. Fillerova et al., “PSMB2 and RPL32 are suitable denominators to normalize gene expression profiles in bronchoalveolar cells,” *BMC Molecular Biology*, vol. 9, article 69, 2008.
- [24] J. Chaisaingmongkol, O. Popanda, R. Warta et al., “Epigenetic screen of human DNA repair genes identifies aberrant promoter methylation of NEIL1 in head and neck squamous cell carcinoma,” *Oncogene*, vol. 31, no. 49, pp. 5108–5116, 2012.
- [25] G. A. Hildrestrand, C. G. Neurauder, D. B. Diep et al., “Expression patterns of Neil3 during embryonic brain development and neoplasia,” *BMC Neuroscience*, vol. 10, article 45, 2009.
- [26] H. Do, N. C. Wong, C. Murone et al., “A critical re-assessment of DNA repair gene promoter methylation in non-small cell lung carcinoma,” *Scientific Reports*, vol. 4, article 4186, 2014.
- [27] S. A. Farkas, V. Vymetalkova, L. Vodickova, P. Vodicka, and T. K. Nilsson, “DNA methylation changes in genes frequently mutated in sporadic colorectal cancer and in the DNA repair and Wnt/ $\beta$ -catenin signaling pathway genes,” *Epigenomics*, vol. 6, no. 2, pp. 179–191, 2014.
- [28] C.-H. Yip and A. Rhodes, “Estrogen and progesterone receptors in breast cancer,” *Future Oncology*, vol. 10, no. 14, pp. 2293–2301, 2014.
- [29] L. M. Roy, P. Jaruga, T. G. Wood, A. K. McCullough, M. Dizdaroglu, and R. S. Lloyd, “Human polymorphic variants of the NEIL1 DNA glycosylase,” *The Journal of Biological Chemistry*, vol. 282, no. 21, pp. 15790–15798, 2007.
- [30] A. Prakash, B. L. Carroll, J. B. Sweasy, S. S. Wallace, and S. Doublié, “Genome and cancer single nucleotide polymorphisms of the human NEIL1 DNA glycosylase: activity, structure, and the effect of editing,” *DNA Repair*, vol. 14, no. 1, pp. 17–26, 2014.
- [31] X. Zhai, H. Zhao, Z. Liu et al., “Functional variants of the NEIL1 and NEIL2 genes and risk and progression of squamous cell carcinoma of the oral cavity and oropharynx,” *Clinical Cancer Research*, vol. 14, no. 13, pp. 4345–4352, 2008.
- [32] P. Broderick, T. Bagratuni, J. Vijayakrishnan, S. Lubbe, I. Chandler, and R. S. Houlston, “Evaluation of NTHL1, NEIL1, NEIL2, MPG, TDG, UNG, and SMUG1 genes in familial colorectal cancer predisposition,” *BMC Cancer*, vol. 6, article 243, 2006.
- [33] M.-A. Seol, I.-S. Chu, M.-J. Lee et al., “Genome-wide expression patterns associated with oncogenesis and sarcomatous transdifferentiation of cholangiocarcinoma,” *BMC Cancer*, vol. 11, article 78, 2011.
- [34] M. Hashimoto, M. Ichihara, T. Watanabe et al., “Expression of CD109 in human cancer,” *Oncogene*, vol. 23, no. 20, pp. 3716–3720, 2004.