

BASIC AND TRANSLATIONAL—ALIMENTARY TRACT

Genomic Landscape of Esophageal Squamous Cell Carcinoma in a Japanese Population



Genta Sawada,^{1,2,*} Atsushi Niida,^{3,*} Ryutaro Uchi,¹ Hidenari Hirata,¹ Teppei Shimamura,³ Yutaka Suzuki,⁴ Yuichi Shiraishi,³ Kenichi Chiba,³ Seiya Imoto,³ Yusuke Takahashi,^{1,2} Takeshi Iwaya,⁵ Tomoya Sudo,¹ Tomoatsu Hayashi,⁶ Hiroki Takai,⁶ Yoshihiro Kawasaki,⁶ Takashi Matsukawa,⁷ Hidetoshi Eguchi,¹ Keishi Sugimachi,¹ Fumiaki Tanaka,¹ Hiromichi Suzuki,⁸ Ken Yamamoto,⁹ Hideshi Ishii,² Makiko Shimizu,¹⁰ Hiroshi Yamazaki,¹⁰ Makoto Yamazaki,² Yuji Tachimori,¹¹ Yoshiaki Kajiyama,¹² Shoji Natsugoe,¹³ Hiromasa Fujita,¹⁴ Kenichi Mafune,¹⁵ Yoichi Tanaka,¹⁶ David P. Kelsell,¹⁷ Claire A. Scott,¹⁷ Shoji Tsuji,⁷ Shinichi Yachida,¹⁸ Tatsuhiro Shibata,^{19,20} Sumio Sugano,⁴ Yuichiro Doki,² Tetsu Akiyama,⁶ Hiroyuki Aburatani,²¹ Seishi Ogawa,⁸ Satoru Miyano,³ Masaki Mori,² and Koshi Mimori¹

¹Department of Surgery, Kyushu University Beppu Hospital, Beppu, Japan; ²Department of Gastroenterological Surgery, Graduate School of Medicine, Osaka University, Suita, Japan; ³Laboratory of DNA Information Analysis, Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan; ⁴Laboratory of Functional Genomics, Department of Medical Genome Sciences, Graduate School of Frontier Sciences, University of Tokyo, Kashiwa, Japan; ⁵Department of Surgery, Iwate Medical University, Morioka, Japan; ⁶Laboratory of Molecular and Genetic Information, Institute of Molecular and Cellular Biosciences, University of Tokyo, Tokyo, Japan; ⁷Department of Neurology, University of Tokyo, Japan; ⁸Department of Pathology and Tumor Biology, Kyoto University, Kyoto, Japan; ⁹Department of Medical Chemistry, Kurume University School of Medicine, Kurume, Japan; ¹⁰Laboratory of Drug Metabolism and Pharmacokinetics, Showa Pharmaceutical University, Tokyo, Japan; ¹¹Esophageal Surgery Division, National Cancer Center Hospital, Tokyo, Japan; ¹²Department of Esophageal and Gastroenterological Surgery, Juntendo University School of Medicine, Tokyo, Japan; ¹³Department of Surgical Oncology and Digestive Surgery, Kagoshima University School of Medicine, Kagoshima, Japan; ¹⁴Department of Surgery, Kurume University School of Medicine, Kurume, Japan; ¹⁵Department of Surgery, Mitsui Kinen Hospital, Tokyo, Japan; ¹⁶Division of Gastroenterological Surgery, Saitama Cancer Center, Saitama, Japan; ¹⁷Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK; ¹⁸Division of Refractory Cancer Research, National Cancer Center Research Institute, Tokyo, Japan; ¹⁹Division of Cancer Genomics, Center for Medical Genomics, National Cancer Center Research Institute, Tokyo, Japan; ²⁰Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan; and ²¹Genome Science Laboratory, Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan

See Covering the Cover synopsis on page 1050;
see editorial on page 1080

BACKGROUND & AIM: Esophageal squamous cell carcinoma (ESCC) is the predominant form of esophageal cancer in Japan. Smoking and drinking alcohol are environmental risk factors for ESCC, whereas single nucleotide polymorphisms in *ADH1B* and *ALDH2*, which increase harmful intermediates produced by drinking alcohol, are genetic risk factors. We conducted a large-scale genomic analysis of ESCCs from patients in Japan to determine the mutational landscape of this cancer. **METHODS:** We performed whole-exome sequence analysis of tumor and non-tumor esophageal tissues collected from 144 patients with ESCC who underwent surgery at 5 hospitals in Japan. We also performed single-nucleotide polymorphism array-based copy number profile and germline genotype analyses of polymorphisms in *ADH1B* and *ALDH2*. Polymorphisms in *CYP2A6*, which increase harmful effects of smoking, were analyzed. Functions of *TET2* mutants were evaluated in KYSE410 and HEK293FT cells. **RESULTS:** A high proportion of mutations in the 144 tumor samples were C to T substitution in CpG dinucleotides (called the CpG signature) and C to G/T substitutions with a flanking 5' thymine (called the APOBEC signature). Based on mutational

signatures, patients were assigned to 3 groups, which associated with environmental (drinking and smoking) and genetic (polymorphisms in *ALDH2* and *CYP2A6*) factors. Many tumors contained mutations in genes that regulate the cell cycle (*TP53*, *CCND1*, *CDKN2A*, *FBXW7*); epigenetic processes (*MLL2*, *EP300*, *CREBBP*, *TET2*); and the NOTCH (*NOTCH1*, *NOTCH3*), WNT (*FAT1*, *YAP1*, *AJUBA*) and receptor-tyrosine kinase–phosphoinositide 3-kinase signaling pathways (*PIK3CA*, *EGFR*, *ERBB2*). Mutations in *EP300* and *TET2* correlated with shorter survival times, and mutations in *ZNF750* associated with an increased number of mutations of the APOBEC signature. Expression of mutant forms of *TET2* did not increase cellular levels of 5-hydroxymethylcytosine in HEK293FT cells, whereas knockdown of *TET2* increased the invasive activity of KYSE410 ESCC cells. Computational analyses associated the mutations in *NFE2L2* we identified with transcriptional activation of its target genes. **CONCLUSIONS:** We associated environmental and genetic factors with base substitution patterns of somatic mutations and provide a registry of genes and pathways that are disrupted in ESCCs. These findings might be used to design specific treatments for patients with esophageal squamous cancers.

BASIC AND
TRANSLATIONAL AT

Keywords: Mutational Signature; Exome Sequencing; Copy Number Profiling; Esophagus.

Esophageal squamous cell carcinoma (ESCC) is one of the most aggressive types of cancer, frequently showing lymph node metastasis and tumor invasion into adjacent organs, even in the early stages.^{1,2} In contrast to the predominance of adenocarcinoma in Western countries, squamous cell histology represents the most prevalent form of esophageal cancer in East Asia, including Japan and China.

In Japan, ESCC is the tenth most common malignancy and the seventh most common cause of cancer-related deaths. Epidemiologic studies have established that drinking and smoking are strong risk factors for developing ESCC.³ Furthermore, genetic polymorphisms that impair the functions of alcohol-metabolizing enzymes have been reported as risk factors. Our previous genome-wide association studies have shown that functional single-nucleotide polymorphisms (SNPs) in 2 alcohol dehydrogenase genes, *ADH1B* (rs1229984 GG) and *ALDH2* (rs671 AG/AA), increase the incidence of ESCC in Japanese cohorts.⁴ These findings are consistent with data from another independent study.⁵ Notably, there is drastic synergy among these genetic and environmental risk factors; our study showed that the incidence odds ratio was as high as 146.4 (95% confidence interval [CI]: 50.5–424.5) in the presence of all 4 risk factors.

In China, the incidence and mortality rates of ESCC are higher than those in Japan. ESCC is the fourth most frequently diagnosed cancer and the fourth leading cause of cancer-related deaths in China. There are substantial differences in the epidemiology³ of ESCC between the Japanese and Chinese populations. The association of smoking and drinking with the risk of developing ESCC appears to be weaker in China than in Japan. Notably, a regional bias exists in the incidence of Chinese ESCC. Although the incidence rate is generally higher in rural areas than in urban areas throughout China, some rural areas have reported particularly high incidences. In the high-incidence areas, the association between the incidence rate and drinking and smoking behavior is especially weak; instead, family history and environmental factors, such as nutritional deficiency and food mutagens, are proposed risk factors.^{6–8}

Recently, several whole-exome sequencing (WES) studies have revealed different landscapes of driver genes, as well as somatically disrupted pathways in ESCC.^{9,10} Notably, 3 large-scale WES studies of ESCC cohorts have been carried out in China; Gao et al¹¹ evaluated a low-incidence cohort from an urban area near Beijing, while Song et al¹² and Zhang et al¹³ examined high-incidence cohorts from the Chaoshan District of Guangdong Province and the Taihang Mountains of north central China, respectively. However, no comprehensive genomic profiling of a Japanese ESCC cohort has been reported.

Here, we present the landscape of genomic alterations in Japanese ESCC patients obtained by WES and/or SNP array-based copy number (CN) analysis in 144 Japanese patients. This study not only expands the registry of somatically disrupted driver genes, but also reveals the association of genotype–environment interactions with mutational signatures that are unique to the Japanese ESCC cohort.

Material and Methods

Sample Collection

Samples from 144 patients with diagnosed ESCC were obtained from 5 hospitals (Juntendo University Hospital, National Cancer Center Hospital, Kurume University Hospital, Saitama Cancer Center, and Kagoshima University Hospital). Written informed consent was obtained from all patients and the study protocol was reviewed and approved by the internal review board of Kyushu University.

Whole-Exome Sequencing

DNA extracted from ESCC and paired normal samples was captured using the SureSelect Human All Exon 50Mb Kit (Agilent Technologies, Santa Clara, CA) following the manufacturer's instructions. Captured DNA was sequenced using the HiSeq2000 and paired-end (75–100 bp) sequencing reads were generated for each sample. Mutation calling was performed using the EBcall algorithm¹⁴ (<http://genomon.hgc.jp/exome/en/index.html>).

Copy Number Profiling

DNA was processed and hybridized to the Human-OmniExpress BeadChip (HumanOmniExpress BeadChip Kit; Illumina, San Diego, CA) following the manufacturer's protocol. GenomeStudio (Illumina) was used to obtain B-allele frequencies and normalized logarithmic probe intensities, which were used as input data for an ASCAT analysis to estimate CN profiles along with tumor ploidy and aberrant cell fraction.¹⁵ For details, please see the *Supplementary Material and Methods*.

Results

Effects of Environmental/Genetic Factors on Mutational Signatures

Tumor and paired normal DNA from 144 Japanese ESCC patients were subjected to WES. The mean read depth was 120× and 91.1% of target bases were covered by >10 independent reads (*Supplementary Tables 1 and 2*). A total of 23,121 somatic events, including 22,175 single-nucleotide substitutions and 946 short insertions and deletions (indels), were identified. The mean number of mutations was 161 (range, 47–644) per sample or 3.10 (range, 0.91–12.4) per megabase across the target exome sequences (*Figure 2A* and *Supplementary Table 3*). Similar to previous findings for many cancer types, the predominant substitution across all ESCC samples was C to T involving the CpG dinucleotide (*Figure 1A*).¹⁶ In addition, C to

*Authors share co-first authorship.

Abbreviations used in this paper: CI, confidence interval; CN, copy number; ESCC, esophageal squamous cell carcinoma; HNSCC, head and neck squamous cell carcinoma; PI3K, phosphoinositide 3-kinase; RTK, receptor-tyrosine kinase; SNP, single-nucleotide polymorphism; WES, whole-exome sequencing.

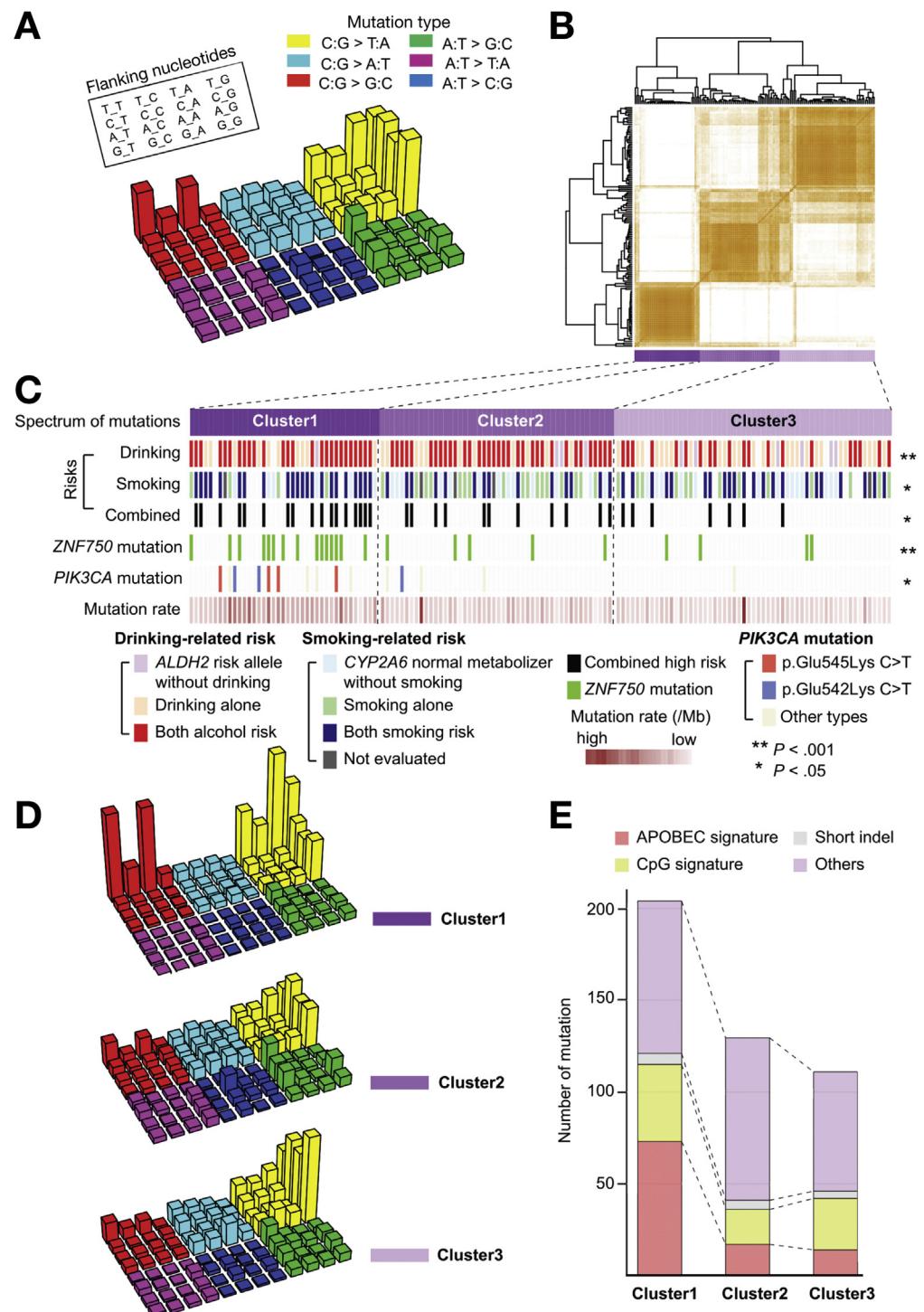
Most current article

G/T substitutions with a flanking 5' thymine characterized the ESCC genomes, particularly when the 3' base was T or A. Recent studies have reported that APOBEC-mediated mutagenesis is associated with this signature.¹⁷ Overall signature profiles based on their trinucleotide context (mutation spectra) were quite similar to those of Chinese ESCC cohorts from high- and low-incidence areas,¹¹⁻¹³ head and neck squamous cell carcinoma (HNSCC),¹⁸ and, to a lesser degree, lung squamous cell carcinoma (Supplementary Figure 1).¹⁹ However, these signatures

are distinct from those reported for esophageal adenocarcinoma,²⁰ suggesting that different carcinogenic mechanisms underlie the 2 histologic cancer types in the same viscera.

To investigate the origin of mutations in individual ESCC cases, we performed a mutation-spectra analysis of the 144 ESCC samples (Supplementary Material and Methods). The samples clustered into 3 distinct groups based on mutation spectra (Figure 1B and C). Cluster 1 and cluster 3 showed a high proportion of APOBEC and CpG signatures, respectively,

Figure 1. Analysis of mutation spectra. (A) ESCC mutation spectra calculated from 144 Japanese ESCC samples subjected to WES. Base substitutions were divided into 96 patterns based on mutation type and nucleotides flanking the mutated base. The height of the bar represents the proportion of each substitution pattern. (B) Stability of the three mutation spectra clusters deduced from negative matrix factorization. (C) Association of the 3 clusters with various factors. To examine the uneven distribution of each factor across the 3 clusters, P values were computed by χ^2 tests and significance is indicated by asterisks. (D) ESCC mutation spectra for each of the 3 clusters. (E) Mean mutation counts for each of the 3 clusters. Stacked bars of different colors represent mean counts of different mutation types in each cluster. The C>G/T substitutions at TpCpN tri-nucleotides are referred to as the APOBEC signature, while the C>T substitutions at NpCpG tri-nucleotides are referred to as the CpG signature.



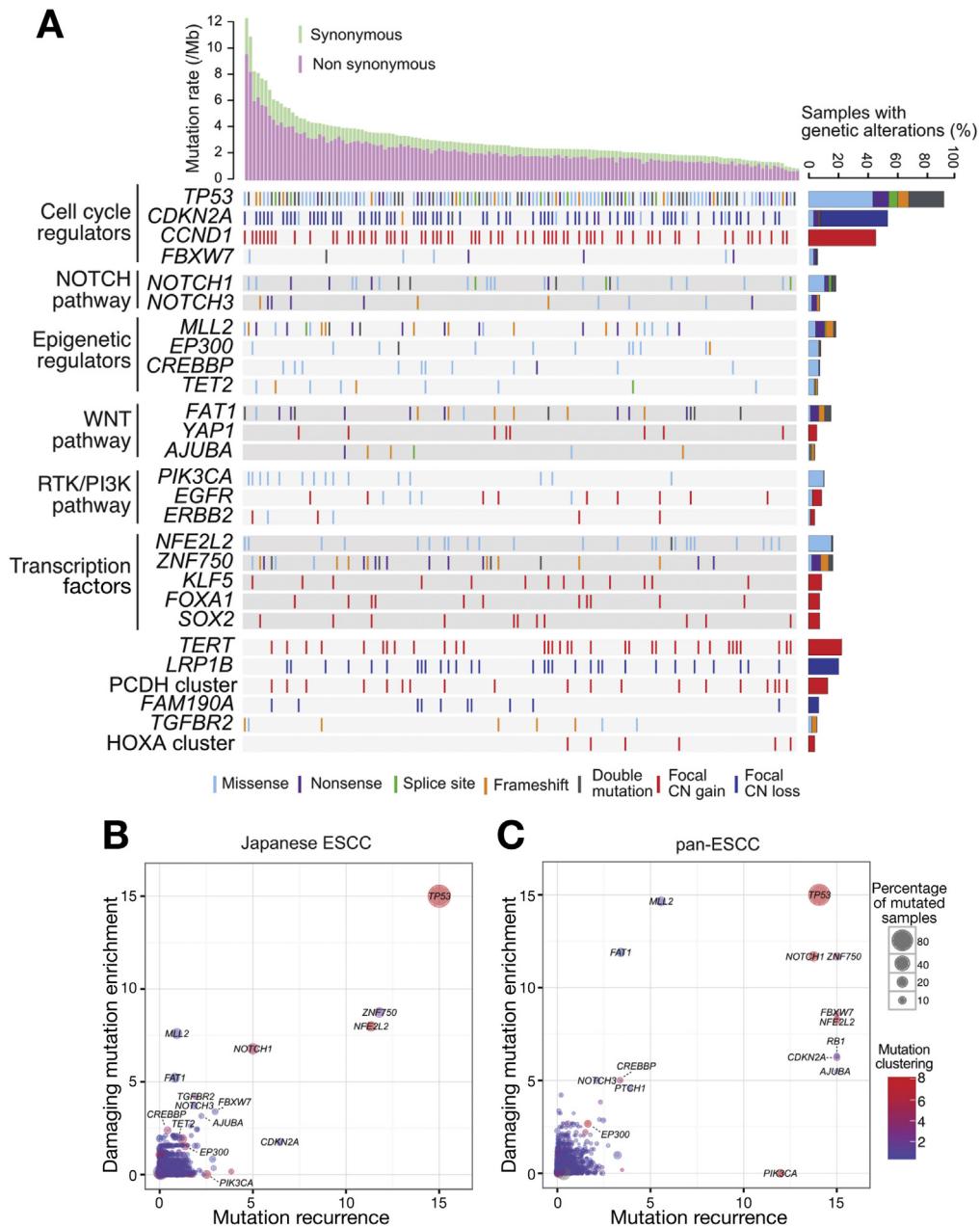


Figure 2. Identification of driver genes. (A) Landscape of genetic alterations across the 144 ESCC samples. The samples were sorted by mutation rates (top bars), while genes were sorted by the proportion of altered samples (left bars) and their functional categories. (B, C) Statistical evaluation of mutated genes for the 144 Japanese ESCC samples (B) and 449 pan-ESCC samples combined with Chinese high- and low-incidence ESCC samples (C). Each gene is represented as a dot, while their horizontal and vertical coordinates indicate negative log₁₀-scaled P values from 2 statistical tests evaluating mutation recurrence and enrichment of mutations that potentially damage protein functions, respectively. Sizes of dots represent the proportion of mutated samples, while colors of dots represent negative log₁₀-scaled P values from a statistical test to evaluate the presence of a mutational hot spot.

whereas neither of the signatures was prominent in cluster 2 (Figure 1D). Cluster 1 showed a relatively high mutation rate, which appeared to be explained by an increase in the APOBEC signature (Figure 1E). Similar clustering based on the mutation spectra was observed for high- and low-incidence ESCC in the Chinese and HNSCC samples (Supplementary Figures 4–8). These clusters were tightly associated with well-known environmental risk factors, including drinking and smoking, as well as genetic factors (Figure 1C, Supplementary Figure 2A, and Supplementary Table 5). Heavy drinkers with drinking-associated *ALDH2* risk alleles (rs671 AG/AA) were significantly enriched in cluster 1 and cluster 2 (odds ratio = 4.06; 95% CI: 1.91–8.89; $P = .000078$, Fisher's exact test) compared with cluster 3. The *ALDH2* SNP alone was also significantly enriched in cluster 1 and cluster 2 ($P = .00086$).

No significant association was observed between the development of drinking-associated ESCC and the presence of an *ADH1B* risk allele (rs1229984 GG).

Although we previously found a significant synergistic effect of drinking and smoking on ESCC risk,⁴ smoking alone was not significantly associated with one or more clusters of the mutation spectra. It has been reported that cytochrome *P450 2A6* (*CYP2A6*) is involved in the metabolic activation of tobacco carcinogens, and its null allele is associated with a lower risk for lung cancer development.²¹ Therefore, we statistically analyzed known polymorphisms of *CYP2A6* and found that smokers with a functional *CYP2A6* allele (*CYP2A6*1*) were significantly enriched in cluster 1 compared with other clusters (OR = 2.47; 95% CI: 1.08–5.73; $P = .021$). The enrichment was even more significant when smokers

with a functional *CYP2A6* allele had additional risk factors, including drinking behavior and the *ALDH2* risk alleles (rs671 AG/AA) (OR = 3.69; 95% CI: 1.52–9.06; $P = .0019$).

We also examined the correlation between drinking-related environmental/genetic factors and 3 mutation spectra clusters in the Chinese ESCC datasets, although we could not examine smoking-related factors, owing to the unavailability of *CYP2A6* genotype information (Supplementary Figures 4 and 5). Compared with Japanese ESCC, the proportion of drinking-related risk factors was generally low in Chinese ESCC, particularly in the high-incidence cohort, which is consistent with previous reports (Supplementary Figure 9).³ In Chinese low-incidence ESCC, we also found that the combined risk of drinking and the *ALDH2* SNP (rs671 AG/AA) was unevenly distributed across the 3 mutation spectra clusters. However, enrichment of the combined risk was observed in cluster 2 only, different from the case of Japanese ESCC. Additionally, in Chinese high-incidence ESCC, all of the few samples showing the combined risk were contained in cluster 2. This observation suggests that a small fraction of Chinese high-incidence ESCC is caused by the same alcohol-related mechanism as that of Chinese low-incidence ESCC.

Major Mutational Targets in Esophageal Squamous Cell Carcinoma

The 23,121 somatic mutations identified by WES of the 144 ESCC samples contained 17,189 nonsynonymous mutations in 10,552 genes (Supplementary Tables 3 and 6). Among these, we identified 15 significantly mutated genes after evaluating the recurrence and clustering of mutations, background mutation rates, and predicted effects of individual mutations on protein functions (Supplementary Material and Methods, Figure 2A and B and Supplementary Table 7). The most frequently mutated gene was *TP53* (mutated in 93.1% of our cohort), followed by *NOTCH1* (18.6%), *MLL2* (18.6%), *NFE2L2* (16.7%), *ZNF750* (16.7%), *FAT1* (14.6%), *PIK3CA* (10.4%), *EP300* (8.3%), *CDKN2A* (8.3%), *CREBBP* (7.6%), *NOTCH3* (7.6%), *TET2* (6.3%), *FBXW7* (5.6%), *TGFBR2* (5.6%), and *AJUBA* (4.2%). *TP53*, *NFE2L2*, *EP300*, *CREBBP*, and *NOTCH1* clearly contained mutational hot spots (Supplementary Figure 10). Except for *NFE2L2* and *PIK3CA*, most of these genes exhibited nonsense and frameshift mutations, suggesting their tumor suppressor roles. Examination of the Chinese datasets showed that most of the 15 genes harbored recurrent mutations, but at lower frequencies (Supplementary Figure 11A, B, and D). In contrast, a pan-ESCC analysis combining the Japanese and Chinese datasets revealed 2 additional genes that were mutated less frequently in our data: *RB1* (4.2%) and *PTCH1* (3.5%; Figure 2C). Our comparative analysis also confirmed that most of the significantly mutated genes in ESCC were also mutated in HNSCC, suggesting that the molecular pathogenesis is shared between the 2 squamous cell carcinomas (Supplementary Figure 11C and D).

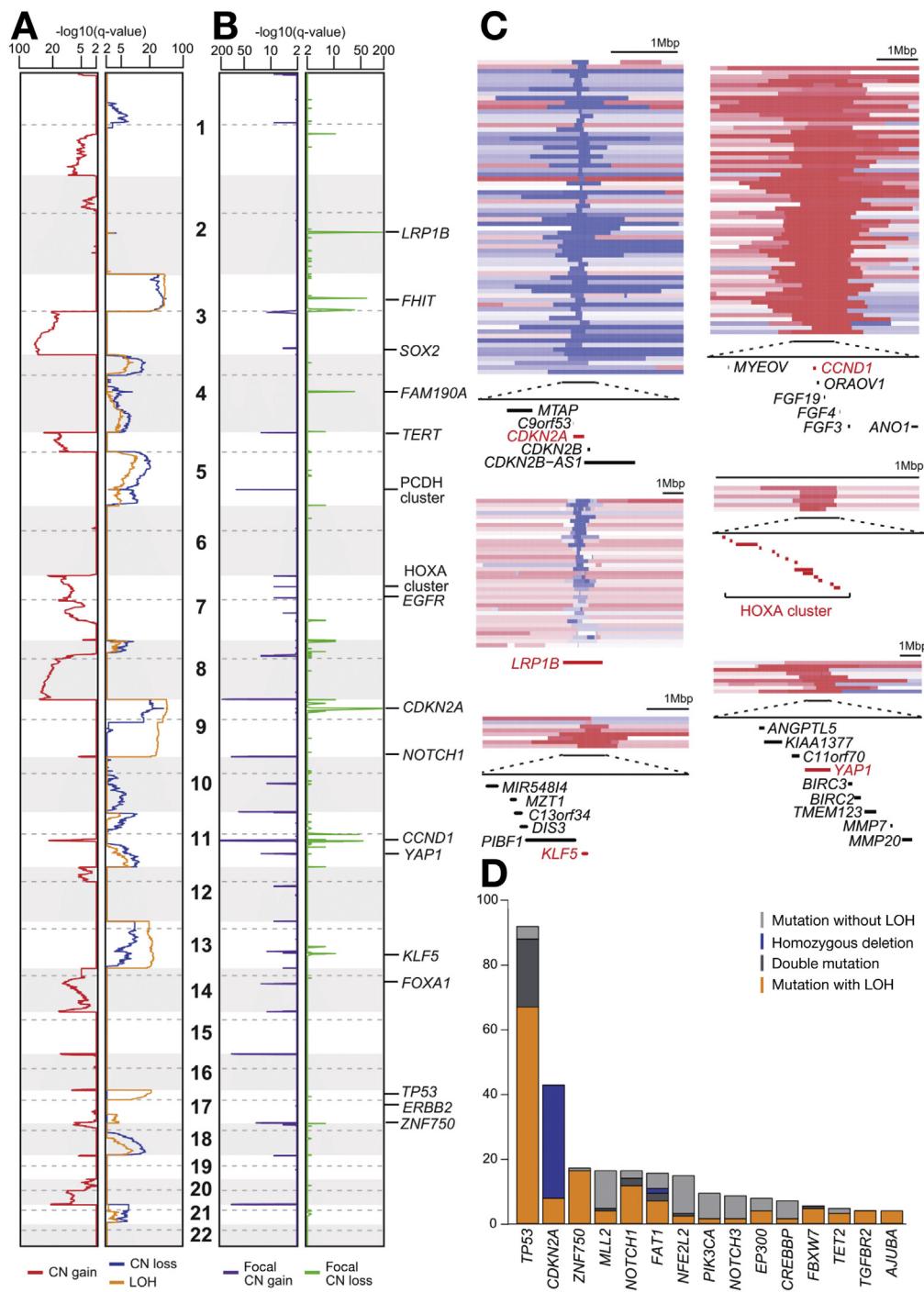
SNP array-based CN profiling to investigate recurrent CN variations was also performed for 123 of the 144 ESCC samples (Supplementary Material and Methods and

Supplementary Figure 12). Recurrent CN variations included amplifications of 3q (69.3%), 8q (64.3%), 5p (53.7%), 7p (47.2%), and 20q (42.3%); deletions of 3p (71.5%), 9p (56.1%), 4p (48.9%), 5q (48.0%), and 13p (47.2%); and loss of heterozygosity in 9p (97.6%), 3p (92.7%), 9q (83.7%), 13q (76.4%), 13p (73.2%), and 17p (68.3%) (Figure 3A).²² A number of candidate driver genes were identified from recurrent focal CN alterations (Figure 3B and Supplementary Figure 13). The genes most frequently affected were *CDKN2A/2B* (deleted in 47.9%) and *CCND1* (amplified in 46.5%), followed by *TERT* (amplified in 22.9%), *LRP1B* (deleted in 20.8%), *PCDH* cluster (amplified in 13.2%), *KLF5* (amplified in 9.0%), *FOXA1* (amplified in 7.6%), *FAM190A* (deleted in 7.6%), *EGFR* (amplified in 6.6%), *YAP1* (amplified in 5.6%), *HOXA* cluster (amplified in 4.2%), and *ERBB2* (amplified in 2.3%) (Figure 3C and Supplementary Figure 14). Notably, *CDKN2A*, *EGFR*, and *ERBB2* also harbored recurrent mutations.

By combining CN profiling and WES data, we found that *TP53*, *NOTCH1*, *ZNF750*, *FAT1*, *TET2*, *FBXW7*, and *CDKN2A* were frequently subjected to biallelic inactivation via double mutations, mutations accompanied by loss of heterozygosity or homozygous deletion (Figure 3D), suggesting that these genes have tumor suppressive roles. In an evaluation of the clinical impacts of these mutations/CN variations, we found that alterations in 4 of these genes were correlated with poor overall survival: *EP300* ($P = .000573$), *TET2* ($P = .0177$), *LRP1B* ($P = .00523$), and *KLF5* ($P = .0373$; Figure 4A). Notably, the *EP300* data were similar to that of low-incidence ESCC in the Chinese population.¹¹ We also examined the correlations between gene mutations and mutational signatures (Figure 4B). Consistent with the results of a previous study examining high-incidence ESCC in Chinese subjects,¹³ we found that samples with a higher APOBEC signature tended to contain 2 types of hot-spot *PIK3CA* mutations ($P = .0012$, Figure 4B); these mutations presumably resulted from a common mutational mechanism, as they are APOBEC signature mutations. Similarly, hot-spot frameshift mutations in *TGFBR2* tended to occur in samples with high short indel rates ($P = .0032$), as previously observed in a gastric cancer study.²³ Notably, *ZNF750* mutations were positively associated with the APOBEC signature ($P = .00030$), consistent with the Chinese ESCC and HNSCC datasets (Figure 4B, Supplementary Figure 15A). *ZNF750*, an epidermal differentiation regulator,²⁴ has been proposed to be a tumor suppressor gene of ESCC.^{10,13} This hypothesis was also supported by our data, which showed that most of the *ZNF750* mutations were null mutations accompanied by loss of heterozygosity (Figures 2A and 3D). This finding suggests that loss of function of *ZNF750* contributes to an unknown mutational mechanism generating the APOBEC signature.

Genetically Disrupted Pathways in Esophageal Squamous Cell Carcinoma

To obtain a more comprehensive understanding of the molecular lesions in Japanese ESCC, we searched for pathways genetically deregulated in our cohort by a literature review (Figure 5). Cell cycle regulators constituted the most



frequently disrupted category (97.9% of our cohort),²⁵ including mutations in *TP53* (93.1%), focal amplification of *CCND1* (46.5%), and mutation and focal deletion of *CDKN2A* (53.5%). Overall, epigenetic regulators comprised one of the largest gene categories in ESCC as well as other types of cancers (disrupted in 59.0% of our cohort).²⁶ Two histone acetyltransferase genes, *CREBBP* and *EP300*, were among the list of significantly mutated genes, in which mutations generally clustered within the HAT domains (Supplementary Figure 10). Multiple trithorax family genes were recurrently disrupted,²⁷ whereas few mutations

occurred in polycomb group complexes, suggesting an important role of repressive epigenetic marks in the pathogenesis of ESCC.

TET2 is involved in epigenetic regulation and is a newly identified mutational target in ESCC. It was mutated in 6.3% of the Japanese cohort. The TET family of proteins encodes α -ketoglutarate-dependent oxygenases that convert 5-methylcytosine to 5-hydroxymethylcytosine, which is the initial step of DNA demethylation.²⁸ *TET2* has been established as a tumor suppressor with prevalent mutations in hematologic malignancies, and *TET2* mutations have

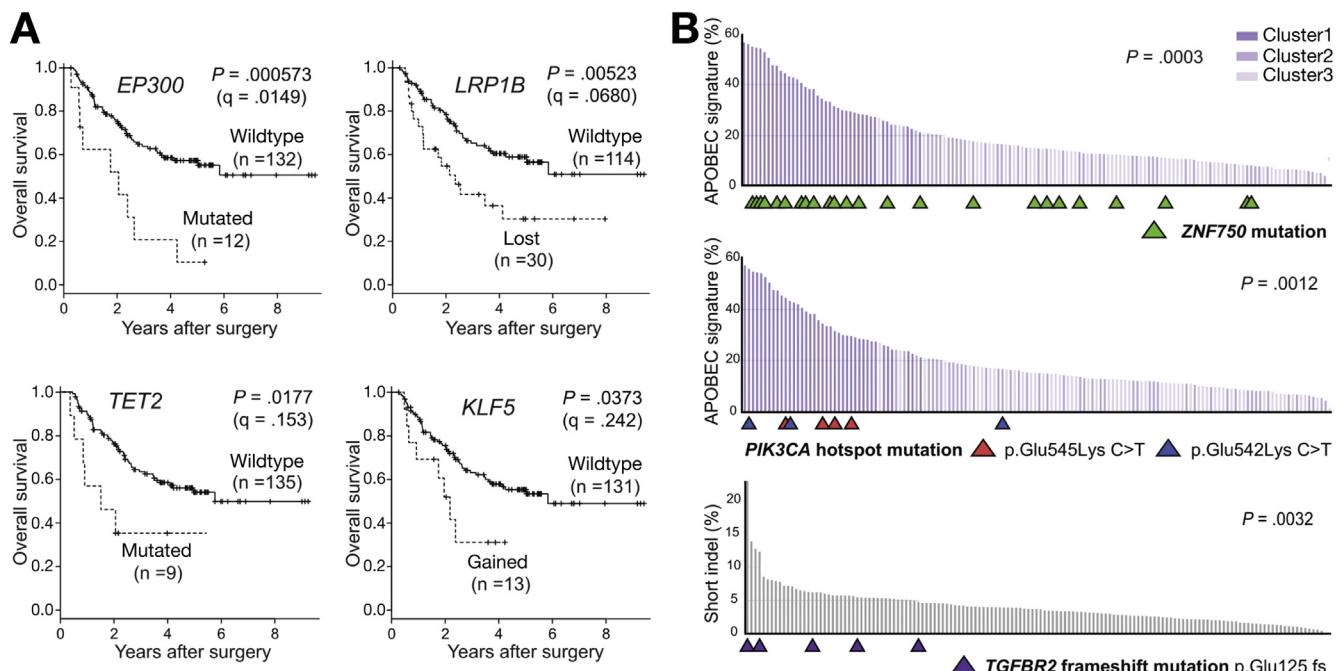


Figure 4. Association of genetic alterations with patient prognosis and mutational signatures. (A) Effect of genetic alterations on overall survival for 4 significant genes. *P* values were calculated by the log-rank test and the Benjamini-Hochberg *q* values were calculated to correct for multiple hypothesis testing performed for multiple genes listed in Figure 2A. (B) Association of mutations in 3 genes with the APOBEC signature and short indels. The 144 samples were sorted by percentages of each types of mutation. *P* values were calculated for the uneven distribution of mutated samples in the sorted list using the Wilcoxon rank sum test.

recently been detected in some solid cancers, such as clear cell renal carcinomas and colorectal cancers.^{29,30} The similarity in the distribution of *TET2* mutations in ESCC with other types of cancers (Figure 6A) suggests that these may be loss-of-function mutations. In fact, when transduced into HEK293FT cells, *TET2* mutants identified in our WES failed to increase cellular 5-hydroxymethylcytosine levels compared to wild-type *TET2* (Figure 6B). In our ESCC cohort, *TET2* mutations were associated with poor survival (*P* = .0177; Figure 4A). In accordance with this, in our in vitro invasion assay, *TET2* knockdown increased the invasive activity of an ESCC cell line (KYSE410) (Figure 6C–F). In addition, transduction of the wild-type *TET2* allele suppressed the increased invasive activity compared with the mutant alleles, suggesting a role for the *TET2* mutation in the invasive phenotype of ESCC.

The NRF2 pathway consisting of *NFE2L2*, *KEAP1*, and *CUL3* was genetically deregulated in 23.6% of our cohort (Figure 7A). *NFE2L2* (also known as *NRF2*) encodes a transcription factor that induces the cellular response to oxidative stress, whereas *KEAP1/CUL*-mediated ubiquitination degrades the *NFE2L2* protein under normal conditions.³¹ Oxidative stress plays important roles in the development of many cancers, including ESCC, and mutations in the NRF2 pathway have been observed in a wide variety of squamous cell carcinomas, including ESCC, lung squamous cell carcinoma, and skin squamous cell carcinoma.³² These mutations were suggested to contribute to tumor development by stabilizing the *NFE2L2* protein.³³ Nearly all *NFE2L2* mutations reportedly involve 2 hot

spots within or near the DLG and ETGE motifs that bind to *KEAP1*, which was also observed in our cohort (Figure 7B). To assess the functional consequence of *NFE2L2* mutations, we predicted the set of genes transcriptionally regulated by *NFE2L2* in ESCC using transcriptome data from a set of 70 ESCC samples (Supplementary Material and Methods and Supplementary Figure 16). We then used a set of *NFE2L2*-regulated genes previously determined in an in vitro experiment³⁴ to extract a coherent subset of genes that were significantly co-expressed in the 70 ESCC samples (Figure 7C).³⁵ As expected, the extracted gene subset, or the “*NFE2L2* module,” was significantly enriched for oxidoreductase activity-associated genes (*P* < 10⁻¹²), such as cytochrome P450, NAD(P)H dehydrogenases, and DP-glucose 6-dehydrogenases (Supplementary Table 9). Accordingly, we found that the mean expression level of the *NFE2L2* module genes, or “*NFE2L2* module activity,” was significantly higher in *NFE2L2*-mutated samples than in wild-type samples (*P* = 1.29 × 10⁻⁷; Figure 7C and D). In addition, a multiple linear regression analysis showed that the *NFE2L2* expression level was significantly correlated with *NFE2L2* module activity in ESCC, regardless of mutation status. Notably, a similar correlation between the mutation/expression status of *NFE2L2* and *NFE2L2* module activity was confirmed in lung squamous cell carcinoma (Supplementary Figure 17).

The NOTCH pathway is also a frequent mutational target in ESCC as well as several other cancer types.³⁶ Combined, *NOTCH1/3* receptors, their ligands (*DLL1*, *DLK1*, and *JAG2*), and downstream regulators (*MAML3*, *NCOR1/2*, and *SPEN*)

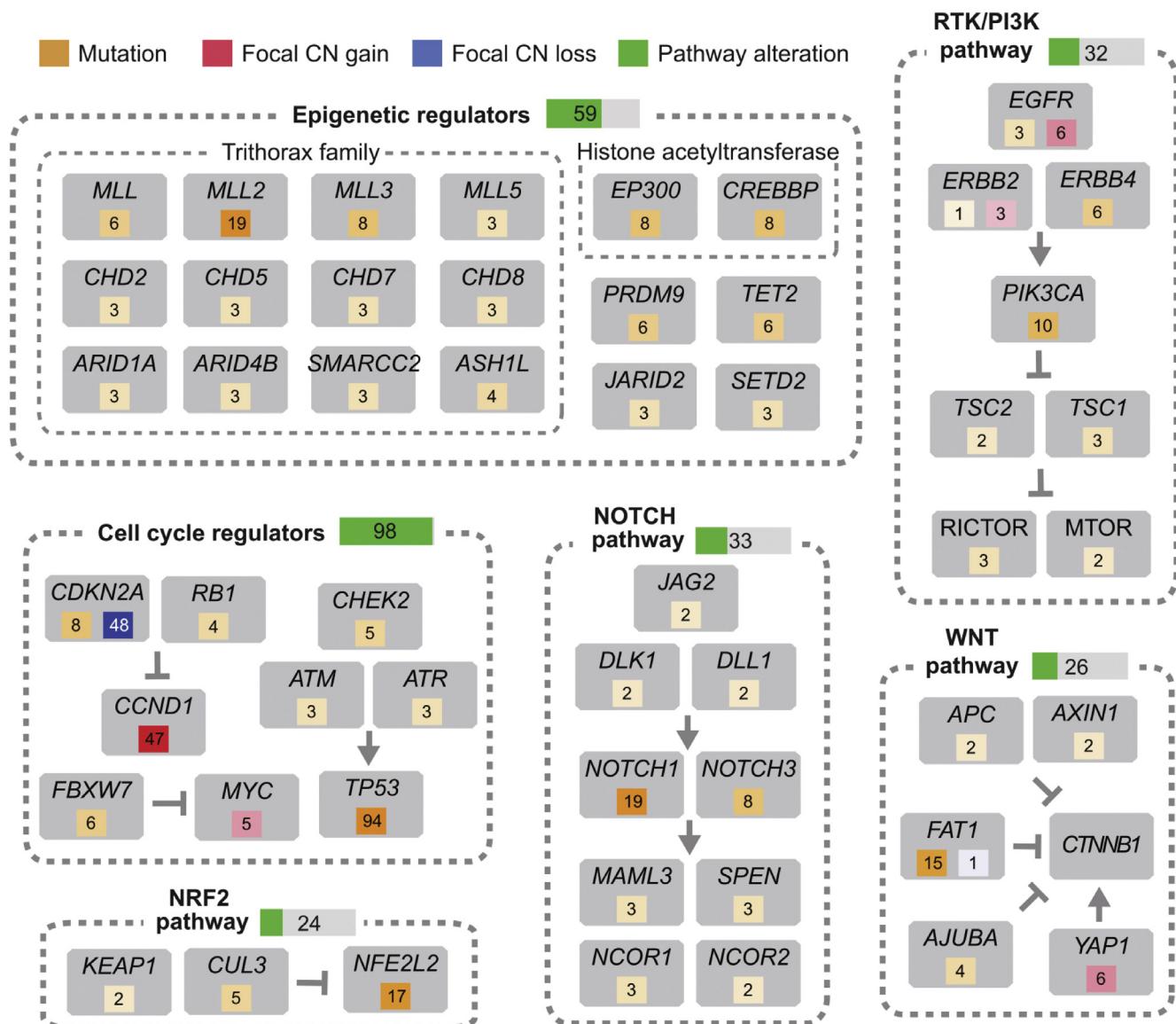


Figure 5. Pathway-level view of genetic alterations. Positive regulation and negative regulation based on a literature review are indicated as *lines with arrowheads* and *blunt ends*, respectively. For each gene, the percentages of altered samples are provided as numerals in colored boxes (mutations: orange; focal CN gains: red; focal CN losses: blue). Color strength corresponds to the percentage. Green bars with numerals indicate the percentage of altered genes in each pathway.

were disrupted in 33.4% of ESCC cases. The *NOTCH1* mutations in ESCC were clearly clustered within epidermal growth factor-like repeats 11–12, which are involved in ligand binding. This mutational hot-spot region is distinct from the HD-L and PEST domains harboring well-known oncogenic mutations in T-lymphoblastic leukemia, but similar to the region harboring presumably tumor-suppressive mutations in other types of squamous carcinomas^{18,20} (Supplementary Figure 10). The other frequently disrupted pathways include the receptor-tyrosine kinase–phosphoinositide 3-kinase (RTK/PI3K) and WNT pathways (disrupted in 31.9% and 26.3% of our cohort, respectively). Several receptor tyrosine kinases (*EGFR*, *ERBB2*, and *ERBB4*) and their downstream signal transducers (*PIK3CA*, *TSC1/2*, *RICTOR*, and *MTOR*) are mutational targets in the RTK/PI3K pathway.^{37,38} Although *CTNNB1*

was not itself affected in the WNT pathway, we found recurrent mutations in *FAT1* and *AJUBA*, which regulate *CTNNB1* stability^{39,40}; *YAP1*, which acts cooperatively with *CTNNB1*⁴¹; *APC*; and *AXIN1*. By comparing these results with published data,^{11–13} we confirmed that mutational landscapes at the pathway level were generally conserved across the different ESCC cohorts (Supplementary Figure 11E).

Discussion

In this study, we determined the landscape of genetic lesions in 144 Japanese ESCC cases by WES and SNP array profiling. Importantly, we identified 3 mutation spectra clusters characterized by different proportions of CpG and APOBEC signatures. Furthermore, these clusters were associated not only with drinking and smoking, but also

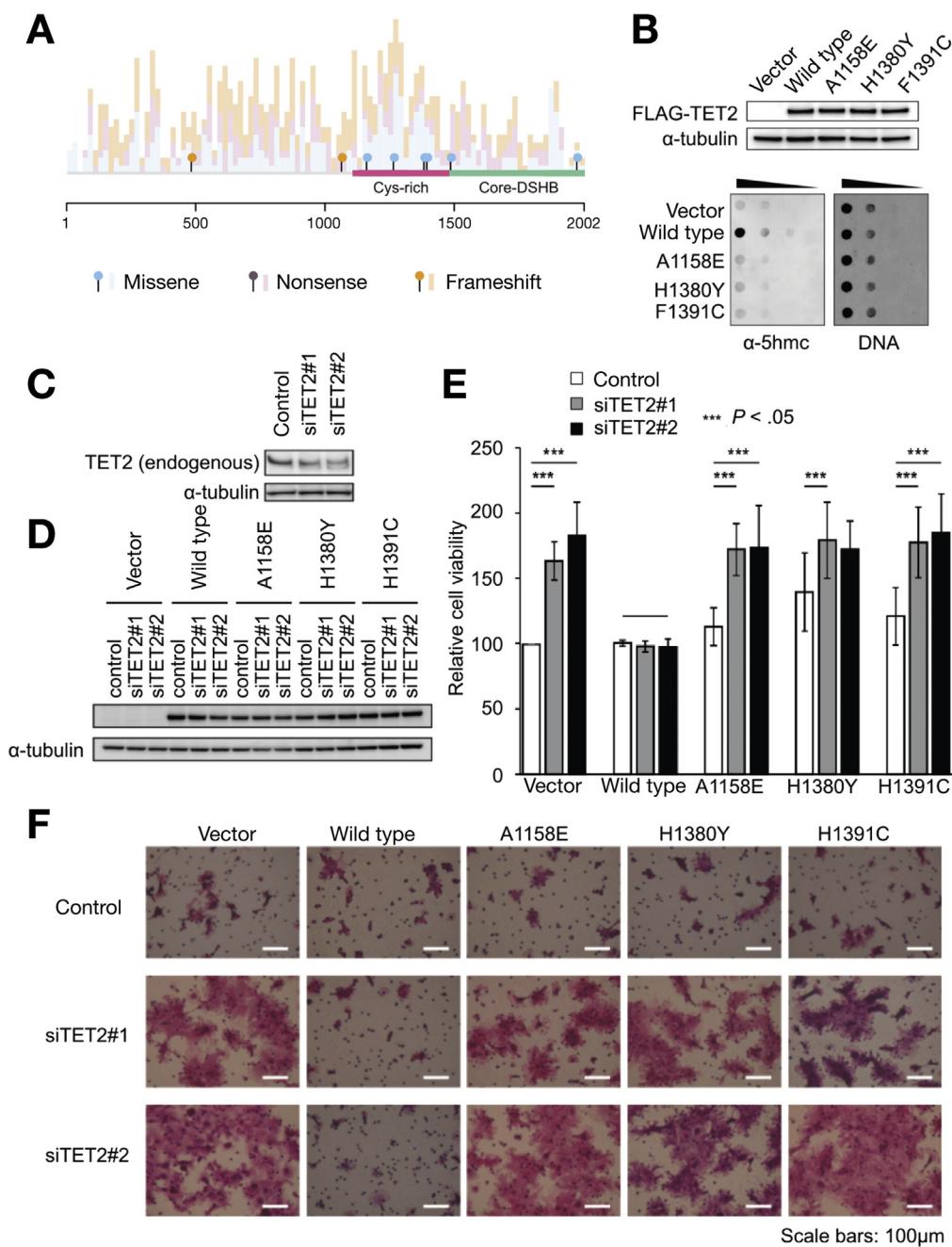


Figure 6. Analysis of *TET2* mutations. (A) Mutational types and positions on the *TET2* proteins. Pins on the horizontal bar indicating the *TET2* protein denote mutations identified by our WES. Colored histograms in the background show distributions of mutational positions in the COSMIC pan-cancer dataset. (B) Evaluation of the expression levels of FLAG-tagged wild-type or mutant *TET2* in KYSE410 cells and dot blot analysis of 5-hydroxymethylcytosine (5hmC) in KYSE410 cells in which FLAG-tagged wild-type or mutant *TET2* was overexpressed. (C) Immunoblot of *TET2* in KYSE410 cells in which *TET2* was knocked down by small interfering RNA (siRNA) targeting *TET2*. (D) Immunoblot of FLAG-tagged *TET2* in KYSE410 cells in which *TET2* was knocked down by siRNA targeting *TET2* and/or the indicated FLAG-tagged *TET2* variant was overexpressed. α -Tubulin was used as a control. (E) Invasion analysis of KYSE410 cells. Viability of the cells that traversed the Matrigel to the lower surface was measured using CellTiter-Glo. Note that ectopic overexpression of wild-type, but not that of a catalytically inactive mutant, *TET2* suppressed the invasion of KYSE410 cells in which *TET2* was knocked down. (F) H&E staining of the invasive cells in panel (E).

with germline polymorphisms in *ALDH2* and *CYP2A6*, which affect alcohol and tobacco metabolism. We also confirmed that Chinese high- and low-incidence ESCC cohorts showed similar mutation spectra clusters, which are associated with drinking and the *ALDH2* SNP; however, the association

pattern differed from that observed in our cohort. Substantial differences in epidemiologic characteristics have been reported between Japanese and Chinese ESCC³ and our unique observations in Japanese ESCC may reflect a strong association between drinking and smoking behaviors and

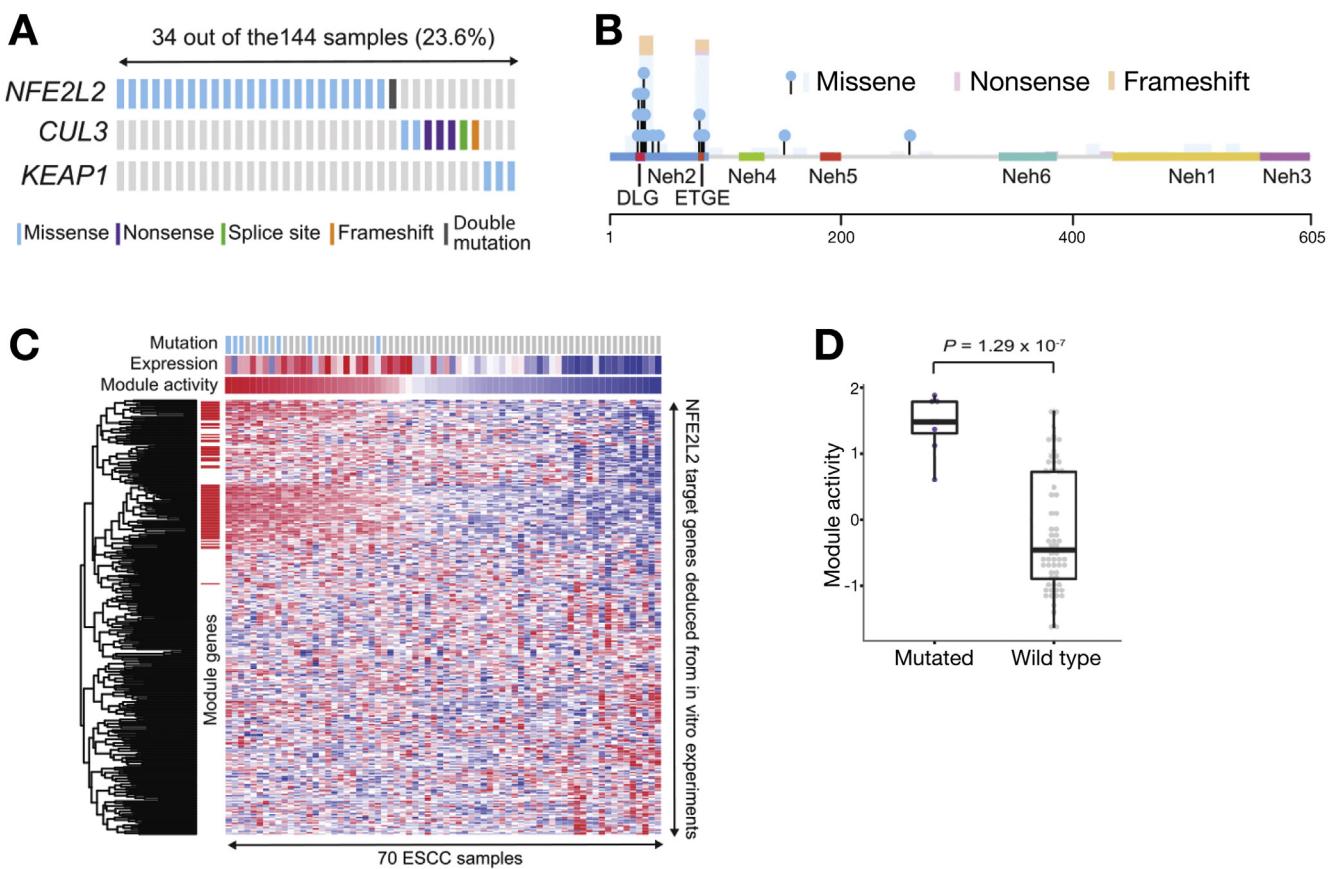


Figure 7. Analysis of *NFE2L2* mutations. (A) Genomic alterations in the NRF2 pathway. (B) Mutational types and positions on the *NFE2L2* protein are the same as in Figure 6A. (C) The *NFE2L2* module in ESCC. For *NFE2L2* target genes prepared from in vitro experiments, expression profiles in 70 ESCC samples are shown as a clustered heat map. A coherent subset identified by extraction of expression modules (EEM)³⁵ and their mean expression profile were obtained as the *NFE2L2* module and module activity, respectively. Note that color bars on the top of the heat map show the *NFE2L2* mutation status. The mutations status and messenger RNA expression level were additively correlated with module activity; the multiple regression coefficients were 0.89 ($P < .001$) and 0.53 ($P < .001$) for mutation and expression, respectively. (D) A box plot of the module activity for samples with or without a mutation.

ESCC incidence in Japan.^{4,5} In addition, our analysis revealed that APOBEC signatures are correlated with mutations in a potential tumor suppressor, *ZNF750*. Our data suggest that ESCC cancer genomes are shaped by complex mutational mechanisms that vary among individual patients, although the details of these mechanisms remain unclear.

By analyzing a large number of ESCC samples, we also unveiled the registry of driver genes somatically disrupted in Japanese ESCC. These genes are dispersed across a number of known cancer-associated pathways, including cell cycle and epigenetic regulators, and the NOTCH, NRF2, WNT, and RTK/PI3K pathways. Biallelic inactivation of *TP53* was observed in most of the samples, suggesting an essential role in tumor initiation. *NOTCH1* is frequently disrupted by loss-of-function hot-spot mutations, supporting that the loss of NOTCH pathway activity is critical for the growth of tumor cells with squamous differentiation characteristics.³⁶ *TET2* also showed recurrent loss-of-function mutations associated with poor patient prognosis. Our results demonstrated that the loss-of-function of *TET2* contributes to the malignant phenotypes of ESCC by enhancing cell invasiveness, although the detailed mechanisms require

further analysis. Notably, *TET1*, a gene belonging to the same family, reportedly suppresses cancer invasion by activating metalloproteinase inhibitors.⁴² Our informatics analysis showed that gain-of-function *NFE2L2* mutations transcriptionally activate oxidoreductase-related genes to oppose oxidative stress, which may play an important role not only in tumor development, but also in the acquisition of chemoradiotherapy resistance.⁴³ Our CN analysis revealed a number of novel driver genes, including *LRP1B*, *KLF5*, and *YAP1*, which frequently underwent focal CN alterations. Notably, *EGFR*, *ERBB2*, *PIK3CA*, and other RTK/PI3K pathway components exhibited recurrent focal amplifications and/or mutations, indicating that they are potential targets of molecular-based therapies.^{37,38} Our data also suggested that although transducers of canonical WNT signaling, such as *APC* and *CTNNB1*, were not major mutational targets, focal amplifications of *YAP1* and loss-of-function mutations in *FAT1* and *AJUBA* may contribute to the activation of WNT signaling in ESCC. A comparison with published Chinese ESCC datasets suggested that driver genes and pathways were overall conserved across cohorts. In summary, this study characterized the genomic landscape

of Japanese ESCC, which increases the understanding of the molecular pathophysiology of ESCC and may be useful for future therapy development and precision medicine.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at <http://dx.doi.org/10.1053/j.gastro.2016.01.035>.

References

- Enzinger PC, Mayer RJ. Esophageal cancer. *N Engl J Med* 2003;349:2241–2252.
- Shimada H, Kitabayashi H, Nabeya Y, et al. Treatment response and prognosis of patients after recurrence of esophageal cancer. *Surgery* 2003;133:24–31.
- Lin Y, Totsuka Y, He Y, et al. Epidemiology of esophageal cancer in Japan and China. *J Epidemiol* 2013;23:233–242.
- Tanaka F, Yamamoto K, Suzuki S, et al. Strong interaction between the effects of alcohol consumption and smoking on oesophageal squamous cell carcinoma among individuals with *ADH1B* and/or *ALDH2* risk alleles. *Gut* 2010;59:1457–1464.
- Cui R, Kamatani Y, Takahashi A, et al. Functional variants in *ADH1B* and *ALDH2* coupled with alcohol and smoking synergistically enhance esophageal cancer risk. *Gastroenterology* 2009;137:1768–1775.
- Gao Y, Hu N, Han X, et al. Family history of cancer and risk for esophageal and gastric cancer in Shanxi, China. *BMC Cancer* 2009;9:269.
- Tran GD, Sun XD, Abnet CC, et al. Prospective study of risk factors for esophageal and gastric cancers in the Linxian general population trial cohort in China. *Int J Cancer* 2005;113:456–463.
- Yokokawa Y, Ohta S, Hou J, et al. Ecological study on the risks of esophageal cancer in Ci-Xian, China: the importance of nutritional status and the use of well water. *Int J Cancer* 1999;83:620–624.
- Agrawal N, Jiao Y, Bettegowda C, et al. Comparative genomic analysis of esophageal adenocarcinoma and squamous cell carcinoma. *Cancer Discov* 2012;2:899–905.
- Lin DC, Hao JJ, Nagata Y, et al. Genomic and molecular characterization of esophageal squamous cell carcinoma. *Nat Genet* 2014;46:467–473.
- Gao YB, Chen ZL, Li JG, et al. Genetic landscape of esophageal squamous cell carcinoma. *Nat Genet* 2014;46:1097–1102.
- Song Y, Li L, Ou Y, et al. Identification of genomic alterations in oesophageal squamous cell cancer. *Nature* 2014;509:91–95.
- Zhang L, Zhou Y, Cheng C, et al. Genomic analyses reveal mutational signatures and frequently altered genes in esophageal squamous cell carcinoma. *Am J Hum Genet* 2015;96:597–611.
- Shiraishi Y, Sato Y, Chiba K, et al. An empirical Bayesian framework for somatic mutation detection from cancer genome sequencing data. *Nucleic Acids Res* 2013;41:e89.
- Van Loo P, Nordgard SH, Lingjaerde OC, et al. Allele-specific copy number analysis of tumors. *Proc Natl Acad Sci U S A* 2010;107:16910–16915.
- Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. *Nature* 2013;500:415–421.
- Roberts SA, Lawrence MS, Klimczak LJ, et al. An APO-BEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat Genet* 2013;45:970–976.
- Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* 2015;517:576–582.
- Comprehensive genomic characterization of squamous cell lung cancers. *Nature* 2012;489:519–525.
- Dulak AM, Stojanov P, Peng S, et al. Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nat Genet* 2013;45:478–486.
- Fujieda M, Yamazaki H, Saito T, et al. Evaluation of *CYP2A6* genetic polymorphisms as determinants of smoking behavior and tobacco-related lung cancer risk in male Japanese smokers. *Carcinogenesis* 2004;25:2451–2458.
- Niida A, Imoto S, Shimamura T, et al. Statistical model-based testing to evaluate the recurrence of genomic aberrations. *Bioinformatics* 2012;28:i115–i120.
- Wang K, Yuen ST, Xu J, et al. Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. *Nat Genet* 2014;46:573–582.
- Sen GL, Boxer LD, Webster DE, et al. *ZNF750* is a p63 target gene that induces *KLF4* to drive terminal epidermal differentiation. *Dev Cell* 2012;22:669–677.
- Evan GI, Vousden KH. Proliferation, cell cycle and apoptosis in cancer. *Nature* 2001;411:342–348.
- Shen H, Laird PW. Interplay between the cancer genome and epigenome. *Cell* 2013;153:38–55.
- Schuetzengruber B, Martinez AM, Iovino N, et al. Trithorax group proteins: switching genes on and keeping them active. *Nat Rev Mol Cell Biol* 2011;12:799–814.
- Solary E, Bernard OA, Tefferi A, et al. The Ten-Eleven Translocation-2 (*TET2*) gene in hematopoiesis and hematopoietic diseases. *Leukemia* 2014;28:485–496.
- Sato Y, Yoshizato T, Shiraishi Y, et al. Integrated molecular analysis of clear-cell renal cell carcinoma. *Nat Genet* 2013;45:860–867.
- Seshagiri S, Stawiski EW, Durinck S, et al. Recurrent R-spondin fusions in colon cancer. *Nature* 2012;488:660–664.
- Shibata T, Ohta T, Tong KI, et al. Cancer related mutations in *NRF2* impair its recognition by Keap1-Cul3 E3 ligase and promote malignancy. *Proc Natl Acad Sci U S A* 2008;105:13568–13573.
- Kim YR, Oh JE, Kim MS, et al. Oncogenic *NRF2* mutations in squamous cell carcinomas of oesophagus and skin. *J Pathol* 2010;220:446–451.
- Sporn MB, Liby KT. *NRF2* and cancer: the good, the bad and the importance of context. *Nat Rev Cancer* 2012;12:564–571.
- Malhotra D, Portales-Casamar E, Singh A, et al. Global mapping of binding sites for Nrf2 identifies novel targets in cell survival response through ChIP-Seq profiling and network analysis. *Nucleic Acids Res* 2010;38:5718–5734.

35. Niida A, Smith AD, Imoto S, et al. Gene set-based module discovery in the breast cancer transcriptome. *BMC Bioinformatics* 2009;10:71.
36. Ntziachristos P, Lim JS, Sage J, et al. From fly wings to targeted cancer therapies: a centennial for notch signaling. *Cancer Cell* 2014;25:318–334.
37. Hynes NE, Lane HA. ERBB receptors and cancer: the complexity of targeted inhibitors. *Nat Rev Cancer* 2005; 5:341–354.
38. Thorpe LM, Yuzugullu H, Zhao JJ. PI3K in cancer: divergent roles of isoforms, modes of activation and therapeutic targeting. *Nat Rev Cancer* 2015;15:7–24.
39. Morris LG, Kaufman AM, Gong Y, et al. Recurrent somatic mutation of *FAT1* in multiple human cancers leads to aberrant Wnt activation. *Nat Genet* 2013; 45:253–261.
40. Haraguchi K, Ohsugi M, Abe Y, et al. Ajuba negatively regulates the Wnt signaling pathway by promoting GSK-3beta-mediated phosphorylation of beta-catenin. *Oncogene* 2008;27:274–284.
41. Rosenbluh J, Nijhawan D, Cox AG, et al. beta-Catenin-driven cancers require a YAP1 transcriptional complex for survival and tumorigenesis. *Cell* 2012;151:1457–1473.
42. Hsu CH, Peng KL, Kang ML, et al. TET1 suppresses cancer invasion by activating the tissue inhibitors of metalloproteinases. *Cell Rep* 2012;2:568–579.
43. Shibata T, Kokubu A, Saito S, et al. *NRF2* mutation confers malignant potential and resistance to chemo-radiation therapy in advanced esophageal squamous cancer. *Neoplasia* 2011;13:864–873.

Author names in bold designate shared co-first authorship.

Received January 27, 2015. Accepted January 20, 2016.

Reprint requests

Address requests for reprints to: Koshi Mimori, MD, PhD, Department of Surgery, Beppu Hospital, Kyushu University, 4546, Tsurumi-hara, Beppu 874-0838, Japan. e-mail: kmimori@tsurumi.beppu.kyushu-u.ac.jp; fax: (81) 977-27-1651; or Masaki Mori MD, PhD, FACS, Department of Gastroenterological Surgery, Graduate School of Medicine, Osaka University, Yamadaoka 2-2, Suita 565-0871, Japan. e-mail: mmori@gesurg.med.osaka-u.ac.jp; fax: (81) 6-6879-3259.

Acknowledgments

This research used computational resources of the K computer provided by the RIKEN Advanced Institute for Computational Science through the HPCI System Research project (Project ID: hp140230). Computational time was also provided by the Supercomputer System, Human Genome Center, Institute of Medical Science, University of Tokyo. The authors would like to thank Prof Keitaro Matsuo for the insightful discussions, Prof Kay Huebner for help in editing the manuscript, and Ms F. Todokoro, Ms K. Imamura, Ms T. Horuchi, and Ms Y. Ishikawa for providing technical assistance with the next-generation sequencing. In addition, the authors would like to thank Ms T. Shimo-oka, Ms M. Kasagi, Ms T. Kohno, Ms K. Oda, Ms M. Aoyagi, and Ms T. Kawano for providing assistance with molecular biology procedures.

Satoru Miyano, Masaki Mori, Koshi Mimori jointly directed this work.

Accession codes: Raw messenger RNA expression data for 70 ESCC samples and array-CGH data for 40 ESCC samples have been deposited at the Gene Expression Omnibus under accession numbers GSE47404 and GSE47630.

Conflicts of interest

The authors disclose no conflicts.

Funding

This work was supported by the following grants and foundations: Grants-in-Aid for Scientific Research, grant numbers 25430111, 25461953, 25861199, 25861200, 24592005, and 21229015; Funding Program for Next Generation World-Leading Researchers (LS094); Grants-in-Aid for Scientific Research on Innovative Areas of MEXT “Systems Cancer Research” (4201); The MEXT Strategic Programs on Innovative Research “Supercomputational Life Science.”