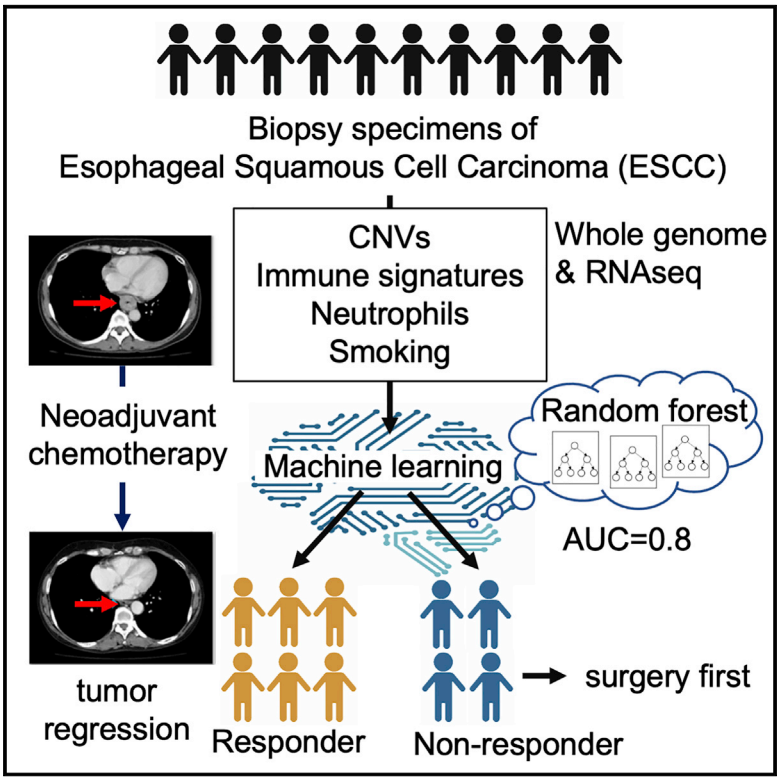
Cell Reports
Medicine

Article

Immuno-genomic profiling of biopsy specimens predicts neoadjuvant chemotherapy response in esophageal squamous cell carcinoma

Graphical abstract



Authors

Shota Sasagawa, Hiroaki Kato, Koji Nagaoka, ..., Kazuhiro Kakimi, Takushi Yasuda, Hidewaki Nakagawa

Correspondence

hidewaki@riken.jp

In brief

Sasagawa et al. show immunogenomic profiles of esophageal cancer biopsy specimens before chemotherapy and suggest interactions between tumor copy-number variants and immunity related with chemotherapy response. Neutrophils infiltration plays an important role in the response to chemotherapy. Machine learning using these immunogenomic large data can predict chemotherapy response.

Highlights

- Four different immune subtypes from RNA-seq of ESCC biopsy specimen
- Neutrophils within tumors are associated with tumor sensitivity to NAC
- Specific copy-number changes and signatures in ESCC are associated with NAC response
- Machine learning prediction for NAC response using immunogenomics of ESCC is useful

Sasagawa et al., 2022, Cell Reports Medicine *3*, 100705
August 16, 2022
https://doi.org/10.1016/j.xcrm.2022.100705

CellPress

Cell Reports Medicine

CellPress
OPEN ACCESS

Article

Immuno-genomic profiling of biopsy specimens predicts neoadjuvant chemotherapy response in esophageal squamous cell carcinoma

Shota Sasagawa,[1] Hiroaki Kato,[2] Koji Nagaoka,[3] Changbo Sun,[3] Motohiro Imano,[2] Takao Sato,[4] Todd A. Johnson,[1] Masashi Fujita,[1] Kazuhiro Maejima,[1] Yuki Okawa,[1] Kazuhiro Kakimi,[3] Takushi Yasuda,[2] and Hidewaki Nakagawa[1,5,*]

[1]Laboratory for Cancer Genomics, RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045, Japan
[2]Department of Surgery, Graduate School of Medicine, Kindai University, Osaka 577-8502, Japan
[3]Department of Immuno-therapeutics, The University of Tokyo Hospital, Tokyo 113-8655, Japan
[4]Department of Pathology, Kindai University Faculty of Medicine, Osaka 577-8502, Japan
[5]Lead contact
*Correspondence: hidewaki@riken.jp
https://doi.org/10.1016/j.xcrm.2022.100705

SUMMARY

Esophageal squamous cell carcinoma (ESCC) is one of the most aggressive cancers and is primarily treated with platinum-based neoadjuvant chemotherapy (NAC). Some ESCCs respond well to NAC. However, biomarkers to predict NAC sensitivity and their response mechanism in ESCC remain unclear. We perform whole-genome sequencing and RNA sequencing analysis of 141 ESCC biopsy specimens before NAC treatment to generate a machine-learning-based diagnostic model to predict NAC reactivity in ESCC and analyzed the association between immunogenomic features and NAC response. Neutrophil infiltration may play an important role in ESCC response to NAC. We also demonstrate that specific copy-number alterations and copy-number signatures in the ESCC genome are significantly associated with NAC response. The interactions between the tumor genome and immune features of ESCC are likely to be a good indicator of therapeutic capability and a therapeutic target for ESCC, and machine learning prediction for NAC response is useful.

INTRODUCTION

Esophageal cancer is one of the most aggressive cancers and the sixth leading cause of cancer-related deaths worldwide.[1] There are two histological types of esophageal cancer: esophageal squamous cell carcinoma (ESCC), which occurs in the middle or upper one-third of the esophagus, and esophageal adenocarcinoma (EAC), which occurs in the lower one-third of the esophagus or at the junction of the esophagus and stomach.[2,3] These two types differ in several aspects, including mortality, clinical features, and etiology. ESCC epidemiology is characterized by a wide variation in the incidence of ESCC among countries and ethnic groups. In Asia and Japan, 90% of esophageal cancers are ESCC, and its etiology is strongly associated with smoking and alcohol intake, while EAC predominates in the Western countries, and its etiology is related to gastroesophageal reflux.[4] The standard treatment for locally advanced and resectable ESCC is platinum-based neoadjuvant chemotherapy (NAC) in Japan[5,6] and neoadjuvant chemoradiation (NACR) in the United States and Europe. Some recent clinical studies have observed a similar treatment effect between NAC and NACR.[5,7] These pre-surgical treatments have been reported to have greater clinical efficacy than surgery alone in patients with resectable esophageal cancer.[8,9] The response rate of ESCC to NAC/NACR is 57%–72%, some of which achieve complete remission. Despite improvements in these pre-surgery treatments and surgical techniques, ESCC has a very low 5-year survival rate of 15%–25%.[10] Some patients cannot undergo surgery or curative resection because of disease progression during and after NAC.[11] Hence, predictive markers for the response to NAC are required.

Several candidate markers for ESCC tissues have been reported to predict the response to NAC, including p53, cell cycle regulators, nucleotide excision repair pathway, and some inflammatory and nutritional biomarkers, such as neutrophil-lymphocyte ratio (NLR) and prognostic nutritional index, have also been reported to predict NAC responsiveness. However, their prediction performance has not yet been validated,[12–16] and biomarkers to predict NAC sensitivity and their mechanism in ESCC remain unclear.

The tumor microenvironment (TME) consists of various cellular and non-cellular components, such as tumor tissue, normal tissues, and immune cells in the surrounding area, and the interactions between tumor cells and immune cells have been known to play a major role in tumor progression and response to chemotherapy or radiation therapies.[17,18] NAC has historically been considered immunosuppressive for cancer patients; however, certain chemotherapeutic agents, such as paclitaxel, cisplatin, gemcitabine, and carboplatin, are known to regulate and modulate anti-tumor immune responses.[19–22] Chemotherapy can

Table 1. Clinical information for ESCC patients (n = 121)

	Responder		Non-responder	
	Complete response	Partial response	Stable disease	Progressive disease
Patients, n	8	67	36	10
Gender				
Male, n (%)	8 (100)	51 (76)	30 (83)	8 (80)
Female, n (%)	0 (0)	16 (24)	6 (17)	2 (20)
Age (years), median (range)	66 (60–71)	69 (64–74)	68 (64–72)	68 (56–74)
Clinical stage				
I, n (%)	0 (0)	0 (0)	1 (2.8)	0 (0)
II, n (%)	2 (25)	7 (10)	5 (14)	1 (10)
III, n (%)	5 (62)	47 (70)	24 (67)	8 (80)
IV, n (%)	1 (12)	13 (19)	6 (17)	1 (10)
Smoking history				
Non-smoker, n (%)	0 (0)	15 (22)	5 (14)	1 (10)
Smoker, n (%)	2 (25)	25 (37)	19 (53)	3 (30)
Ex-smoker, n (%)	6 (75)	27 (40)	12 (33)	6 (60)
Brinkman index, median (range)	750 (596–908)	700 (278–1,132)	910 (485–1,200)	590 (320–772)
Alchohol drinking history				
Yes	8 (100%)	60 (90%)	32 (89%)	8 (80%)

induce immunogenic cell death and the subsequent release of tumor-associated neoantigens, which triggers immune activation. These neoantigens activate antigen-presenting cells, such as tumor-associated macrophages and dendritic cells, through Toll-like receptors.[23–26]

Here, we demonstrate a hierarchical approach to the comparative analysis of genomes of pre-treatment biopsies from 121 ESCC responders and non-responders, starting with immune signature, copy-number variation, and copy-number signature, followed by a machine learning approach to predict the response to NAC by combining several immunogenomic and clinical information.

RESULTS

Clinical characteristics and response assessment

We performed comprehensive genomic and transcriptomic tumor analysis for 143 ESCC patients as part of the University Hospital Medical Information Network Clinical Trials Registry of Japan (identification no. UMIN000004555/000004616; Figure S1, Table S1). Unless there was a special reason, NAC was indicated and treated for NAC according to the guidelines for the Japanese Esophageal Association. Fresh-frozen pre-treatment core tumor biopsies were collected from 121 patients with platinum-based NAC. DNA and RNA were extracted and profiled by deep whole-genome sequencing (dWGS, ×30∼; 20 cases), shallow whole-genome sequencing (sWGS, ×1∼; 72 cases) and RNA sequencing (RNA-seq) (79 samples). Most patients were male (73.9%), and the median age was 67 years (range, 41–82 years). The response was determined according to the criteria of the Japanese Esophageal Association's code of practice. The objective responses of the 121 patients who received NAC were complete response (CR) in 8 patients, partial response (PR) in 67, stable disease (SD) in 36, and progressive disease (PD) in 10. For this study, patients with CR and PR were defined as "responders," and patients with SD and PD were defined as "non-responders." There was no significant difference in the NAC response among several drug protocols, all of which were platinum-based (Table 1). Smoking is a major environmental factor associated with ESCC. However, the prevalence of these exposures varies widely between men and women.[27] In fact, in our cohort, males had a significantly higher Brinkman smoking index than women (Figure S2A). We found no significant influence of Brinkman index on NAC response in either male and/or female patients (Figures S2B and S2C). We compared overall survival (OS) and disease-free survival (DFS) excluding treatment-related deaths and postoperative-related deaths of less than 3 months in responders and non-responders. (Figures S2D–S2I). The responders tended to have a better OS than the non-responders ($p = 0.066$, log rank test for trend, Figure S2D), while the non-responders had a significantly lower DFS rate than the responders ($p = 0.012$ by log rank test for trend; Figure S2G). We found no difference in the OS ($p = 0.28$ by log rank test for trend; Figure S2E) or DFS periods ($p = 0.17$ by log rank test for trend; Figure S2H) between male responders and male non-responders. In contrast, female patients showed a significant difference between the responders and non-responders in both OS ($p = 0.0021$ by log rank test for trend; Figure S2F) and DFS ($p = 0.0005$ by log rank test for trend: Figure S2I), which is consistent with reports of gender differences in response to treatment in patients with esophageal cancer.[28]

Differential expression of immune-related genes is associated with response to NAC

To gain a deeper understanding of the potential mechanisms of therapeutic response to NAC, we performed RNA-seq in

CellPress
OPEN ACCESS

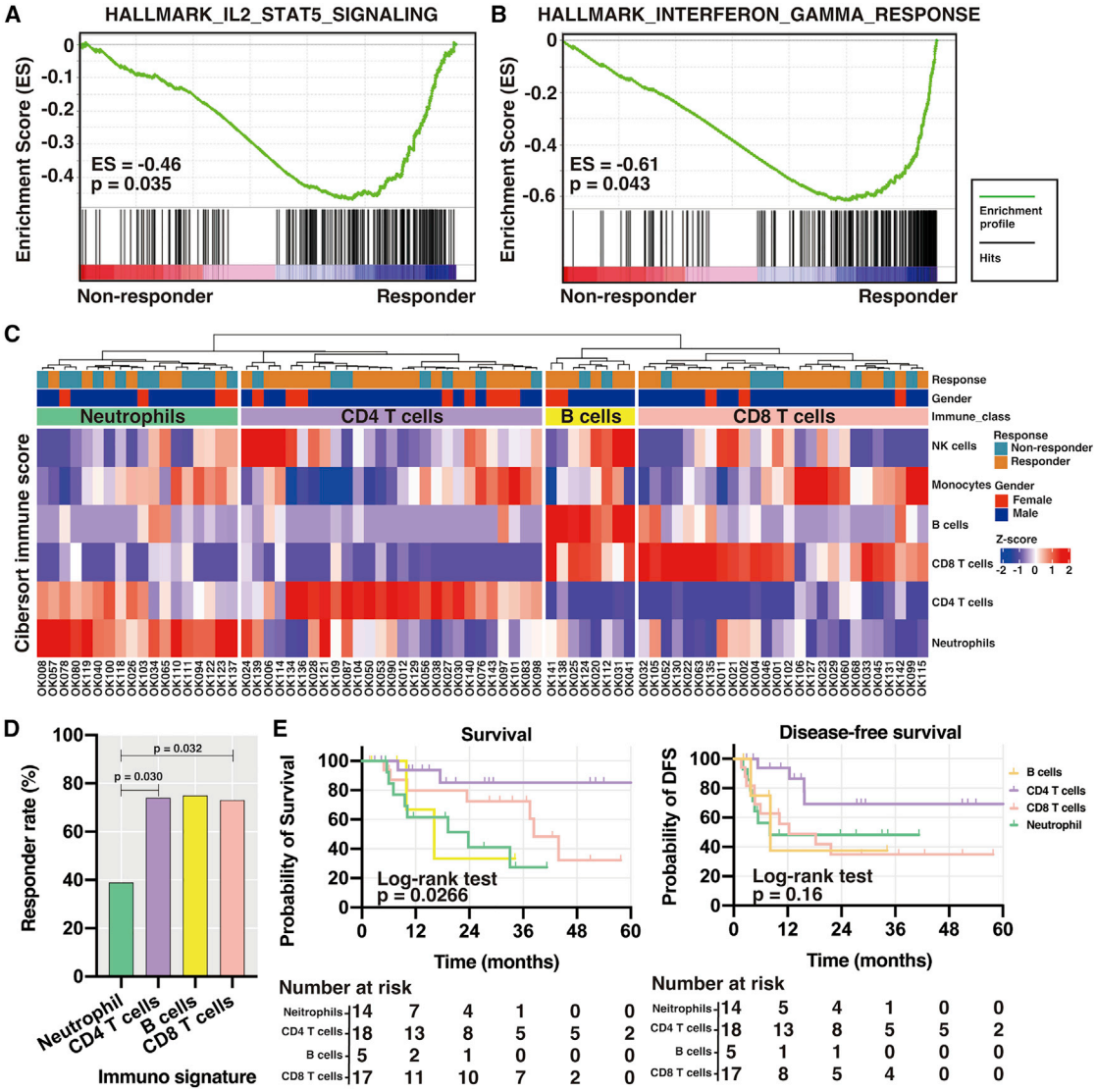


Figure 1. Classification of ESCC based on immune signatures

(A and B) GSEA of IL2-STAT5 signaling (A) and interferon-γ response (B) genes in ESCC RNA-seq data.

(C) Unsupervised clustering of ESCC patients by six gene expression signatures related to immune cell fractions (NK cells, monocytes, B cells, CD8+ T cells, CD4+ T cells, and neutrophils).

(D) Comparison of the responder rate: p = 0.030: neutrophils (N = 7/18) versus CD4 T cells (N = 20/27); and p = 0.032: neutrophils (N = 7/18) versus CD8 T cells (N = 19/26) by Fisher exact test.

(E) OS (Left) and DFS (Right) in each immune cell group. *p < 0.05.

tumor biopsy specimens before NAC treatment and analyzed the relationship between tumor gene expression and the response to NAC. Overall, we did not find any statistical differences in the expression of 19,652 genes between responders and non-responders (10% false discovery rate; Figure S3A). However, when we performed gene set enrichment analysis (GSEA) using hallmark gene sets of MSigDB between the responders and non-responders, several pathways involved in the immune response, such as IL2 STAT5 SIGNALING and INTERFERON_GAMMA_RESPONSE (p = 0.035 and 0.043, respectively; Figures 1A and 1B), were significantly enriched in the responders. In many cancers, tumor-infiltrating $CD8^+$ T cells predict patient survival and response to immunotherapy or chemotherapy.[29–31] Therefore, we analyzed T cell signatures using 13 known T cell signature transcripts[32] to distinguish between tumors with hot and cold $CD8^+$ T cell infiltration (Figure S3B). As a result, we categorized them into three subclasses based on their T cell signature characteristics: hot, middle, and cold. Of the 25 hot class patients, 22 patients responded well to NAC treatment, while one did not. NAC response rates were 88.0%, 60.5%, and 57.1% for the hot class, middle class, and cold class, respectively,

indicating a significant difference between the hot, middle, and cold classes (p = 0.023 and 0.047 by Fisher exact test; Figure S3C). No significant difference in OS and DFS was observed between the hot and cold classes (p = 0.37, 0.96 by log rank test for trend; Figure S3D). These results suggest that the differences in responsiveness to NAC may involve the immune system, and tumor immunity could determine the response to cytotoxic chemotherapy in ESCC. Therefore, we investigated a broader range of immune profiles from the RNA-seq data of the biopsy specimens.

Subclassification of immune cells in the tumor

To further elucidate the relationship between the genes and immune response to differential responsiveness to NAC, we estimated the relative abundance of constituent immune cell types in the mixed cell populations present in the TME using CIBERSORT with ESCC gene expression data. As a result, 79 ESCC were classified into 4 immune subclasses (Figure 1C) corresponding to those with a predominance of $CD8^+$ T cells, $CD4^+$ T cells, neutrophils, or B cells, based on their characteristic immune signatures. When the response rate was calculated for each subclass, the $CD8^+$ T cells, $CD4^+$ T cells, and B cell classes showed response rates of 73.0%, 74.0%, and 75.0%, respectively, while the neutrophil class had a response rate of 38.8%, which was significantly lower (p = 0.032 and 0.029 by Fisher exact test; Figure 1D). Tissue staining of specimens classified in the neutrophil subclass had significantly more neutrophils than tissue staining of specimens not classified in the neutrophil subclass (Figures S3E and S3F). Interestingly, when the $CD4^+$ T cell and $CD8^+$ T cell subclasses, which had significantly different proportions in the responders, were compared with the neutrophil subclasses, only the $CD4^+$ T cell subclass had a significantly prolonged OS (p = 0.029 by log rank test for trend; Figure 1E) and a tendency for prolonged DFS (p = 0.062 by log rank test for trend; Figure 1E). These results suggest that multiple immune cell subsets within the TME can impact NAC response and patient prognosis.

Neutrophil depletion promotes tumor sensitivity to NAC

To further investigate the biological association between neutrophils and NAC response, we transplanted mouse SCC ASB-XIV cells subcutaneously into syngeneic C57BL/6 mice and treated them with cisplatin and neutrophil depletion (STAR Methods). Five days after implantation, the SCC tumor reached a volume of 25 mm^3, and 7–10 mice per group were treated with saline (control), cisplatin alone, anti-Ly6G plus anti-kappa (neutrophil depletion), or anti-Ly6G plus anti-kappa plus cisplatin (for study design, see Figure 2A). Cisplatin alone or anti-Ly6G plus anti-kappa (neutrophil depletion) alone did not significantly inhibit tumor growth in this model. However, the combination of cisplatin with anti-Ly6G and anti-kappa (neutrophil depletion) resulted in marked tumor reduction and delayed outgrowth, which measures ANOVA on study days 9 and 12 (p = 0.0493 and 0.0013; Figure 2B), indicating that neutrophil infiltration in the tumor was related to tumor sensitivity to NAC (Figure 2C). We also performed an RNA-seq of these mouse tumors (n = 5 or 6 for each group) and analyzed the relationship between tumor gene expression and neutrophil depletion. Interestingly, GSEA using the MSigDB hallmark gene sets from RNA-seq data showed that only Notch signaling pathways were significantly enriched in the tumors treated with neutrophil depletion than those in the controls and the tumors treated with CDDP (p = 0.041 and 0.029, respectively; Figure 2D). Epithelial Notch signaling has been reported to promote metastasis by attracting neutrophils,[33] which is consistent with our data showing that depletion of neutrophils inhibits tumor growth. These results suggested that neutrophils could respond to chemotherapy through interaction with the Notch signaling pathway in SCC tumors.

Mutational profiles associated with response to NAC

Previously, we performed dWGS (×30∼) of 20 ESCC patients.[34] Here, we re-analyzed the dWGS data of the 20 ESCCs using another computational method and compared the genomes of the responders (n = 10) and non-responders (n = 10). We investigated the overall mutation burden in responders and non-responders at a genome-wide level and identified a median of 42.5 (0–165) non-synonymous mutations per tumor genome in the responders, compared with 30 (6–84) mutations in the non-responders, but the mutation burden per megabase (Mb) was not different in each group (p = 0.268 by Wilcoxon rank-sum test; Figure S4A). To further investigate the mutational profile of SNVs in the trinucleotide context, we analyzed mutational signatures of these samples using deconstructSigs,[35] and the identified mutational signatures were annotated using COSMIC mutation signatures (version 3.2) (Figure S4B). Interestingly, there was a significant difference between responders and non-responders in SBS92 reported as a tobacco smoking signature (Figure S4C). Brinkman index did not correlate with NAC responsiveness (Figures S2B and S2C), although smoking may have an indirect effect on NAC responsiveness.

Copy-number alterations (CNAs) of ESCC and response to NAC

Recent studies have shown that CNAs may be associated with chemotherapy sensitivity and immune response in several types of cancers.[36–38] Therefore, focusing on CNAs of ESCC, we performed shallow whole-genome sequencing (sWGS) of 72 cases in addition to dWGS of 20 cases (91 ESCC cases in total) and examined the relationship between CNAs and NAC response. There were no statistical differences in clinical information between dWGS and sWGS (Table S2). We found that 79% of ESCC had CNVs at the chromosome arm level, including gain at 1q, 3q, 5q, 8q, 20p, and 20q and loss at 3p, 4p, 5q, 9p, 13q, and 21q (Figure S5A). We also identified 13 focal regions with a recurrent gain of copy number and 26 regions with recurrent loss of copy number (both $q < 0.25$; Figure S5B), in which 471 and 1,858 genes were involved, respectively, and many have previously been identified as tumor-associated genes (Table S3). Consistent with previous reports, GISTIC2 (Figure S5B) detected somatic copy-number amplifications at 11q13.3 (*CCND1*), 3q26.33 (*TP63*/*SOX2*), and 8p11.23 (*FGFR1*), and a deletion at 9p21.3 (*CDKN2A*). There was no significant difference in purity (Figure S5C), ploidy (Figure S5D), and CNA burden (Figure S5E) between responders and non-responders. Among them, chromosomes 9p and 12q showed differentiated segment mean values and alteration counts between the responder and non-responder groups (p = 0.036 and 0.002 by

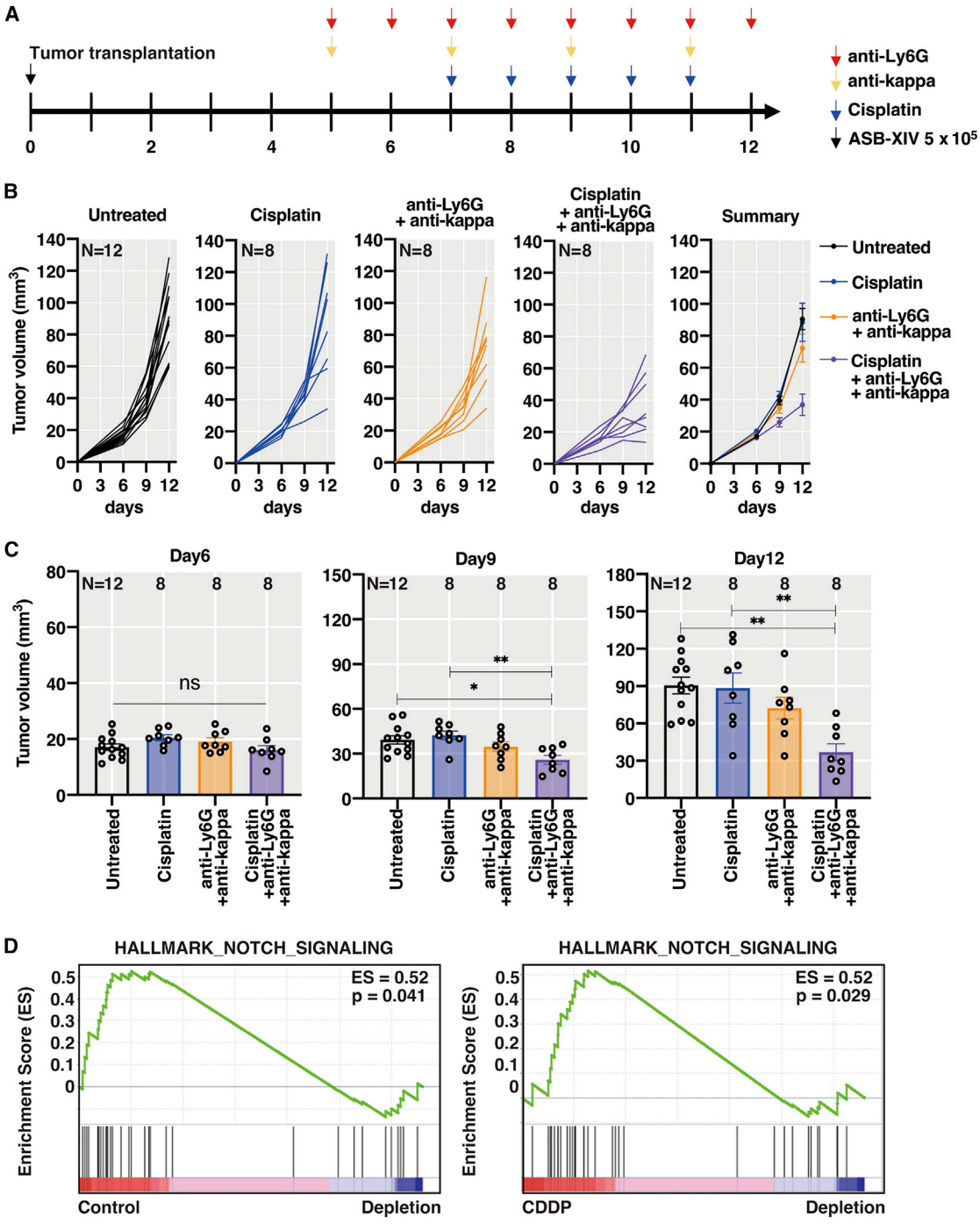


Figure 2. Anti-tumor efficacy of cisplatin and depletion of neutrophils in an SCC syngeneic mouse model

(A) Scheme of the cisplatin and depletion of the neutrophils test set. Humanized mice were generated by injecting ASB-XIV cells into mice.

(B and C) Tumor growth in each treatment group and summary (B) and comparison of tumor size on each day (C).

(D) Tumor RNA expression analysis was performed in the control non-treated group (n = 5), the neutrophil-depleted group (n = 6), and the CDDP-treated group (n = 5). GSEA between the control non-treated group (and neutrophil-depleted groups (left) and CDDP-treated and -depleted groups (right). GSEA found that only Notch signal pathways were significantly enriched in tumors treated with neutrophil depletion more than those in the control and tumors treated with CDDP ($p = 0.041$ and 0.029, respectively). *$p < 0.05$ and **$p < 0.01$ by Dunn's multiple comparisons test.

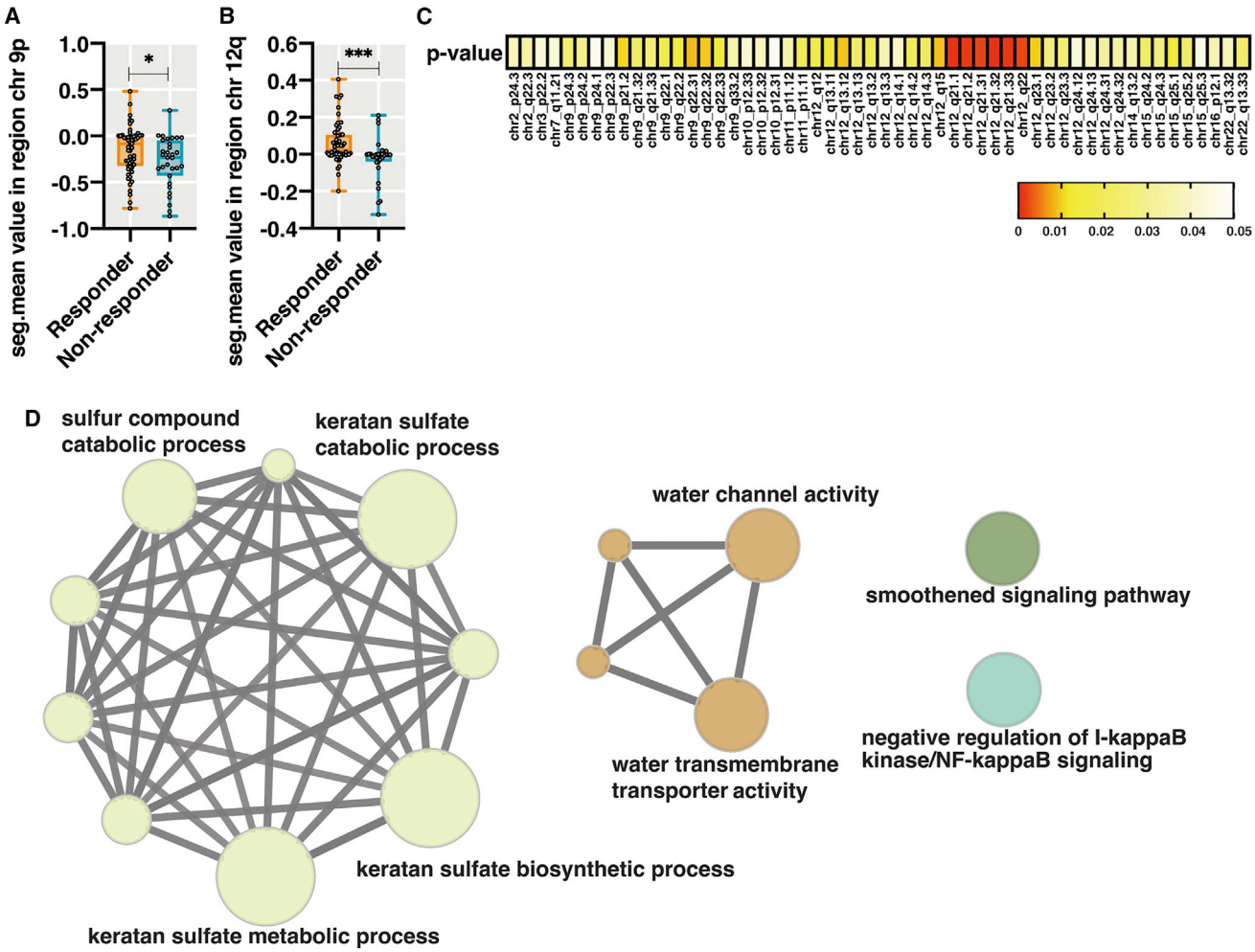


Figure 3. Somatic copy-number alterations in specific regions of ESCC

(A and B) Comparison of chromosome 9p (A) and 12q (B) segment means between NAC responders (N = 55) and non-responders (N = 31). Chromosomes 9p and 12q are significantly smaller in NAC non-responders than in responders (p = 0.036 and 0.002 by Student's test). *p < 0.05 and **p < 0.01 by Student's test

(C) Fifty-four recurrent focal CNA events were significantly different between the responders and non-responders. Locations with p values smaller than 0.01 are highlighted in red.

(D) Pathways of gene groups of locations with p value $\leq$ 0.001 among the sub-locations shown in (C). ClueGO Cytoscape visualizes the interaction of gene clusters in a functionally grouped network using enrichment maps. Nodes in the same cluster are assigned the same node color, and the node size indicates the number of genes mapped to each GO term. Node labels are determined based on the common themes among the processes in the cluster. Additional information on the clusters can be found in Table S3.

Student's test; Figure 3A). Among them, 54 recurrent focal CNA events were significantly different between responders and non-responders (p < 0.05 by Student's test; Figure 3C). To identify biological processes enriched in genes that have significant copy-number gain in responders compared with non-responders, we used ClueGO, a bioinformatics tool to identify biological functions associated with given gene sets.[39] ClueGO identified 20 biological pathways significantly enriched in genes that had significant copy-number gain in the responder (Figure 3C and Table S4). The 20 biological pathways were clustered into four groups: negative regulation of I-κB/NF-κB signaling, keratan sulfate, water channel activity, and smoothened signaling pathway. The human immune interferon-γ gene is located on chromosome 12. Consistent with the 12q amplification observed in the responder, RNA-seq data showed enriched INTERFERON_GAMMA_RESPONSE (Figure 1A). These data agree with recent reports that CNA may be associated with chemotherapy sensitivity and the immune response of ESCC.

Copy-number signatures and response to NAC

It has been reported that measurement of exposure to copy-number signatures provides a rational framework for choosing combination treatments that target multiple mutational processes.[40,41] Therefore, we next examined copy-number signatures using previously reported methods using non-negative matrix factorization.[40,41] Using random permutations of the data and four model selection measures, we found the optimal number of CN signatures in the ESCC dataset (n = 91) to be

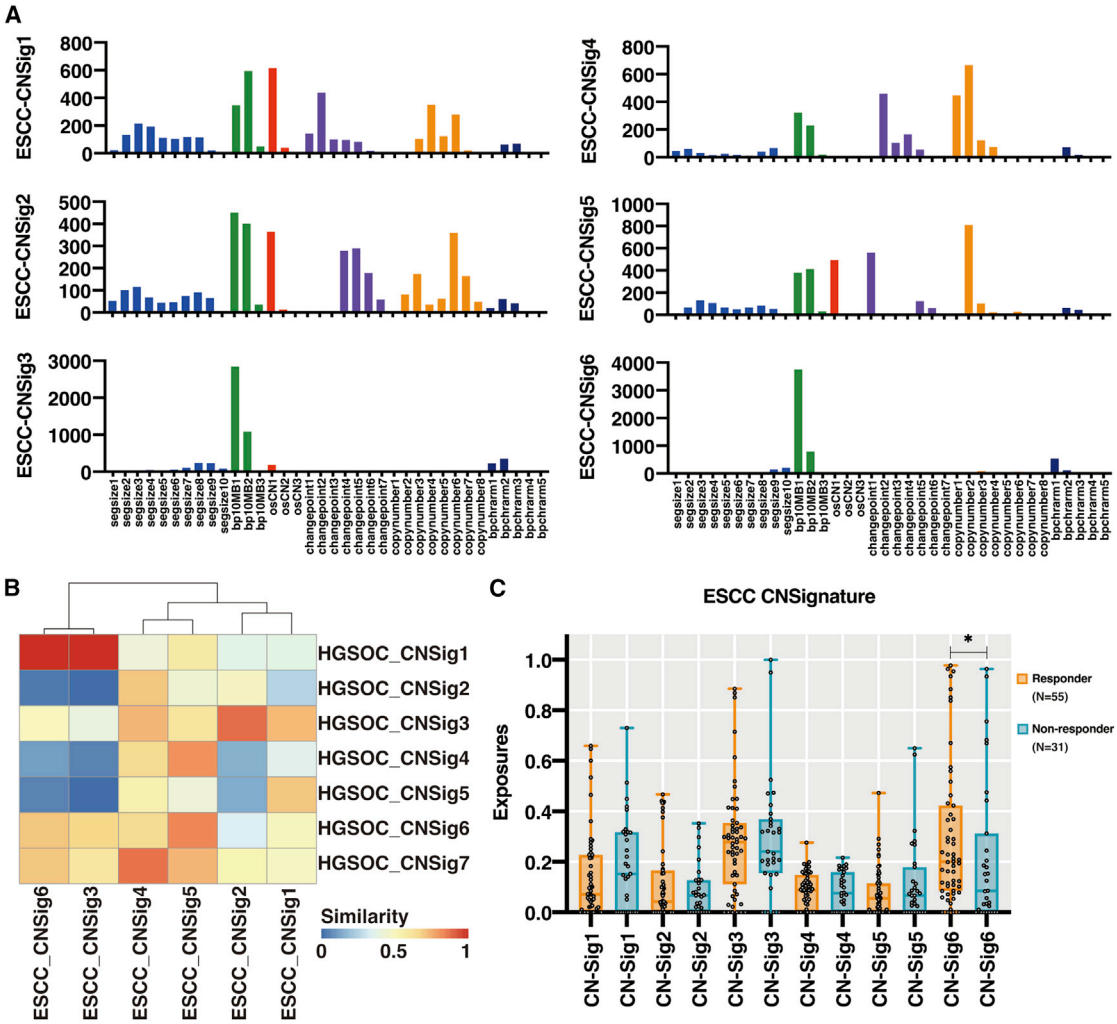


Figure 4. Copy number signatures in ESCC

Six copy-number signatures (ESCC-CNSig1, ESCC-CNSig2, ESCC-CNSig3, ESCC-CNSig4, ESCC-CNSig5, and ESCC-CNSig6) were identified using ESCC shallow and deep WGS data (n = 92).

(A) Defining features of the CN signatures, showing each feature (segsize, bp10MB, osCN, changepoint, copy number, bpcharm) split into 36 constituent components, as defined in Macintyre et al.[40] The mean value for each component is shown on the x axis, with the component weights shown on the y axis. Features are defined as follows: segment size (Mb); bp10MB, number of breakpoints (10 Mb−1); ocCN, region length with neighboring oscillating copy-number segments (Mb); changepoint, the difference in copy number between neighboring segments; copy number, the absolute copy number of a segment; bpcharm, breakpoints per chromosome arm.

(B) Identified copy-number signatures of ESCC were compared with HGSOC copy-number signatures using cosine similarity scoring.

(C) Comparison of copy-number signatures of NAC responders and non-responders. ESCC CNSig6 was significantly reduced in non-responders compared with responders (p = 0.034 by Mann-Whitney test). *p < 0.05.

six, namely ESCC-CNSig1 to ESCC-CNSig6 (Figure 4A). The CN signature in ESCC was then compared with the signature of ovarian cancer (HGSOC-CNSig1 to HGSOC-CNSig7) [40] (Figure 4B), and the identified ESCC-CNSig1 was similar to HGSOC-CNSig7 (cosine similarity 0.865), in which single copy-number changes were observed from a tetraploid rather than a diploid state. ESCC-CNSig2 was characterized by high copy-number states (four to eight copies) and predominant copy-number change points. This signature was similar to HGSOC-CNSig4 and HGSOC-CNSig6 (cosine similarity 0.804 and 0.865, respectively), while ESCC-CNSig3 and ESCC-CNSig6 were similar to HGSOC-CNSig1 (cosine similarity 0.990 and 0.998, respectively). None of the samples in this study contained the highest levels of ESCC-CNSig4, similar to HGSOC-CNSig1 (cosine similarity 0.871). ESCC-CNsig5, defined by single copy changes, had low similarity with all of the HGSOC-CNSigs (all cosine similarity <0.75). We compared these ESCC-CNSig in the responders and non-responders (Figure 4C), and the responders tended to have higher exposure to ESCC-CNSig6, which was related to chemotherapy response as HGSOC-CNSig1, than the non-responders (p = 0.034 by Mann-Whitney test; Figure 4C). CN signatures may not only be connected to genomic alterations

but may also be closely related to NAC response in ESCCs as well as ovarian cancers.[40]

Integrative prediction model by incorporating smoking status, transcriptomic, CNV, and immune characteristics

Tumor-immune cell interactions during NAC treatment are complicated and involve many factors, and a more comprehensive and integrated approach may lead to a better understanding of chemotherapy response and, hopefully, the capacity to predict NAC responsiveness. Based on this supposition, we developed a combined analysis associating NAC responsiveness and each of the four immune-signature subclasses and the CNV signatures identified in our dataset. This multi-omics information was used to perform a decision tree analysis, random forest method (RFM), and develop a model to predict the response after NAC treatment. The main idea of a decision tree is to identify the features that contain the most information regarding the target feature and then split the dataset along the values of these features such that the target feature values at the resulting nodes are as pure as possible. A feature that best separates the uncertainty from information about the target feature is the most informative feature. On the other hand, RFM has been used to recognize cancer-associated biomarkers from clinical trial data[42] to predict protein-protein interactions and to identify informative genes for a disease from microarray gene expression data.[42] RFM has many advantages: it is fast in both model training and evaluation, is robust to outliers, can capture complex nonlinear associations, cope with class imbalance data, and produces competitive performance for high-dimensional data.[42] It has also been shown to handle challenges arising from small sample sizes.[42]

We randomly split the cohort data, excluding patients with complex and difficult-to-understand diagnoses, into a 70% training set and a 30% test set to compensate for missing values. An additional sample of 20 new cases was used to examine the models created with these training and test data (Figure S1, Table S5). Figure 5A shows the final decision tree model of the recursive partitioning analysis for predicting the NAC responsiveness of ESCC patients, which was finally generated into seven layers and eight classes. Classes 1 to 4 were mainly classified as non-responders, and classes 5 to 8 were mainly classified as responders. Interestingly, there were more patients with CR in class 7 with higher chromosome 12q scores. The analysis identified chromosome 12q as the most important discriminating factor, followed by neutrophils, chromosome 9p, CD4 T cells, ESCC-CNSignature6, the Brinkman index, and CD8 T cells (Figure 5B). For validation, we used ROC to evaluate the discrimination of the ESCC prediction model, and the AUC for this decision tree model was 0.73 in the test dataset (Figure S6A). Based on these models, machine learning with RFM was used to predict responders and non-responders using the test data. The relationship between the percentage distribution of test data and NAC responsiveness is shown in Figure 5C. Classification of test data by the diagnostic prediction model showed 84.4% accuracy, 66.7% sensitivity, 66.7% specificity, 75% PPV, and 87.5% NPV. Patients classified in the responder category had a significantly longer OS compared with patients in the non-responder category ($p = 0.01$ by log rank test for trend; Figure S6B). Also, we found no difference in the DFS periods ($p = 0.14$ by log rank test for trend; Figure S6C) between responder category and non-responder category. A more detailed classification of NAC reactivity can lead to a therapeutic strategy. We created a decision tree model to predict pathologic CR (pCR) or PD (Figures S6D and S6E), which are more important issues in the clinical decision of ESCC treatment. As a result, patients classified as subclass 3 with high chromosome 12q score (amplification of chromosome 12q) and low monocyte expression were responders (pCR, 66.7%; responder, 100%; Figure S6D). On the other hand, patients with high monocyte expression and low ESCC-Sig6 scores were non-responders who did not respond well to NAC (PD, 100%; Figure S6E). In this model, perhaps NAC should not be administered to patients classified as such non-responder and surgical resection should be prioritized. Finally, diagnostics were also performed on another cohort of 20 new cases with the aim of validating this model. As a result, the AUC for the validation data with the predictive diagnostic model was 0.81 (Figure 5F) and the accuracy rate was 84%. Therefore, the diagnostic factors detected by this model are expected to be used to determine treatment strategies in clinical practice.

DISCUSSION

Here, we comprehensively profiled immune signatures, mutation signatures, CNVs, and copy-number signatures from pre-treatment biopsy specimens of ESCC in a cohort and powered to identify differences between responders and non-responders. We detected distinct mutational characteristics of ESCC between responders and non-responders across the spectrum, from large-scale chromosomal alterations to point mutations. Based on these results, we created a model to predict the responsiveness of NAC using machine learning. In our model, we found that neutrophils, ESCC-CNSig6, and smoking were negative predictors for non-responders to NAC.

The clinical prognosis of cancer patients depends on both the characteristics of the tumor itself and the host response. These include cells of the innate immune system, notably macrophages, neutrophils, mast cells, and myeloid progenitors, which infiltrate premalignant lesions and advanced tumors and collect at the margins of such lesions.[43] We used both genetic and pharmacologic approaches to deplete neutrophils in a mouse model of syngeneic tumors and found that neutrophil depletion by either method increased tumor growth delay after cisplatin treatment. Interestingly, patients who received granulocyte-macrophage colony-stimulating factor or granulocyte colony-stimulating factor to promote neutrophil recovery reported significantly decreased local tumor control.[44] Neutrophils can exert both pro-tumoral or anti-tumoral functions. In non-small cell lung cancer tissue, tumor-associated neutrophils can induce genetic instability, favor tumor growth, promote the remodeling of the extracellular matrix and tumor cell invasive capabilities, support angiogenesis and lymphangiogenesis, and suppress anti-tumoral adaptive immunity.[45] On the other hand, hyperactive Notch signaling in colorectal cancer cells triggers production of the chemokine, which recruits neutrophils via its receptor to the tumor. Notch signaling also drives the production of TGF-β, which subsequently engages

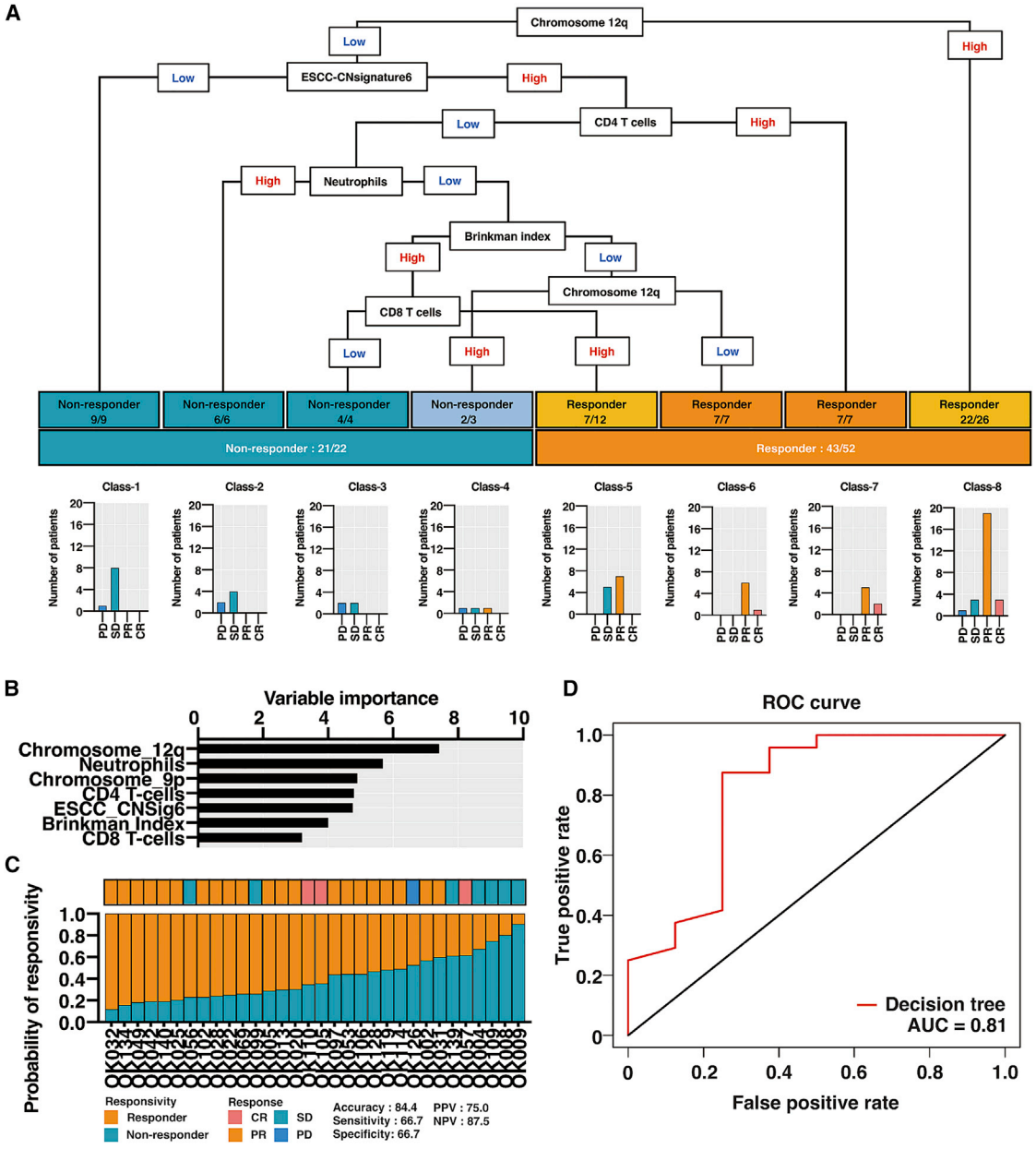


Figure 5. Multi-parameter integrative modeling accurately predicts the therapeutic outcome

(A) The diagnostic models to discriminate between responder and non-responder on the learning dataset. The decision tree had eight layers and eight nodes. The bar graphs show the respective number of patient responses at each node (class).

(B) Seven features with the highest weighting scores.

(C) Probability of each case (n = 32) being classified as responders or non-responders in the test set. This discriminating rule achieved 84.4% accuracy, 66.7% sensitivity, and 66.7% specificity.

(D) The responsiveness model with the validation set (n = 20) indicates that the area under the curve is 81%.

the TGF-β receptor and the associated kinase ALK5 on neutrophils to activate an inhibitory program that downregulates T cell responses in the tumor microenvironment (TME) and thereby generates an immune-suppressed niche that enables the metastatic process.[33] Taken together, these results indicate that neutrophils are important mediators of NAC resistance in ESCC, possibly through the Notch signaling pathway of the tumor.[33] The NLR in peripheral blood has been reported to be a potential predictor of chemotherapy sensitivity in ESCC.[46] These results suggest that neutrophils within the TME can impact NAC response and patient prognosis. However, the NLR has been reported to be associated with the patient's general condition or nutritional status,[47] which is a different concept from neutrophils within the tumor.

The mechanism of primary drug resistance due to smoking remains unknown. Smoking is associated with an increased mutational burden,[48] which may be one possible reason for this observation. In contrast, other studies have suggested overexpression of DNA repair enzymes in long-term smokers.[49] Our data showed a very close relationship between gender and smoking status. One or both of these mechanisms may be the reason for the failure of NAC. The burden of ESCC attributable to smoking and heavy alcohol consumption was much higher in men. In females, the burden of ESCC attributable to these factors was lower, and undernutrition may be a contributing factor.

Gene ontology of the genes in the region significantly associated with NAC response ($p < 0.01$) was primarily NF-κB signaling, smoothened, aquaporins, and keratan sulfate. One of the significant physiological roles of NF-κB is in the immune system. In particular, NF-κB family members regulate various aspects of innate and adaptive immune responses by controlling the transcription of genes that control cellular differentiation, survival, and proliferation, as well as cytokines and antimicrobial effectors. Furthermore, NF-κB contributes to the development and survival of cells and tissues responsible for the mammalian immune response.[50] Many chemotherapeutic agents, such as platinum-based drugs, have been shown to promote activation of the NF-κB pathway, a key transcription factor that plays a role in the development and progression of cancer and chemotherapy resistance by activating many mediators, including anti-apoptotic genes.[51] These factors may be related to the differences in the NAC response. In addition, we found that the copy-number signature of ESCC was significantly different with respect to NAC responsiveness and was highly similar to the copy-number signature of ovarian cancer related to platinum responsiveness.[40] CNAs are the most frequently observed alterations in immune-related genes, such as *TGFB2* and *IL10*,[52] and cancers harboring many CNAs, tend to show less immune involvement and worse response to immunotherapies.[53] CNAs or genomic instability might be one of the mechanisms by which CTLs and IFN-γ immuno-edits tumors in mouse models[54] and CNV signatures can potentially represent the mechanism of NAC response related to immune editing and the interaction between tumor genome and immune cells.

In summary, we identified genetic features and differences in immune reactivity that are uniquely associated with response to NAC. This indicates the presence of a subset of patients with pre-existing mutations that confer resistance to NAC. Importantly, these mutations may be clinically valid and can support targeted treatment strategies that have been successful in patients with metastatic ESCC, as various agents are in clinical trials. We envision a treatment prediction model that combines copy-number changes and immune profiling of ESCC with clinical data to improve the response to NAC and patient prognosis.

Limitations of the study

Certain limitations are noted in this study. First, we do not have any syngeneic mouse model of ESCC in the world as far as we know. ASB-XIV cell is a syngeneic mouse SCC originated from lung, but lung SCC is biologically, embryologically, and epidemiologically similar to ESCC. We used this syngeneic mouse model for this *in vivo* experiment. Moreover, due to the short follow-up period of the validation data, further validation with larger sample sizes is needed to improve this diagnostic model and overcome the challenges of its application to clinical practice. Further investigation is warranted to dissect the biological mechanisms of the NAC response, especially related to neutrophils.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Assessment of clinical response
- METHOD DETAILS
  - Tissue sample characteristics

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xcrm.2022.100705.

ACKNOWLEDGMENTS

We thank Professor James Brenton and Dr. Florian Markowetz at Cancer Research UK Cambridge Institute, Cambridge, UK, for their discussion and advice on copy-number signature analysis. The super-computing resource "SHIROKANE" was provided by the Human Genome Center, University of Tokyo (http://supcom.hgc.jp/). We would also like to acknowledge the technical staff at RIKEN IMS for their technical support and Ms. Hiroko Tanaka and the Human Genome Center staff, The University of Tokyo, for their data management in SHIROKANE. We would like to thank Editage (www.editage.com) for English language editing. This study was supported by P-CREATE (Project for Cancer Research and Therapeutic Evolution) of AMED Japan. Graphic abstract created in Bio Render (https://biorender.com/).

AUTHOR CONTRIBUTIONS

Conception and design, S.S. and H.N.; clinical and pathological sample/data collection, H.K., S.I., T.S., and T.Y.; development of methodology, S.S., H.K., K.N., C.S., T.A.J., M.F., K.M., Y.O., K.K., T.Y., and H.N.; analysis and interpretation of data, S.S., K.N., T.A.J., M.F., K.K., H.N.; writing, review, and/or revision of the manuscript, S.S., T.A.J., and H.N.; study supervision, H.N.

DECLARATION OF INTERESTS

All authors declare no conflicts of interest.

INCLUSION AND DIVERSITY

We worked to ensure ethnic or other types of diversity in the recruitment of human subjects.

Received: November 23, 2021
Revised: April 15, 2022
Accepted: July 11, 2022
Published: August 8, 2022

REFERENCES

1. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries CA Cancer. CA Cancer J. Clin. *68*, 394–424.
2. Jemal, A., Center, M.M., DeSantis, C., and Ward, E.M. (2010). Global patterns of cancer incidence and mortality rates and trends. Cancer Epidemiol. Biomarkers Prev. *19*, 1893–1907.
3. Zhang, H.-Z., Jin, G.F., and Shen, H.B. (2012). Epidemiologic differences in esophageal cancer between Asian and Western populations Chin. Chin. J. Cancer *31*, 281–286.
4. Yang, C.S., Chen, X., and Tu, S. (2016). Etiology and prevention of esophageal cancer. Gastrointest. Tumors *3*, 3–16.
5. Ando, N., Kato, H., Igaki, H., Shinoda, M., Ozawa, S., Shimizu, H., Nakamura, T., Yabusaki, H., Aoyama, N., Kurita, A., et al. (2012). A randomized trial comparing postoperative adjuvant chemotherapy with cisplatin and 5-fluorouracil versus preoperative chemotherapy for localized advanced squamous cell carcinoma of the thoracic esophagus (JCOG9907). Ann. Surg. Oncol. *19*, 68–74.
6. Ando, N., Iizuka, T., Ide, H., Ishida, K., Shinoda, M., Nishimaki, T., Takiyama, W., Watanabe, H., Isono, K., Aoyama, N., et al. (2003). Surgery plus chemotherapy compared with surgery alone for localized squamous cell carcinoma of the thoracic esophagus: a Japan Clinical Oncology Group Study—JCOG9204. J. Clin. Oncol. *21*, 4592–4596.
7. Stahl, M., Walz, M.K., Riera-Knorrenschild, J., Stuschke, M., Sandermann, A., Bitzer, M., Wilke, H., and Budach, W. (2017). Preoperative chemotherapy versus chemoradiotherapy in locally advanced adenocarcinomas of the oesophagogastric junction (POET): long-term results of a controlled randomised trial. Eur. J. Cancer *81*, 183–190.
8. Cao, X.F., He, X.T., Ji, L., Xiao, J., and Lv, J. (2009). Effects of neoadjuvant radiochemotherapy on pathological staging and prognosis for locally advanced esophageal squamous cell carcinoma. Dis. Esophagus *22*, 477–481.
9. Allum, W.H., Stenning, S.P., Bancewicz, J., Clark, P.I., and Langley, R.E. (2009). Long-term results of a randomized trial of surgery with or without preoperative chemotherapy in esophageal cancer. J. Clin. Oncol. *27*, 5062–5067.
10. Pennathur, A., Gibson, M.K., Jobe, B.A., and Luketich, J.D. (2013). Oesophageal carcinoma. Lancet *381*, 400–412.
11. Okamura, A., Watanabe, M., Mine, S., Kurogochi, T., Yamashita, K., Hayami, M., Imamura, Y., Ogura, M., Ichimura, T., Takahari, D., and Chin, K. (2017). Failure of neoadjuvant chemotherapy for resectable esophageal squamous cell carcinoma. Dis. Esophagus *30*, 1–8.
12. Ishibashi, Y., Tsujimoto, H., Hiraki, S., Kouzu, K., Tsuchiya, S., Itazaki, Y., Yaguchi, Y., Horiguchi, H., Nomura, S., Ito, N., et al. (2020). Predictive value of immuno-inflammatory and nutritional measures modulated by neoadjuvant chemotherapy on the response of neoadjuvant chemotherapy and long-term outcomes in patients with esophageal cancer. Oncol. Lett. *19*, 487–497.
13. Huang, Y., Sun, Y., Peng, P., Zhu, S., Sun, W., and Zhang, P. (2017). Prognostic and clinicopathologic significance of neutrophil-to-lymphocyte ratio in esophageal squamous cell carcinoma: evidence from a meta-analysis. OncoTargets Ther. *10*, 1165–1172.
14. Zhou, X.-L., Li, Y.Q., Zhu, W.G., Yu, C.H., Song, Y.Q., Wang, W.W., He, D.C., Tao, G.Z., and Tong, Y.S. (2017). Neutrophil-to-lymphocyte ratio as a prognostic biomarker for patients with locally advanced esophageal squamous cell carcinoma treated with definitive chemoradiotherapy. Sci. Rep. *7*, 42581.
15. Han, L., Song, Q., Jia, Y., Chen, X., Wang, C., Chen, P., Min, R., and Cheng, Y. (2016). The clinical significance of systemic inflammation score in esophageal squamous cell carcinoma. Tumour Biol. *37*, 3081–3090.
16. Xie, X., Luo, K.J., Hu, Y., Wang, J.Y., and Chen, J. (2016). Prognostic value of preoperative platelet-lymphocyte and neutrophil–lymphocyte ratio in patients undergoing surgery for esophageal squamous cell cancer. Dis. Esophagus *29*, 79–85.
17. Fridman, W.H., Zitvogel, L., Sautès-Fridman, C., and Kroemer, G. (2017). The immune contexture in cancer prognosis and treatment. Nat. Rev. Clin. Oncol. *14*, 717–734.
18. Bonavita, E., Gentile, S., Rubino, M., Maina, V., Papait, R., Kunderfranco, P., Greco, C., Feruglio, F., Molgora, M., Laface, I., et al. (2015). PTX3 is an extrinsic oncosuppressor regulating complement-dependent inflammation in cancer. Cell *160*, 700–714.
19. Bracci, L., Schiavoni, G., Sistigu, A., and Belardelli, F. (2014). Immune-based mechanisms of cytotoxic chemotherapy: implications for the design of novel and rationale-based combined treatments against cancer Cell Death. Cell Death Differ. *21*, 15–25.
20. Wu, X., Feng, Q.M., Wang, Y., Shi, J., Ge, H.L., and Di, W. (2010). The immunologic aspects in advanced ovarian cancer patients treated with paclitaxel and carboplatin chemotherapy Cancer. Cancer Immunol. Immunother. *59*, 279–291.
21. Chen, C.-A., Ho, C.M., Chang, M.C., Sun, W.Z., Chen, Y.L., Chiang, Y.C., Syu, M.H., Hsieh, C.Y., and Cheng, W.F. (2010). Metronomic chemotherapy enhances anti-tumor effects of cancer vaccine by depleting regulatory T lymphocytes and inhibiting tumor angiogenesis. Mol. Ther. *18*, 1233–1243.
22. Rettig, L., Seidenberg, S., Parvanova, I., Samaras, P., Curioni, A., Knuth, A., and Pascolo, S. (2011). Gemcitabine depletes regulatory T-cells in human and mice and enhances triggering of vaccine-specific cytotoxic T-cells. Int. J. Cancer *129*, 832–838.
23. Ghiringhelli, F., Apetoh, L., Tesniere, A., Aymeric, L., Ma, Y., Ortiz, C., Vermaelen, K., Panaretakis, T., Mignot, G., Ullrich, E., et al. (2009). Activation of the NLRP3 inflammasome in dendritic cells induces IL-1beta-dependent adaptive immunity against tumors. Nat. Med. *15*, 1170–1178.
24. Apetoh, L., Ghiringhelli, F., Tesniere, A., Obeid, M., Ortiz, C., Criollo, A., Mignot, G., Maiuri, M.C., Ullrich, E., Saulnier, P., et al. (2007). Toll-like receptor 4-dependent contribution of the immune system to anticancer chemotherapy and radiotherapy. Nat. Med. *13*, 1050–1059.
25. Michaud, M., Martins, I., Sukkurwala, A.Q., Adjemian, S., Ma, Y., Pellegatti, P., Shen, S., Kepp, O., Scoazec, M., Mignot, G., et al. (2011). Autophagy-dependent anticancer immune responses induced by chemotherapeutic agents in mice. Science *334*, 1573–1577.
26. Ma, Y., Adjemian, S., Mattarollo, S., Yamazaki, T., Aymeric, L., Yang, H., Portela Catani, J., Hannani, D., Duret, H., Steegh, K., et al. (2013). Immunity *38*, 729–741.
27. Pandeya, N., Olsen, C.M., and Whiteman, D.C. (2013). Sex differences in the proportion of esophageal squamous cell carcinoma cases attributable to tobacco smoking and alcohol consumption. Cancer Epidemiol. *37*, 579–584.
28. Liu, F., Feng, H., Guo, S., Chen, Y., Liu, Q., Wu, F., Gu, W., and Shan, B. (2019). Esophageal cancer: should gender be considered as an influential factor for patient safety in drug treatment? J. Oncol. *2019*, 6340567.
29. Jin, Y.W., and Hu, P. (2020). Tumor-infiltrating CD8 T cells predict clinical breast cancer outcomes in young women. Cancers *12*, 1076.
30. Seo, A.N., Lee, H.J., Kim, E.J., Kim, H.J., Jang, M.H., Lee, H.E., Kim, Y.J., Kim, J.H., and Park, S.Y. (2013). Tumour-infiltrating CD8+ lymphocytes as an independent predictive factor for pathological complete response to primary systemic therapy in breast cancer. Br. J. Cancer *109*, 2705–2713.
31. Eriksen, M., Lebreton, L.C.M., Carson, H.S., Thiel, M., Moore, C.J., Borerro, J.C., Galgani, F., Ryan, P.G., and Reisser, J. (2014). Plastic pollution in the World's oceans: more than 5 trillion plastic pieces weighing over 250, 000 tons afloat at sea. PLoS One *9*, e111913.
32. Spranger, S., Bao, R., and Gajewski, T.F. (2015). Melanoma-intrinsic β-catenin signalling prevents anti-tumour immunity. Nature *523*, 231–235.

33. Ruland, J. (2019). Epithelial Notch signaling recruits neutrophils to drive metastasis. Cancer Cell *36*, 213–214.

34. Dutta, M., Nakagawa, H., Kato, H., Maejima, K., Sasagawa, S., Nakano, K., Sasaki-Oku, A., Fujimoto, A., Mateos, R.N., Patil, A., et al. (2020). Whole genome sequencing analysis identifies recurrent structural alterations in esophageal squamous cell carcinoma. PeerJ *8*, e9294.

35. Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B.S., and Swanton, C. (2016). DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution Genome. Genome Biol. *17*, 31.

36. Lu, Z., Chen, H., Li, S., Gong, J., Li, J., Zou, J., Wu, L., Yu, J., Han, W., Sun, H., et al. (2020). Tumor copy-number alterations predict response to immune-checkpoint-blockade in gastrointestinal cancer. J. Immunother. Cancer *8*, e000374.

37. Sakimura, S., Nagayama, S., Fukunaga, M., Hu, Q., Kitagawa, A., Kobayashi, Y., Hasegawa, T., Noda, M., Kouyama, Y., Shimizu, D., et al. (2021). Impaired tumor immune response in metastatic tumors is a selective pressure for neutral evolution in CRC cases. PLoS Genet. *17*, e1009113.

38. Safonov, A., Jiang, T., Bianchini, G., Győrffy, B., Karn, T., Hatzis, C., and Pusztai, L. (2017). Immune gene expression is associated with genomic aberrations in breast cancer Cancer. Cancer Res. *77*, 3317–3324.

39. Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.H., Pagès, F., Trajanoski, Z., and Galon, J. (2009). ClueGO: a cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics *25*, 1091–1093.

40. Macintyre, G., Goranova, T.E., De Silva, D., Ennis, D., Piskorz, A.M., Eldridge, M., Sie, D., Lewsley, L.A., Hanif, A., Wilson, C., et al. (2018). Copy number signatures and mutational processes in ovarian carcinoma. Nat. Genet. *50*, 1262–1270.

41. Loveday, C., Litchfield, K., Proszek, P.Z., Cornish, A.J., Santo, F., Levy, M., Macintyre, G., Holryod, A., Broderick, P., Dudakia, D., et al. (2020). Genomic landscape of platinum resistant and sensitive testicular cancers. Nat. Commun. *11*, 2189.

42. Han, S., Williamson, B.D., and Fong, Y. (2021). Improving random forest predictions in small datasets from two-phase sampling designs. BMC Med. Inform. Decis. Mak. *21*, 322.

43. Hanahan, D., and Weinberg, R. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646–674.

44. Wisdom, A.J., Hong, C.S., Lin, A.J., Xiang, Y., Cooper, D.E., Zhang, J., Xu, E.S., Kuo, H.C., Mowery, Y.M., Carpenter, D.J., et al. (2019). Neutrophils promote tumor resistance to radiation therapy. Proc. Natl. Acad. Sci. USA *116*, 18584–18589.

45. Kargl, J., Busch, S.E., Yang, G.H.Y., Kim, K.H., Hanke, M.L., Metz, H.E., Hubbard, J.J., Lee, S.M., Madtes, D.K., McIntosh, M.W., and Houghton, A.M. (2017). Neutrophils dominate the immune cell composition in non-small cell lung cancer. Nat. Commun. *8*, 14381.

46. Sato, H., Tsubosa, Y., and Kawano, T. (2012). Correlation between the pretherapeutic neutrophil to lymphocyte ratio and the pathologic response to neoadjuvant chemotherapy in patients with advanced esophageal cancer World. World J. Surg. *36*, 617–622.

47. Kaya, T., Açıkgöz, S.B., Yıldırım, M., Nalbant, A., Altaş, A.E., and Cinemre, H. (2019). Association between neutrophil-to-lymphocyte ratio and nutritional status in geriatric patients. J. Clin. Lab. Anal. *33*, e22636.

48. Alexandrov, L.B., Ju, Y.S., Haase, K., Van Loo, P., Martincorena, I., Nik-Zainal, S., Totoki, Y., Fujimoto, A., Nakagawa, H., Shibata, T., et al. (2016). Mutational signatures associated with tobacco smoking in human cancer. Science *354*, 618–622.

49. Guo, M., Akiyama, Y., House, M.G., Hooker, C.M., Heath, E., Gabrielson, E., Yang, S.C., Han, Y., Baylin, S.B., Herman, J.G., and Brock, M.V. (2004). Hypermethylation of the GATA genes in lung cancer. Clin. Cancer Res. *10*, 7917–7924.

50. Hayden, M.S., West, A.P., and Ghosh, S. (2006). NF-κB and the immune response. Oncogene *25*, 6758–6780.

51. Godwin, P., Baird, A.M., Heavey, S., Barr, M.P., O'Byrne, K.J., and Gately, K. (2013). Targeting nuclear factor-kappa B to overcome resistance to chemotherapy. Front. Oncol. *3*, 120.

52. Mizuno, S., Yamaguchi, R., Hasegawa, T., Hayashi, S., Fujita, M., Zhang, F., Koh, Y., Lee, S.Y., Yoon, S.S., Shimizu, E., et al. (2021). Immunogenomic pan-cancer landscape reveals immune escape mechanisms and immunoediting histories. Sci. Rep. *11*, 15713.

53. Davoli, T., Uno, H., Wooten, E.C., and Elledge, S.J. (2017). Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. Science *355*, eaaf8399.

54. Takeda, K., Nakayama, M., Hayakawa, Y., Kojima, Y., Ikeda, H., Imai, N., Ogasawara, K., Okumura, K., Thomas, D.M., and Smyth, M.J. (2017). IFN-γ is required for cytotoxic T cell-dependent cancer genome immunoediting. Nat. Commun. *8*, 14607.

55. International Cancer Genome Consortium (2010). International network of cancer genome projects. Nature *464*, 993–998.

56. Shiraishi, O., Yamasaki, M., Makino, T., Motoori, M., Miyata, H., Shinkai, M., Kimura, Y., Hirao, M., Fujitani, K., Tamura, S., et al. (2017). Feasibility of preoperative chemotherapy with docetaxel, cisplatin, and 5-fluorouracil versus adriamycin, cisplatin, and 5-fluorouracil for resectable Advanced esophageal cancer. Oncology *92*, 101–108.

57. Hayata, K., Ojima, T., Nakamori, M., Nakamura, M., Katsuda, M., Kitadani, J., Takeuchi, A., Tabata, H., Maruoka, S., and Yamaue, H. (2018). Neoadjuvant chemotherapy with docetaxel, cisplatin and S-1 for resectable Advanced esophageal cancer. Anticancer Res. *38*, 5267–5273.

58. Japan Esophageal Society (2009). Japanese classification of esophageal cancer, tenth edition: parts II and III. Esophagus *6*, 71–94.

59. Fonseca, N.A., Petryszak, R., Marioni, J.C., and Brazma, A. (2014). iRAP—an integrated RNA-Seq analysis pipeline. Preprint at bioRxiv. https://doi.org/10.1101/005991.

60. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

61. Chen, B., Khodadoust, M.S., Liu, C.L., Newman, A.M., and Alizadeh, A.A. (2018). Profiling tumor infiltrating immune cells with CIBERSORT methods. Methods Mol. Biol. *1711*, 243–259.

62. Fan, X., Luo, G., and Huang, Y.S. (2021). Accucopy: accurate and fast inference of allele-specific copy number alterations from low-coverage low-purity tumor sequencing data. BMC Bioinf. *22*, 23.

63. Poell, J.B., Mendeville, M., Sie, D., Brink, A., Brakenhoff, R.H., and Ylstra, B. (2019). ACE: absolute copy number estimation from low-coverage whole-genome sequencing data. Bioinformatics *35*, 2847–2849.

64. Rigatti, S.J. (2017). Random forest. J. Insur. Med. *47*, 31–39.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Tumor tissue from participants in the University Hospital Medical Information Network Clinical Trials Registry of Japan	(identification number UMIN000004555/ 000004616)	(identification number UMIN000004555/ 000004616)
Deposited data		
Whole genome sequence data	This paper	JGA Accession#: JGAS000535 https://humandbs.biosciencedbc.jp/en/hum0316-v1
RNA sequencing data	This paper	JGA Accession#: JGAS000535 https://humandbs.biosciencedbc.jp/en/hum0316-v1
Chemicals, peptides, and recombinant proteins		
DMEM	Nacalai Tesque, Kyoto, Japan	09893-05
10% heat-inactivated fetal bovine serum	Sigma-Aldrich, St. Louis, Missouri, USA	MFCD00132239
streptomycin	FUJIFILM Wako Pure Chemical Corporation, Osaka, Japan	N/A
penicillin	FUJIFILM Wako Pure Chemical Corporation, Osaka, Japan	N/A
anti-Ly6G	clone 1A8, Bio X Cell, Lebanon, New Hampshire, USA	BP0075-1, RRID: AB_1107721
anti-rat κ immunoglobulin light chainIgG	clone MAR18.5, Bio X Cell	BE0122, RRID: AB_10951292
Cisplatin	AdipoGen, San Diego, California, USA	15663-27-1
QIAamp DNA Mini Kit	QIAGEN	ID51304
TruSeq Nano DNA Library Prep Kit	Illumina	20015965
Software and algorithms		
GSEA ver 4.1.0	N/A	https://www.gsea-msigdb.org/gsea/index.jsp
ClueGO	Bernhard Mlecnik et al. 2019	https://apps.cytoscape.org/apps/cluego
Cytoscape	Lotia S et al. 2013	https://apps.cytoscape.org/
ComplexHeatmap (version 2.4.3)	Zuguang Gu et al. 2016	https://jokergoo.github.io/ComplexHeatmap-reference/book/
CIBERSORT	Newman et al.	https://cibersort.stanford.edu/
Picard	N/A	http://broadinstitute.github.io/picard/
QDNAseq	Ilari Scheinin et al.	https://github.com/msfuji/qdnaseq-pipeline
GISTIC2 (v7)	N/A	https://www.genepattern.org/modules/docs/GISTIC_2.0
CNApp	Sebastia Franch-Exposito et al.	https://tools.idibaps.org/CNApp/
Accucopy	X Fan et al.	https://github.com/polyactis/Accucopy
rascal-absolute copy number scaling	N/A	https://bioinformatics.cruk.cam.ac.uk/rascal/
ACE	Jos B Poell et al.	http://bioconductor.org/packages/release/bioc/html/ACE.html
CNSignature	Geoff Macintyre et al.	https://bitbucket.org/britroc/cnsignatures/src/master/
Rpart	Terry Therneau et al.	https://cran.r-project.org/web/packages/rpart

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
random forest	N/A	https://cran.r-project.org/web/packages/randomForest
The ROCR R package	Tobias Sing et al.	http://cran.r-project.org/web/packages/ROCR/
All code	This paper	https://github.com/ssgwshota/ESCC_chemotherapy_response-paper.git
Experimental models: Cell lines		
ASB-XIV cell line	Eppelheim, Germany	400120, Eppelheim, Germany

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Hidewaki Nakagawa (hidewaki@riken.jp).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- WGS and RNA-seq reads for matched tumor, and non-tumor tissues data have been deposited at the National Bioscience Database Center (NBDC) and are publicly available under accession number JGAS000535 (NDBC:JGAS000535, "https://humandbs.biosciencedbc.jp/en/hum0316-v1").
- The package and analysis pipeline were obtained from the following URLs, respectively and all original code has been deposited at Github and is publicly available as of the date of publication. URLs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this work paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Full details of the eligibility criteria and pre-treatment evaluation were reported previously.[34] Briefly, Patients with histologically proven ESCC that was considered potentially resectable were eligible for inclusion in the study. The additional eligibility criteria were as follows: eligible patients were aged 20 years or older with performance status ≥ 1, histologically confirmed squamous cell carcinoma of the thoracic esophagus, and adequate bone marrow, renal, hepatic, and pulmonary function. Patients with clinical stage Ib to IIIB, stage IIIC without T4b, or stage IV disease based on only supraclavicular lymph node metastasis were eligible. Biopsy specimens of tumors and normal samples were obtained from patients were treated with NAC at Kindai University Hospital, Osaka, Japan. All patients agreed to participate in the study and provided written informed consent following ICGC guidelines.[55] The study was approved by the institutional review boards of the Kindai University Hospital and RIKEN. Clinicopathological data are available in Table 1, and detail information of each case is available in Tables S1 and S5.

Assessment of clinical response

The "Response Evaluation Criteria for Chemotherapy in Solid Tumors" published by the Japan Society of Clinical Oncology (JSCO), which were based on the RECIST evaluation criteria, was adopted as the response evaluation criteria[56–58] (Table S6). Patients with complete (CR) and partial (PR) response were classified as responders while patients with stable (SD) or progressive disease (PD) were classified as non-responders. PFS and OS were defined as the time from the surgery to the time when documented evidence of progressive disease or death were obtained.

METHOD DETAILS

Tissue sample characteristics

Three biopsy specimens were obtained from the center of the non-necrotic tumorous regions without positive lugol staining and two specimens from the normal esophageal mucosa, apart from the tumor region with positive lugol staining, by performing gastrointestinal endoscopy prior to NAC. One tumor biopsy specimen was pathologically examined after Hematoxylin-eosin (HE) staining. Other specimens were snap-frozen and stored at -80°.

RNA sequencing

RNA was extracted from frozen biopsy specimens using TRIzol reagent. After checking the quality and quantity of total RNA using a Bioanalyzer (Agilent Technologies), RNA sequencing libraries were generated using the TruSeq Stranded mRNA Library Preparation Kit (Illumina). RNA sequencing was performed using a HiSeq2500 (Illumina). Read mapping to the human reference genome (GRCh37), and the mouse reference genome (GRCm38) with TopHat2 and per-gene read counting using GENCODE release 19 with HTseq were orchestrated by the iRAP pipeline.[59] Fragments per kilobase of exon per million fragments mapped with upper quartile normalization (FPKM-UQ) were computed and used as the gene expression levels throughout the study. When log expression was required, 0.01 was added to FPKM-UQ as an offset. For mouse tumor models, RNA was extracted from tumor specimens using TRIzol, and RNA sequencing libraries were generated using the TruSeq Stranded mRNA Library Preparation Kit (Illumina). RNA sequencing was performed using a HiSeq2500 (Illumina).

Molecular pathway analysis

GSEA was performed using GSEA ver 4.1.0. We selected Hallmarks h.all.v7.2.symbol.gmt as the gene set database and performed GSEA using the default settings. We sorted them in the NES order and selected the categories in detail. To identify Gene Ontology for genes enriched in CNAs in responders and non-responders to NAC treatment, we used ClueGO[39] in Cytoscape[60] with default settings.

T-cell signature analysis and immune clustering

Thirteen known genes (*CD8A*, *CCL2*, *CCL3*, *CCL4*, *CXCL9*, *CXCL10*, *ICOS*, *GZMK*, *IRF1*, *HLA-DMA*, *HLA-DMB*, *HLA-DOA*, and *HLA-DOB*[32]) were selected from the FPKM-UQ file, and heat maps were generated using the R Bioconductor package ComplexHeatmap (version 2.4.3). FPKM-UQ expression levels were scaled by gene, then Euclidean distances were calculated, and hierarchical clustering was performed using the Ward method. CIBERSORT uses the LM22 and LM6 databases to identify immune subclasses; since the 22 immune cells are microarray-based and the 6 immune cells are RNAseq-based, our data is based on the CIBERSORT scores of the 6 RNAseq-based immune cells Immune signatures were calculated. To immunologically classify ESCC, we estimated the proportions of six gene expression signatures of immune cell subsets (NK cells, monocytes, B cells, CD8 T cells, CD4 T cells, and neutrophils) using CIBERSORT with LM6,[61] an RNAseq gene signature matrix based on GSE60424. The estimated relative proportions of immune cell subsets were converted to Z-scores, and hierarchical clustering was performed using the Ward method. Dendrograms were cut after visual inspection into four clusters, corresponding to $CD4^+$ T cells, $CD8^+$ T cells, NK cells, and B cell immune classes. These four signature clusters were scaled across the samples, and the distance between samples was defined by Pearson's correlation.

Syngeneic mice model

C57BL/6 mice were inoculated with 5×10^5 ASB-XIV cells subcutaneously into the right flank on day 0.

ASB-XIV cell is a syngeneic mouse SCC originated from lung, but lung SCC is similar to ESCC biologically, embryologically, and epidemiologically. This cell lines were obtained from CLS (400120, Eppelheim, Germany) and maintained in Dulbecco's modified Eagle's medium (DMEM, Nacalai Tesque, Kyoto, Japan) with 10% heat-inactivated fetal bovine serum (Sigma-Aldrich, St. Louis, Missouri, USA), 100 μg/mL streptomycin, 100 U/mL penicillin (FUJIFILM Wako Pure Chemical Corporation, Osaka, Japan). To deplete the neutrophils, anti-Ly6G (50 μg, clone 1A8, Bio X Cell, Lebanon, New Hampshire, USA) and/or anti-rat kappa immunoglobulin light chainIgG (100 μg, clone MAR18.5, Bio X Cell) antibodies were injected intraperitoneally into tumor-bearing mice on 5–12 days or 5, 7, 9 and 11 days. Cisplatin (0.3 mg/kg, AdipoGen, San Diego, California, USA) was injected intraperitoneally on days 7–11 days. These concentrations and dosing periods were validated in a prior experiment. Tumor growth was monitored every three days using calipers, and tumor volume was calculated using the formula $\pi/6^*L1L2H$, where L1 is the long diameter, L2 is the short diameter, and H is the height of the tumor. Tumors were harvested on day 14 for RNAseq. RNAseq was performed on tumors excluding the smallest and largest tumors to exclude outliers.

Whole genome sequencing and CNV analysis

DNA was extracted from the frozen tumor and normal mucosa biopsy specimens using the QIAamp DNA Mini Kit (QIAGEN). The libraries were prepared using the TruSeq Nano DNA Library Prep Kit (Illumina) following the manufacturer's protocol. Paired-end sequencing of 126-bp reads was performed using HiSeq2500. For 20 cases, we previously performed dWGS (tumor 40x and normal 30x),[34] and for 84 matched pairs of tumor and normal specimens, we performed sWGS (tumor and normal 1.0x). Sequence reads were mapped to the human reference genome GRCh37 using BWA-0.7.8, and the PCR duplicates were removed using the Picard tool. The segment files generated from the QDNAseq pipeline were used to calculate copy number amplified and deleted regions, using the default settings of the GISTIC2 (v7) and CNApp bioinformatics tools.

Copy number signature identification

The BAM files generated from sWGS and dWGS were segmented into segment files using the QDNAseq pipeline. Purity and ploidy were calculated using AccuCopy.[62] These data were used to calculate the absolute copy number using two bioinformatics tools: rascal-absolute copy number scaling and ACE.[63] The absolute copy number was used to calculate the copy number signature of ESCC using the advanced source code of CNSignature. To explain in detail, we used the following genome-wide distribution of six different copy number features. (1) Segment size: the length of each genome segment; (2) breakpoint count per 10 Mb: the number of genome breaks appearing in 10-Mb sliding windows across the genome; (3) Changepoint copy number: the absolute difference in copy number between adjacent segments across the genome; (4) segment copy number: the observed absolute copy number state of each segment; (5) breakpoint count per chromosome arm: the number of breaks occurring per chromosome

arm; (6) length of segments with oscillating copy number: a traversal of the genome counting the number of contiguous copy number segments alternating between two copy number states, rounded to the nearest integer copy number state.[40] The chooseNumberSignatures function used the R package NMF (v0.21.0) to deconvolve the sum-of-posteriors matrix of per-patient components into a per-patient signature matrix and a per-signature component matrix. As a result, six signatures were identified as the appropriate numbers. Identified ESCC-CNSignatures were compared with Geoff Macintyre's CNSignatures (HGSOC-CNSignatures) using cosine similarity scoring.

Decision tree and prediction model

Integrated machine learning analysis of genomic, transcriptomic, copy number signature, and immunologic features was performed using the "rpart" and "random forest" R packages. We randomly split our cohort data into an 70% training set and a 30% testing set. The 70/30 split is a common practice of splitting ratio for moderately sized samples in machine learning applications. We chose this ratio to obtain sufficient training samples to build a good model and sufficient test samples to evaluate the model. In addition to these data, we prepared nine new cases of validation data. A partition tree (RPART) is a classification model based on a top-down methodology in which, starting from a root node, binary splits of data are generated until a specific criterion (i.e., the minimization of the node impurity) is encountered. This method is prone to overfitting the training data. The cross-validation or bootstrapping procedure is a useful method to limit overfitting, leading to the proper tuning of the decision tree (DT) parameters and optimizing the model accuracy. Random Forest is a tree-based algorithm that involves the computation of hundreds to thousands of RPART trees and merges the DT output to increase the generalizability of the model.[64] The ROCR R package was applied for AUC evaluation of response classification accuracy based on out-of-bag predictions.

Statistical analyses

The statistical comparison between responder and non-responder groups for a given continuous variable was performed using the two-sided Mann–Whitney U-test. The association between two continuous variables was assessed using Spearman's rank correlation coefficient. Univariate and multivariate analyses predicting the response to NAC were performed using logistic regression modeling. We used R (ver 3.6.0) and GraphPad Prism 8 for analyses and graphical plotting.