Identification of novel prognostic and predictive biomarkers in salivary duct carcinoma via comprehensive molecular profiling

Shinji Kohsaka1,2,3, Yuichiro Tada4,5, Mizuho Ando4,5, Masato Nakaguro4, Yukina Shirai1, Toshihide Ueno1, Shinya Kojima1, Hideaki Hirai5, Natsuki Saigusa4, Satoshi Kano9, Kiyoko Sukurara7, Takafumi Togashi9, Hiroyuki Ozawa9, Takahito Kondo10, Kenji Okami11, Hideaki Takahashi12, Daisuke Kawakita13, Chihiro Fushimi1, Takayoshi Suzuki6, Akira Shimizu7, Isaku Okamoto7, Takuro Okada1, Yuichiro Sato14, Yoshihisa Imamish6, Yoshihiro Watanabe9, Akiko Sakai11, Koji Ebisumoto11, Yukiko Sato12, Makoto Urano16, Yoshihisa Honma17, Keisuke Yamazaki18, Yushi Ueki18, Toyoyuki Hanazawa19, Yuki Saito20, Tomotaka Shimura21, Toshitaka Nagao5 and Hiroyuki Mano1,2

Molecular targets and predictive biomarkers for prognosis in salivary duct carcinoma (SDC) have not been fully identified. We conducted comprehensive molecular profiling to discover novel biomarkers for SDC. A total of 67 SDC samples were examined with DNA sequencing of 464 genes and transcriptome analysis in combination with the clinicopathological characteristics of the individuals. Prognostic biomarkers associated with response to combined androgen blockade (CAB) treatment were explored using mRNA expression data from 27 cases. Oncogenic mutations in receptor tyrosine kinase (RTK) genes or genes in the MAPK pathway were identified in 55 cases (82.1%). Alterations in the phosphatidylinositol 3-kinase (PI3K)/AKT signaling pathway were identified in 38 cases (56.7%). Interestingly, patient prognosis could be predicted using mRNA expression profiles, but not genetic mutation profiles. The risk score generated from the expression data of a four-gene set that includes the ADAMTS1, DSC1, RNF39, and IGLL5 genes was a significant prognostic marker for overall survival in the cohort (HR = 5.99, 95% confidence interval (CI) = 2.73–13.1, p = 7.8 × 10−5). Another risk score constructed from the expression of CD2E and LOXB3 was a strong prognostic marker for progression-free survival for CAB treatment (p = 0.03). Mutations in RTK genes, MAPK pathway genes, and PI3K/AKT pathway genes likely represent key mutations in SDC tumorigenesis. The gene expression profiles identified in this study may be useful for stratifying patients who are good candidates for CAB treatment and may benefit from additional systemic therapies.

INTRODUCTION
Salivary duct carcinoma (SDC) is an aggressive malignancy that resembles high-grade mammary ductal carcinoma1–3. SDC was once regarded to account for 1–4% of all salivary gland carcinomas (SGCs)4–6, although recent studies have reported a higher incidence than previously recorded4,7. SDC usually occurs in major salivary glands such as the parotid gland and is often detected in males greater than 50 years old8,9. SDC can arise de novo or as the malignant component of carcinoma ex pleomorphic adenoma (PA)10. The current treatments for SDC include complete surgical resection of the primary site followed by adjuvant chemotherapy/radiotherapy2,3,8,9. However, recurrence and distant metastasis frequently occur, resulting in a poor prognosis. With conventional therapy, more than 50% of SDC patients die of their disease in 3–5 years, and they are at substantial risk of recurrence even after complete surgical resection. The 5-year overall survival (OS) ranges from 20 to 30%2,3,6,8,9,11. Efficacy data are limited for systemic therapy or radiation therapy for individuals with advanced SGC following resection1,2,8,9. Reliable prognostic biomarkers are desperately needed to select patients who are at high-risk for recurrence and who may benefit from additional systemic or radiation therapy1–3.

Similar to breast carcinoma, conventional immunohistochemistry has identified that SDC often exhibits overexpression of androgen receptor (AR), HER2 (encoded by ERBB2 gene) and epidermal growth factor receptor (EGFR)1,2,11,12. However, SDC usually lacks estrogen receptor- or progesterone receptor-expression which is often observed in breast carcinoma13,14.

1Division of Cellular Signaling, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan. 2Department of Head and Neck Oncology and Surgery, International University of Health and Welfare, Mita Hospital, Tokyo, Japan. 3Department of Otolaryngology-Head & Neck Surgery, Okayama University Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama, Japan. 4Department of Pathology and Laboratory Medicine, Nagoya University Hospital, Nagoya, Japan. 5Department of Anatomic Pathology, Tokyo Medical University, Tokyo, Japan. 6Department of Otolaryngology Head and Neck Surgery, Faculty of Medicine and Graduate School of Medicine, Hokkaido University, Sapporo, Japan. 7Department of Otolaryngology Head and Neck Surgery, Tokyo Medical University, Tokyo, Japan. 8Department of Head and Neck Surgery, Niigata Cancer Center Hospital, Niigata, Japan. 9Department of Otolaryngology Head and Neck Surgery, Keio University School of Medicine, Tokyo, Japan. 10Department of Otolaryngology Head and Neck Surgery, Tokyo Medical University Hachioji Medical Center, Tokyo, Japan. 11Department of Otolaryngology and Head Neck Surgery, Tokai University School of Medicine, Isehara, Japan. 12Department of Otolaryngology, Head and Neck Surgery, Yokohama City University School of Medicine, Yokohama, Japan. 13Department of Otolaryngology Head and Neck Surgery, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan. 14Department of Otolaryngology and Head and Neck Surgery, School of Life Dentistry at Niigata, The Nippon Dental University, Tokyo, Japan. 15Division of Pathology, Cancer Institute Hospital, Japanese Foundation for Cancer Research, Tokyo, Japan. 16Department of Diagnostic Pathology Bantane Hospital Fujita Health University, School of Medicine, Nagoya, Japan. 17Department of Head & Neck, Esophageal Medical Oncology, National Cancer Center Hospital, Tokyo, Japan. 18Department of Otolaryngology Head and Neck Surgery, Niigata University Graduate School of Medical and Dental Sciences, Niigata, Japan. 19Department of Otorhinolaryngology/Head & Neck Surgery, Chiba University Graduate School of Medicine, Chiba, Japan. 20Department of Otolaryngology - Head and Neck Surgery, Faculty of Medicine, The University of Tokyo, Tokyo, Japan. 21Department of Otorhinolaryngology, Showa University Fujigaoka Hospital, Yokohama, Japan. Email: skohsaka@ncc.go.jp; ytada@iuhw.ac.jp; hmano@ncc.go.jp

Published in partnership with The Hormel Institute, University of Minnesota

www.nature.com/npjprecisiononcology

npj Precision Oncology (2022)6:82; https://doi.org/10.1038/s41698-022-00324-1
Combined androgen blockade (CAB) with LHRH analog and bicalutamide, a common androgen deprivation therapy (ADT), has a relatively mild toxicity profile and is considered a promising treatment option for patients with unresectable AR–positive SGC. Several clinical trials have demonstrated a favorable response in patients treated with ADT. In prostate cancer, generation of AR-V7 encoded by a splice variant of AR may confer resistance to androgen inhibitor treatment, whereas few specific biomarkers predicting the efficacy of CAB treatment have been identified in SGC. For HER2-targeted therapy, a combination of trastuzumab and chemotherapy resulted in impressive response rates in clinical trials.

Due to the development of the next-generation sequencing, genetic, transcriptomic, and proteomic profiles have been investigated in SDC. Tumor mutational burden is higher than other tumor types of SGC, and the phosphatidylinositol 3-kinase (PI3K)/AKT/mTOR pathway is the most frequently altered pathway. Other mutations are frequently found in TP53, HRAS/NRAS, ERBB2, EGFR, BRAF, and the CCND1/CDK pathway genes. Reports have described prognostic factors of the clinical outcomes of SDC. The degree of immune cell infiltration in tumor microenvironment is a predictive marker for immune checkpoint inhibitors. However, studies have not presented abundant integrative analysis combining comprehensive molecular profiling data with detailed clinical data of SDC patients. Therefore, we performed an integrative analysis combining comprehensive cancer-gene panel along with RNA sequencing (RNA-seq) and analyzed these data in combination with detailed clinical features in a cohort of SDC patients.

**RESULTS**

**Patient characteristics**

The study cohort comprised 76 SDC patients who underwent surgical resection between 2005 and 2017 at hospitals participating in the Japan SDC consortium. Nine cases were excluded because of poor DNA quality isolated from the specimens. The demographic and clinical data of the remaining 67 individuals, eight patients were female and 59 were male, are summarized in Table 1. The median age was 62.1 years old and was similar between female and male patients (mean, 59.4 vs. 62.6 years; \( p = 0.5 \)). The most common location of the primary tumors was the parotid gland (76.1%) followed by the submandibular gland (20.9%). SDC occurred as a malignant component of carcinoma ex pleomorphic adenoma in 14 (20.9%) patients. We used the TOP cancer gene panel to evaluate somatic single nucleotide variations (SNVs) and insertions/deletions (indels) for 464 genes in the panel, and to determine tumor mutation burden (TMB) and copy number variations. We identified frequent oncogenic mutations within the genes for the receptor tyrosine kinase (RTK)/MAPK signaling pathway including 24 cases with HRAS activating mutations, 46 cases with BRAF mutations, 4 cases with MAP2K4 mutations, and 13 cases with NF1 mutations (Fig. 1a). TP53 was altered in approximately half of the cases. Copy number analysis further identified six cases with ERBB2 amplification. In total, oncogenic

---

**Table 1. Demographic features of the 67 patients with salivary duct carcinoma.**

| Feature                        | No. of patients (N = 67) |
|--------------------------------|--------------------------|
| Age in years, median (IQR)     | 62.1 (55–70)             |
| Range, years                   | 30–94                    |
| Age (years), N (%)             |                          |
| <50                            | 15 (22.4)                |
| 51–60                          | 11 (16.4)                |
| 61–70                          | 25 (37.3)                |
| 71–80                          | 14 (20.9)                |
| >80                            | 2 (3)                    |
| Sex, N (%)                     |                          |
| Female                         | 8 (11.9)                 |
| Male                           | 59 (88.1)                |
| Location of primary tumor, N (%)|                          |
| Parotid gland                  | 51 (76.1)                |
| Submandibular gland            | 14 (20.9)                |
| Sublingual gland               | 0 (0)                    |
| Minor salivary gland           | 2 (3)                    |
| Tumor arising from pleomorphic adenoma, N (%) |          |
| Yes                            | 49 (73.1)                |
| No                             | 12 (17.9)                |
| Not determined                 | 6 (9.0)                  |
| T-stage, N (%)*                |                          |
| I                              | 12 (17.9)                |
| 2                              | 22 (32.8)                |
| 3                              | 8 (11.9)                 |
| 4                              | 25 (37.3)                |
| N-stage, N (%)*                |                          |
| 0                              | 32 (47.8)                |
| 1                              | 4 (6.0)                  |
| 2                              | 31 (46.3)                |
| 3                              | 0 (0)                    |
| M-stage, N (%)*                |                          |
| 0                              | 59 (88.1)                |
| 1                              | 8 (11.9)                 |
| Overall stage, N (%)*          |                          |
| I                              | 7 (10.4)                 |
| II                             | 11 (16.4)                |
| III                            | 7 (10.4)                 |
| IV                             | 42 (62.7)                |
| Initial treatment, N (%)       |                          |
| Surgery alone                  | 22 (32.8)                |
| Surgery and (chemo-) radiotherapy | 38 (56.7)               |
| Chemotherapy and/or radiotherapy | 7 (10.4)                |
| Recurrence after initial treatment, N (%)  |                      |
| No persistence or recurrence   | 18 (26.9)                |
| Recurrence after initial treatment | 49 (73.1)             |
| Survival status at last census, N (%) |                      |
| Alive, no evidence of disease  | 16 (23.9)                |
| Alive with disease             | 20 (29.9)                |
| Died of disease                | 28 (41.8)                |
| Died of other cause            | 3 (4.5)                  |
| Follow-up time in months, median (IQR) | 40 (24, 61)          |

*The 8th edition of the UICC (Union for International Cancer Control) staging system, IQR interquartile range.*
mutations in RTK genes or genes in the MAPK pathway were present in 55 of the cases (82.1%). The list of all mutations found in our cohort is shown as Supplementary Data 1. We included the previous study results by Dalin et al. to indicate the mutation frequency of individual genes (Supplementary Fig. 1).

In addition, the PI3K pathway was dysregulated in 38 tumors (56.7%) and mutations in PIK3CA (29 cases), PIK3R1 (5 cases), PTEN (7 cases) and AKT1 (1 case) were found in a mutually exclusive manner. HRAS mutations were significantly concurrent with PIK3CA mutations (83.3%). Other gene mutation identified in multiple cases were ZFHX3 (8 cases), FOXA1 (6 cases), ARID1A (6 cases), and KMT2A/C/D (5, 9, 3 cases, respectively). The average TMB was 2.3 mut/Mb and two cases (2.7%) exhibited hypermutation (>10 mut/Mb).

Detection of fusion genes in SDC by RNA-seq
RNA-seq identified four cases with tyrosine kinase fusion genes (FGFR2-CCDC6, FGFR2-KIAA1217, FGFR2-NOL4, and ETV6-NTRK3) and six cases with other fusions (CHCHD7-PLAG1, FGFR1-PLAG1, MAPK14-ZFAND3, NFIK-MAST1, SLC45A3-ELK4, TMCC1-PLXND1, and ERBB2-CTTN) (Fig. 1b and Supplementary Fig. 2).

In total, driver oncogenes, which we defined as oncogenic mutations in receptor tyrosine kinase (RTK) genes or genes in the MAPK pathway, were identified in 55 cases (82.1%).

mRNA expression analysis in SDC
The expression profile of ERBB2 established by RNA-seq was compared with the copy number data obtained from the TOP panels, protein expression by immunohistochemistry, and its copy number determined by FISH, and resulted in good concordance among the four assays (Supplementary Fig. 3). Thus, RNA-seq analysis is likely robust for the evaluation of ERBB2 aberrations. Similarly, the expression of AR (p = 7×10^-4, low vs. high, Student’s t test) and EGFR (p = 0.02, 3+ vs. the others, Student’s t test) both correlated to the results obtained by immunohistochemical analysis. In contrast, CD274 encoding PD-L1 was more abundantly expressed in driver-negative compared with driver-positive tumors (p = 0.01, Student’s t test).
Fig. 2  Gene expression profile of salivary duct carcinoma. a K-means clustering analysis conducted with RNA-seq data. Clinical information (gender, age, primary site, lymph node involvement, and mutational status) is shown in the upper part. Fisher’s test identified factors associated with either group stratified by k-means clustering. b Kaplan–Meier curves of recurrence-free survival in the cohort stratified by k-means clustering as clusters 1–3. c GSEA results with the indicated gene sets differentially enriched among clusters defined by k-means clustering. Full GSEA results can be found in Supplementary Data 4.

Association of genetic profiles with clinicopathological features

Approximately 40% of the tumors in the parotid gland harbored HRAS mutations and 40% of the submandibular tumors were positive for ERBB2 gene alterations (Fig. 1c). All HRAS-mutant tumors (23 cases) were de novo PDC, whereas 50% (6/12) of the tumors from ex PA had ERBB2 gene alterations.

To evaluate whether the driver mutations were the trunk ones, the variant allele frequency (VAF) of the driver genes was compared with the highest VAF among all mutated genes in each case. For the cases positive for driver mutations of SNVs and indels, the average ratio of driver VAF to the highest VAF was 75.5 (95% CI = 12.6–134), suggesting that the driver mutations are the founding ones in SDC (Supplementary Fig. 4).

There was no significant difference in the highest VAF between the driver-positive and -negative cases (an average of 36.4% vs. 44.8%, respectively, p = 0.05, Student’s t test) (Fig. 1d). The mean age of the driver-positive cases was slightly older compared with that of driver-negative cases (63.5 years old vs. 55.2 years old, respectively, p < 0.05, Student’s t test) (Fig. 1d). All clinicopathological features are shown as Supplementary Data 2.

Transcriptional profiles of SDC

We conducted a k-means cluster analysis of the tumors based on the gene expression profiles from RNA-seq data, and compared the clusters with clinicopathological features or gene mutational profiles (Fig. 2a). Clustering using the top 100 genes with the most variation of expression divided the cohort into three groups: cluster 1 (the left cluster in Fig. 2a), cluster 2 (middle) and cluster 3 (right) (Supplementary Data 3). Female patients were enriched in cluster 2 (p < 0.01, cluster 2 vs. the others, Fisher’s exact test), whereas male patients were enriched in cluster 3 (p < 0.05, cluster 3 vs. the others, Fisher’s exact test). There was no significant difference in progression-free survival (PFS) among the clusters (Fig. 2b). The median follow-up time was 40.3 months (range, 2.4–135 months) and the median PFS was 16.4, 11, 23.6 months for cluster 1–3, respectively (95% Cl = 16–42.8 months, 6.6–41.1 months and 17.7–30.3 months). The median OS was 40.5, 36.5 and 46.5 months for cluster 1–3, respectively (95% Cl = 38.7–64.2 months, 26.2–57.8 months and 37.5–55.6 months).

Gene set enrichment analysis (GSEA) among the three clusters identified 2, 9 and 12 gene sets were upregulated in cluster 1–3, respectively (q ≤ 0.01, one cluster vs. the others). Among these gene sets, “E2F_TARGETS” and “HALLMARK_INTERFERON_RESPONSE” were enriched in cluster 2 while “HALLMARK_OXIDATIVE_PHOSPHORYLATION” was enriched in cluster 3 and “HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION” was enriched in cluster 1. The data are bulk transcriptomes and the three clusters might well reflect the microenvironment composition (Fig. 2c and Supplementary Data 4).

Expression of salivary gland marker genes including HTN1/3 and PRB1/2/3/4 encoding salivary glycoproteins was elevated specifically in cluster 1, suggesting salivary differentiation in the tumors of cluster 1 (Supplementary Data 3).
Risk assessment was conducted using RNA-seq data from 61 cases. The median follow-up time was 40.5 months (range: 10–135 months) and median OS was 40.5 months (95% confidence interval (CI) = 40.3–54.6 months). For the OS analysis, genes which satisfy the following conditions were selected: gene expression is >0 in more than 80% (49/61) of total samples and sd >1. Univariate Cox proportional hazards regression analysis showed a significant correlation between 9 genes and OS ($p < 2 \times 10^{-3}$).

To generate an OS prognostic signature, these 9 genes were used for forward conditional stepwise regression with multivariable Cox analysis in the cohort. This procedure established a prognostic model with four genes including ADAMTS1, DSC1, RNF39, and IGLL5. We constructed a risk score with the regression coefficients from this model and performed manual selection with a suitable threshold at the 75th percentile, which was −0.6 (Fig. 3a). High-risk patients, as defined by the four-gene signature-based risk score, had significantly worse OS ($p = 5 \times 10^{-7}$, log-rank test) in the cohort (Fig. 3b).

To understand the biology underlying the high-risk tumors, we identified the top 100 significantly overexpressed genes and the top 100 underexpressed genes in the high-risk tumors (Fig. 3c and Supplementary Data 3). There were no differing clinical features between the low- and high-risk groups. However, TP53 mutations were significantly enriched in the high-risk group as determined by Fisher's exact test ($p < 0.01$) (Supplementary Fig. 5). The gene sets related to cell proliferation, such as "G2M_CHECKPOINT" and "E2F_TARGETS," were significantly enriched in the high-risk tumors, whereas gene sets including "MYOGENESIS" and "TNFA_SIGNALING_VIA_NFKB" were enriched in the low-risk group (Fig. 3d and Supplementary Data 4).

Using multivariable Cox analysis, the risk score for OS remained significant (HR = 5.62, $p = 7.7 \times 10^{-3}$) in the cohort independent of gender, age, primary site, response to first-line treatment, and gene mutations (Table 2). Additional risk factors for OS by univariate analysis included response to first-line treatment (SD; HR = 17.2, $p = 6.5 \times 10^{-4}$, PD; HR = 7.94, $p = 2.1 \times 10^{-4}$), PTEN mutations (HR = 2.51, $p = 0.02$), TP53 mutations (HR = 2.74, $p = 3.0 \times 10^{-3}$), the presence of fusion genes (HR = 2.6, $p = 0.04$), and the N2b/N2c of lymph node status (HR = 2.31, $p = 0.01$); however, none of these factors were significantly different in the multivariate analysis.
Risk assessment for CAB treatment

Biomarkers related to the response to CAB were derived from the mRNA expression data of 27 cases with CAB treatment. The median follow-up time was 22.2 months (range, 4.2–61.4 months) and the median PFS for CAB treatment was 9.9 months (95% CI = 6.7–13.0 months). Genes which satisfy the following conditions were selected for the analysis: the highest gene expression among the samples is >50. Univariate Cox proportional hazards regression analysis revealed 14 genes that were significantly correlated with PFS (p ≤ 2 × 10^{-3}).

Fourteen genes were used to generate a prognostic signature using a forward conditional stepwise regression with multivariable Cox analysis for the cohort. This procedure established a prognostic model with two genes, CD3E and LDB3. We constructed a risk score with the regression coefficients from this model and conducted a manual selection of a suitable threshold at the 75th percentile (Fig. 4a). High-risk patients, as defined by the two-gene signature-based risk score, had a significantly worse PFS (p = 0.03, log-rank test) compared with the rest of the cohort (Fig. 4b). Neither the AR-V7 variant expression nor total AR expression correlated with patient response to CAB treatment (Fig. 4c and Supplementary Fig. 6).

We repeated the analysis of the differentially expressed genes using the set of patients defined by the two-gene signature to further understand the biology of high-risk tumors (Fig. 4d and Supplementary Data 3). Similar to the OS prognostic signature analyses, gene sets including "MYC_TARGETS_V1," "FATTY ACID METABOLISM" and "ANDROGEN_RESPONSE" were significantly enriched in the high-risk tumors, whereas gene sets including "ALLOGRAFT_REJECTION," "INTERFERON_GAMMA_RESPONSE" and "INFLAMMATORY_RESPONSE" were enriched in the low-risk group (Fig. 4e and Supplementary Data 4).

Immunohistochemistry of ADAMTS1, DSC1, RNF39, CD3 and LDB3 was performed to evaluate protein expression in tumor cells (Supplementary Fig. 7). The concordance between the FPKM and the protein expression intensity or the ratio of protein expression positive cells was not observed in respective genes (Supplementary Data 5).

Table 2. Cox proportional hazard models in salivary duct carcinoma.

| Variables                        | Categories | No. | Univariate analysis | Multivariate analysis |
|----------------------------------|------------|-----|---------------------|----------------------|
|                                  |            |     | Hazard ratio        | 95% CI               | P value     | Hazard ratio | 95% CI               | P value     |
| Sex                              | Female     | 10  | 0.82               | 0.38–2.18            | 0.82       |             |                      |             |
|                                  | Male       | 66  | 0.99               | 0.50–1.94            | 0.97       |             |                      |             |
| Age                              | <60        | 28  | 0.95               | 0.43–2.08            | 0.89       |             |                      |             |
|                                  | ≥60        | 48  | 2.25               | 0.54–9.43            | 0.27       |             |                      |             |
| Primary site                     | Parotid gland | 56  |                    |                      |            |             |                      |             |
|                                  | Submandibular gland | 17  |                    |                      |            |             |                      |             |
| Stage                            | I          | 7   |                    |                      |            |             |                      |             |
|                                  | II         | 13  | 0.63               | 0.11–3.80            | 0.63       |             |                      |             |
|                                  | III        | 8   | 0.87               | 0.12–6.23            | 0.89       |             |                      |             |
|                                  | IV         | 48  | 2.25               | 0.54–9.43            | 0.27       |             |                      |             |
| Results of 1st line treatment    | CR         | 62  |                    |                      |            |             |                      |             |
|                                  | PR         | 7   | 0.94               | 0.22–3.97            | 0.94       | 1.41        | 0.15–12.9           | 0.76       |
|                                  | SD         | 2   | 17.2               | 3.31–87.1            | 6.5E–04    | NA          | NA                  | NA         |
|                                  | PD         | 5   | 7.94               | 2.64–23.6            | 2.1E–04    | 3.82        | 0.63–23.3           | 0.15       |
| Mutations                        | ERBB2      | 9   | 0.29               | 0.07–1.20            | 0.088      |             |                      |             |
|                                  | HRAS       | 25  | 0.53               | 0.25–1.12            | 0.097      |             |                      |             |
|                                  | BRAF       | 5   | 0.53               | 0.13–2.20            | 0.38       |             |                      |             |
|                                  | NF1        | 9   | 1.24               | 0.44–3.53            | 0.69       |             |                      |             |
|                                  | PIK3CA     | 26  | 0.58               | 0.28–1.24            | 0.16       |             |                      |             |
|                                  | PTEN       | 11  | 2.51               | 1.17–5.35            | 0.02       | 2.11        | 0.69–6.43           | 0.19       |
|                                  | TP53       | 35  | 2.74               | 1.41–5.34            | 3.0E–03    | 1.16        | 0.44–3.07           | 0.77       |
|                                  | Other Genes| 15  | 0.66               | 0.23–1.87            | 0.43       |             |                      |             |
|                                  | Fusion     | 10  | 2.6                | 1.04–6.52            | 0.04       | 2.1         | 0.67–6.53           | 0.2        |
| CAB                              | F          | 43  | 1.62               | 0.84–3.14            | 0.15       |             |                      |             |
|                                  | T          | 32  |                    |                      |            |             |                      |             |
|                                  | Tmab/DTX   | F   | 66  | 0.33               | 0.05–2.45  | 0.28       |             |                      |             |
|                                  | Tmab/DTX   | T   | 9   |                    |                      |            |             |                      |             |
| Lymph nodes status              | N0/N1      | 41  |                    |                      |            |             |                      |             |
|                                  | N2b/N2c    | 35  | 2.31               | 1.19–4.48            | 0.01       | 1.67        | 0.69–4.07           | 0.26       |
| Risk Score (OS)                 | Low risk   | 46  | 5.99               | 2.73–13.1            | 7.8E–06    | 5.62        | 1.58–20.0           | 7.7E–03    |
|                                  | High risk  | 15  |                    |                      |            |             |                      |             |

Two-sided likelihood ratio test.

CAB combined androgen blockade, Tmab/DTX trastuzumab and docetaxel, OS overall survival, CR complete response, PR partial response, SD stable disease, PD progressive disease, CI confidence interval.
Fig. 4  Two-gene predictive signature associated with CAB treatment of SDC. a Two-gene expression and risk score distribution in the cohort by z-score. Red and light blue indicate higher and lower expression, respectively. Risk scores for all patients are plotted in ascending order and marked as low-risk (blue) or high-risk (red), as divided by the threshold (vertical black line). The risk score threshold was −4.0. b Kaplan–Meier curves of recurrence-free survival in the cohort stratified by the two-gene predictive signature in patients at high or low-risk. The log-rank test was used to evaluate p values. c Kaplan–Meier curves representing progression-free survival (PFS) in the cohort stratified by AR and AR-V7 expression. d Heatmap of the top 200 differentially expressed genes between patients at high or low-risk with red and blue indicating higher and lower expression, respectively. e Significantly different gene sets identified by GSEA as differentially overexpressed or underexpressed in high-risk tumors. Supplementary Data 4 presents the full GSEA results. NES normalized enrichment score, FDR false discovery rate.
DISCUSSION
This is the first study to identify predictive biomarkers for CAB treatment and prognostic biomarkers of SDC using molecular profiling with next-generation sequencing combined with clinicopathological data. Our analysis identified driver mutations in the RTK/ MAPK signaling pathway in 55 out of 67 tumors (82.1%). Given that the driver mutations were truncal, SDC likely depends on these driver genes for growth. There was no difference in the highest VAF of gene mutations between driver mutation-positive and -negative cases; therefore, the latter cases were the true negatives rather than the false negatives caused by low tumor content. Previous comprehensive genomic analysis of 31 SDC cases also revealed driver mutations in ~74% of the cases. Our RNA-seq analysis identified RTK fusions and ERBB2 overexpression in multiple cases. Furthermore, we precisely depicted the involvement of genes that relate to histone modification, such as KMT2A/C/D, and to transcription factors, ZFHK3 and FOXA1, whose expression was identified as a favorable prognosis in our previous study. The TMB of SDC was lower than those found in other solid tumors, which is consistent with a previous study.

Further analysis with whole-genome sequencing and long-read sequencing may be useful in deciphering the transformation mechanism of driver-negative SDC.

The utility of RNA-seq was demonstrated by the detection of fusion genes and through expression analyses. The presence of fusion genes was an independent marker for poor OS (HR = 2.6, p = 0.04) in this study. Especially, we identified targetable FGFR1/2 and NTRK3 fusions in multiple cases. It is promising that there has been recent developments and approval of targeted therapies for cholangiocarcinomas with FGFR2 fusions and for solid tumors with NTRK fusions. Evaluation of mRNA expression with RNA-seq was well-correlated with the immunohistochemical staining for HER2, AR, and EGFR. The expression of CD274 was upregulated in the driver-negative group suggesting that PD-L1 inhibitors may represent a viable treatment option for these patients.

The two-gene signature and the expression of CD3E and LDLB3, but not AR, successfully stratified patients by outcome to CAB treatment. CD3E encodes CD3-epsilon polypeptide which forms the T-cell receptor-CD3 complex and plays an important role in coupling antigen recognition to several intracellular signal transduction pathways. The immunohistochemistry of CD3 confirmed that the CD3 expression was observed in lymphocytes but tumor cells suggesting CD3E RNA expression probably reflect the tumor immune contexture. The association of infiltrating CD8+ cytotoxic T cells, as well as that of CD3+ T cells, with favorable prognosis has been widely demonstrated in cancers with different histological features and anatomical location, in both primary and metastatic settings, including melanoma, most squamous cell carcinomas, large cell lung cancer and several types of adenocarcinoma. The epsilon polypeptide is also involved in T-cell development. LDL3 encodes a PDZ domain-containing adapter protein in striated muscle to couple protein kinase C-mediated signaling through its LIM domains to the cytoskeleton. What our analysis suggested was that the T-cell infiltration may be a predictive maker for CAB treatment and it was consistent with the previous studies which investigated the immune microenvironment and neoantigen landscape in SDC.

Neither CD3E and LDL3 has been reported to be involved in the AR signaling pathway or the metabolism of dihydrotestosterone, a potent agonist of AR. Our GSEA analysis comparing high- and low-risk groups against CAB treatment revealed that the genes involved in androgen pathways are enriched in the low-risk group. This suggests that hormone signal dependency may be another predictive biomarker of the treatment.

We also defined a four-gene set (ADAMTS1, DSC1, RNF39, and IGLL5) for predicting aggressive SDC. This gene set may act as a valuable predictive biomarker to stratify patients who may benefit from additional systemic or radiation therapies. These four genes are related to multiple processes, such as inflammation and cell adhesion.

ADAMTS1 is a well-characterized matrix metalloprotease-related enzyme. However, mammalian cell studies report conflicting roles of ADAMTS1 in cancer development, and different mechanisms have been linked to its protumor/antitumor activities, including the regulation of angiogenesis, lymphangiogenesis, cell proliferation, adhesion, migration, and degradation or interactions with extracellular matrix components. For example, ADAMTS1 is downregulated in prostate and colorectal cancers by promoter hypermethylation, but is overexpressed in pancreatic cancer, fibrosarcoma, and renal cell carcinoma. Presently, the role of ADAMTS1 in SDC progression remains unclear.

DSC1 is a desmosomal cadherin that has an important role in cell–cell adhesion. These cadherins include DSC1–3 and four desmogleins. DSC1 was reported to be involved in progression, metastatic potential, and cell adhesion processes in breast cancer. Currently, little is known regarding the involvement of RNF39 and IGLL5 in tumorigenesis.

The GSEA analyses identified upregulated expression of genes involved in the G2/M checkpoint, cell cycle-related targets of E2F transcription factors, and epithelial–mesenchymal transition (EMT) in the high-risk group, whereas upregulated genes in the low-risk group were involved in myogenic differentiation. During EMT, epithelial cells lose their cell polarity and cell–cell adhesion, and gain migratory and invasive properties to become mesenchymal stem cells. The enriched gene set related to the cell cycle and EMT may indicate the capability of rapid cell growth, high invasion, and migration of aggressive SDC. Upregulated ADAMTS1 and DSC1 expression may have an important role in the survival and invasion of SDC cells, and may explain their association in patients with poor prognosis. Considering that mutations in TP53 or PTEN are also independent markers for poor prognosis, this combination of genetic and epigenetic changes may promote the progression and recurrence of SDC.

The discordance between protein expression and RNA-seq data is because the sensitivity of RNA-seq to quantify RNA expression might be higher compared with immunohistochemistry to quantify protein expression. In fact, the protein expression levels of DSC1, LDL3 and RNF39 in most samples were too low to be detected by immunohistochemistry while ADAMTS1 expression was moderately positive in all samples. The CD3 protein expression was observed in lymphocytes but tumor cells suggesting CD3E RNA expression probably reflect the tumor immune contexture. Although the concordance was not observed in these genes, we are confident that our RNA-seq is suitable for the evaluation of RNA expression as good correlations were observed in HER2, AR and EGFR expression (Supplementary Fig. 3). RNA-seq may be preferable method to evaluate expression of genes whose protein level are low or when the immunohistochemical analysis is not well optimized.

A major limitation of this study is that this is a retrospective study without area under the curve assessment supporting the specificity of the biomarker signature. Therefore, prospective studies of independent cohorts should be conducted to validate the proof of concept. Another limitation is about TOP panel that it can’t detect all the mutations exiting in the cancer and doesn’t necessarily elucidate all tumor features. Therefore, comprehensive assays such as whole-genome sequencing and genome wide methylation analysis are needed to identify additional novel biomarkers.

In conclusion, our discoveries may have the potential for application in a clinical setting. Expression profiling may directly predict the efficacy of CAB treatment and the prognosis of SDC. Our genomic and transcriptomic analyses highlight the importance of precise tumor profiling to provide increased treatment options for SDC patients.
METHODS
Study design and patient specimens
The study cohort consisted of 76 SDC patients who underwent surgical resection between October 2005 and September 2017 at hospitals in the Japan SDC consortium, including the International University of Health and Welfare, Mita Hospital, Tokyo Medical University Hospital, Tokyo Medical University Hachioji Medical Center, Hokkaido University, Niigata Cancer Center Hospital, Keio University, and Tokai University. Nine patients were excluded because of poor quality DNA and the analysis was conducted in the remaining 67 patients. Several cases were previously reported by Otsuka et al.9, Takase et al.16, Masubuchi et al.17, Shimura et al.41, Fushimi et al.20, Takahashi et al.29, and Okada et al.63. A board-certified pathologist (TN) specializing in salivary gland tumors reviewed the SDC histological features based on the criteria of the current World Health Organization classification. Fresh frozen specimens or formalin-fixed paraffin-embedded (FFPE) of surgically resected tumors were obtained from all patients. Approval for this study was obtained from the Ethics Committee of National Cancer Center (No. 2019-271), International University of Health and Welfare, Mita Hospital (No. 5-19-6), Tokyo Medical University (No. SH2563), Faculty of Medicine and Graduate School of Medicine, Hokkaido University (No. 017-0487), Keio University School of Medicine (No. 20120083), School of Medicine, Tokai University (No. 20R-204), Tokyo Medical University Hachioji Medical Center (No. SH2563) and Niigata Cancer Center Hospital (No. 2021-300). All subjects provided written informed consent, except for those who could not be reached because of loss of follow-up or death at registration. In these cases, the Institutional Review Board at each participating institution granted permission for the existing tissue samples to be used for research purposes. No samples from the patients who had opted out of participation were used in this study.

Immunohistochemistry and fluorescence in-situ hybridization (FISH)
Immunohistochemistry and FISH were conducted as previously described16,41. Briefly, FFPE tumor tissues were cut into 3 µm-thick sections, and a polymer-based detection system with heat-mediated antigen retrieval was used with primary antibodies against HER2 (Polyclonal, cat# A0485, Agilent Technologies, Santa Clara, CA) with 1:400 dilution, AR (clone AR441, cat# PM109AA, BIOCARE Medical LLC, CA), EGFR (clone 31G7, cat# 423701, NICHIREI BIOSCIENCES INC, Tokyo, Japan) in undiluted form, PD-L1 (clone 22C3, cat# M3653, Agilent Technologies) with 1:50 dilution, DSC1 (clone A-4, cat# sc-398590, Santa Cruz Biotechnology, Dallas, TX) with 1:100 dilution, RNF39 (cat# HPA047115, Atlas Antibodies, Bromma, Sweden) with 1:200 dilution, CD3 (clone SP-7, cat# 413601, Nichirei Biosciences, Tokyo, Japan) with 1:100 dilution, and LB33 (cat# 11004-1-AP, Proteintech) with 1:100 dilution. Diaminobenzidine was used to detect antigen-antibody reactions. Appropriate positive and negative controls were applied for all conditions. ERBB2 amplification was evaluated using PathVysion HER-2 DNA Probe Kit (cat# 02J01-031, Abbott Molecular, Des Plaines, IL). For the PD-L1 IHC 22C3 pharmDX assay, the slides were stained using a Dako Autostainer Link 48 platform with an automated staining protocol and the pathologists A. U. and T. H. used a light microscope to score the percentage of positive tumor cells in each sample.

Evaluation of HER2 status
HER2 positivity was defined as either immunohistochemically 3+ or ERBB2 amplification according to the American Society of Clinical Oncology/College of American Pathologists guidelines for breast cancer64. In Her2 Immunohistochemistry, 3+ stain was defined as a circumferential membrane staining that was complete and intense and exhibited >10% frequency of tumor cells. For ERBB2 copy number prediction with FISH analysis, 100 non-overlapping, intact interphase tumor nuclei stained with 4’6-diamidino-2-phenylindole (DAPI) were evaluated. The ERBB2 gene (red signal) and CEP17 (green signal) copy numbers in each nucleus were assessed. ERBB2 amplification is annotated when the average ERBB2/CEP17 ratio was ≥2.0 in all nuclei or when the ERBB2 signals formed a tight gene cluster.

Assessment of immunohistochemistry of AR, EGFR, ADAMTS1, DSC1, RNF39, CD3 and LDB3
A case was considered positive for AR when ≥20% of tumor cell nuclei exhibited strong staining. The percentage of EGFR immunostaining cells was scored from 0 to 3+ as follows: 0; 0%; 1+, 1–10%; 2+, 11–30%; and 3+, >30%. Score 3+ was considered positive for EGFR. For ADAMTS1, DSC1, RNF39, CD3 and LDB3, staining intensity and ratio of positive cells were assessed. The staining intensity was scored from 0 to 3+ as follows: 0; negative; 1+, mildly positive; 2+, moderately positive; 3+, strongly positive.

DNA sequencing with the TOP cancer gene panel for mutation call and copy number analysis
Fresh frozen or FFPE SDC specimens were analyzed with TOP panel version 344. This evaluates nucleotide variants and insertions/deletions for 464 genes to calculate TMB and to infer copy number variation. Genomic DNA was extracted from FFPE samples using the GeneRead DNA FFPE Kit (Qiagen, Hilden, Germany) and 500 ng of each DNA sample was used for target fragment enrichment with an Agilent Kit (v6) (Agilent Technologies). Next-generation sequencing was conducted with a HiSeq2500 (Illumina) using a paired-end option and sequencing reads were independently aligned to the human reference genome (hg38) using BWA65, Bowtie2 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml), and NovoAlign (http://www.novocraft.com/products/novALIGN). Somatic mutations were called using MuTect (http://www.broadinstitute.org/cancer/cga/mutect), SomaticIndelDetector (http://www.broadinstitute.org/cancer/cga/node/87), and VarScan (http://varscan.sourceforge.net). The exclusion criteria for mutation analysis were (i) the read depth was <20 or the VAF was <0.1; (ii) mutations were supported by only one strand of the genome; or (iii) they were SNP in either the 1000 Genomes Project dataset (http://www.internationalgenome.org/) or our in-house database. Gene mutations were annotated by Snpeff (http://snpeff.sourceforge.net). Copy number was evaluated using our in-house pipeline to calculate the log2 ratio (LRR) as follows: (i) used selected homozygous (VAF ≤ 0.05 or ≥ 0.95) or a heterozygous (VAF, 0.4–0.6) SNP in the 1000 Genomes Project database; (ii) normal and tumor read depths at the selected SNP position were counted and adjusted based on GpC percentage around 100 bp from the position66; (iii) calculated the LRR = log2 (t/n), where n and t are the normal and tumor-adjusted depths at the SNP position; and (iv) each representative LRR was determined by the median of a moving window (1 Mb) around the SNP position. The values of the LRR of the copy number of the major allele and the minor allele were determined for every region of the entire genome. The p values for the gain or loss of respective genomic regions were determined from the LRRs with a permutation test (100,000 iterations) according to the algorithm used in GISTIC67, and q values were calculated using the R package qvalue (http://github.com/jdstorey/qvalue).

Transcriptome sequencing, expression analysis, and detection of fusion genes and exon skipping
Total RNA was extracted from fresh, frozen, or FFPE samples using RNA-Bee (Tel-Test Inc., Gainesville, FL) and treated with DNase I.
Survival analysis and Cox regression analyses were conducted using the "survival" (v2.44.1.1) package. OS and PFS were analyzed using the Kaplan–Meier method and curve differences were evaluated using the log-rank test according to either the risk score or the driver mutation subtypes. GSEA was conducted using Java GSEA software (http://software.broadinstitute.org/gsea/index.jsp) (v2.2.4).

Reporting summary
Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY
We have deposited the raw sequencing data of all new and unpublished DNA and RNA sequencing data presented in this paper under accession number hum0094 in the Japanese Genotype-Phenotype Archive which is hosted by the DNA Data Bank of Japan, under accession number JGAS000534.

CODE AVAILABILITY
No unreported custom computer code or algorithm was used to generate the results of this paper. The references of all source codes are included within Materials and Methods.

Received: 3 September 2021; Accepted: 18 October 2022; Published online: 04 November 2022

REFERENCES
1. Nagao, T., Licitra, L., Loening, T., Vielh, P. & Williams, M.D. Salivary duct carcinoma in WHO Classification of Head and Neck Tumours, 4th edition (eds I. Naggar, A.K., Chan, J.K.C., Grandis, J.R., Takata, T. & Sloomweg, P.J.) 173–174 (IARC, 2017).
2. Nakaguro, M. et al. Salivary duct carcinoma: Updates in histology, cytology, molecular biology, and treatment. Cancer Cytopathol. 128, 693–703 (2020).
3. D.Ehryge, E., Meulemans, J. & Vander Poorten, V. Salivary duct carcinoma. Curr. Opin. Otolaryngol. Head. Neck Surg. 26, 142–151 (2018).
4. Boukhers, H., Curtis, R. E., Land, C. E. & Dores, G. M. Incidence of carcinoma of the major salivary glands according to the WHO classification, 1992 to 2006: a population-based study in the United States. Cancer Epidemiol. Biomark. Prev. 18, 2899–2906 (2009).
5. Tamaki, T. et al. The burden of rare cancer in Japan: application of the RARECARE definition. Cancer Epidemiol. 38, 490–495 (2014).
6. Jayaprakash, V. et al. Survival rates and prognostic factors for infiltrating salivary duct carcinoma: Analysis of 228 cases from the Surveillance, Epidemiology, and End Results database. Head. Neck 36, 694–701 (2014).
7. El-Naggar AK, C.J., et al (eds). Salivary duct carcinoma, in WHO Classification of Head and Neck Tumours, 4th edition (eds l-Naggar, A.K., Chan, J.K.C., Grandis, J.R., Takata, T. & Sloomweg, P.J.) 173 (IARC, Lyon, France, 2017).
8. Boon, E. et al. Androgen deprivation therapy for androgen receptor-positive advanced salivary duct carcinoma: A nationwide case series of 35 patients in The Netherlands. Head. Neck 40, 605–613 (2018).
9. Otsuka, K. et al. Clinical Outcomes and Prognostic Factors for Salivary Duct Carcinoma: A Multi-Institutional Analysis of 141 Patients. Ann. Surg. Oncol. 23, 2038–2045 (2016).
10. Gilbert, M. R. et al. A 20-Year Review of 75 Cases of Salivary Duct Carcinoma. JAMA Otolaryngol. Head. Neck Surg. 142, 489–495 (2016).
11. Osborn, V. et al. Characterization, treatment and outcomes of salivary ductal carcinoma using the National Cancer Database. Oral. Oncol. 71, 41–46 (2017).
12. Williams, M. D. et al. Differential expression of hormonal and growth factor receptors in salivary duct carcinomas: biologic significance and potential role in therapeutic stratification of patients. Am. J. Surg. Pathol. 31, 1645–1652 (2007).
13. Hosal, A. S., Fan, C., Barnes, L. & Myers, E. N. Salivary duct carcinoma. Otolaryngol. Head. Neck Surg. 129, 720–725 (2003).
14. Nasser, S. M., Faquin, W. C. & Dayal, Y. Expression of androgen, estrogen, and progesterone receptors in salivary gland tumors. Frequent expression of androgen receptor in a subset of malignant salivary gland tumors. Am. J. Clin. Pathol. 108, 801–806 (2003).
15. Fan, C. Y., Melhem, M. F., Hosal, A. S., Grandis, J. R. & Barnes, E. L. Expression of androgen receptor, epidermal growth factor receptor, and transforming growth factor alpha in salivary duct carcinoma. Arch. Otolaryngol. Head. Neck Surg. 127, 1075–1079 (2001).
10. Takahashi, H. et al. Phase II Trial of Trastuzumab and Docetaxel in Patients With Metastatic Salivary Duct Carcinoma: A Multicenter, Open-label, Phase 2 Study. Lancet Oncol. 21, 671–684 (2020).

11. Drilon, A. et al. Efficacy of Larotrectinib in TRK Fusion-Positive Cancers in Adults and Children. N. Engl. J. Med. 378, 731–739 (2018).

12. Laetsch, T. W. et al. Larotrectinib for paediatric solid tumours harbouring NTRK gene fusions: phase 1 results from a multicentre, open-label, phase 1/2 study. Lancet Oncol. 19, 705–714 (2018).

13. Doebele, R. C. et al. Entrectinib in patients with advanced or metastatic NTRK fusion-positive solid tumours: integrated analysis of three phase 1–2 trials. Lancet Oncol. 21, 271–282 (2020).

14. Barber, E. K., Dasgupta, J. D., Schlossman, S. F., Treviñán, J. M. & Rudd, C. E. The CD4 and CD8 antigens are coupled to a protein-tyrosine kinase (p56lck) that phosphorylates the CD3 complex. Proc. Natl Acad. Sci. USA 86, 3277–3281 (1989).

15. Galon, J. et al. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. Science 313, 1606–1609 (2006).

16. Friedman, W. H., Pages, F., Sauver-Friedman, C. & Galon, J. The immune contexture in human tumours: impact on clinical outcome. Nat. Rev. Cancer 12, 298–306 (2012).

17. Friedman, W. H., Zitvogel, L., Sauver-Friedman, C. & Kroemer, G. The immune contexture in cancer prognosis and treatment. Nat. Rev. Clin. Oncol. 14, 717–734 (2017).

18. Bruni, D., Angell, H. K. & Galon, J. The immune contexture and Immunoscore in cancer prognosis and therapeutic efficacy. Nat. Rev. Cancer 20, 662–680 (2020).

19. Pathak, P. et al. Myophagitis associated LDB3 mutation causes Z-disc disassembly and protein aggregation through PKCalpha and TSC2-mTOR downregulation. Commun. Biol. 4, 355 (2021).

20. Cal, S. & Lopez-Otin, C. ADAMTS proteases and cancer. Matrix Biol. 44–46, 77–85 (2015).

21. Tan Ide, A., Ricciardelli, C. & Russell, D. L. The metalloprotease ADAMTS1: a comprehensive review of its role in tumorigenic and metastatic pathways. Int. J. Cancer 133, 2263–2276 (2013).

22. Masui, T. et al. Expression of METH-1 and METH-2 in pancreatic cancer. Clin. Cancer Res. 7, 3437–3443 (2001).

23. Casal, C. et al. ADAMTS1 contributes to the acquisition of an endothelial-like phenotype in plastic tumor cells. Cancer Res. 70, 4676–4686 (2010).

24. Wen, Y. C. et al. Melatonin-triggered post-transcriptional and post-translational modifications of ADAMTS1 coordinately retard tumorigenesis and metastasis of renal cell carcinoma. J. Pineal Res. 69, e12666 (2020).

25. Faktor, J. et al. Proteomics Identification and Validation of Desmocollin-1 and Catechol-O-Methyltransferase as Proteins Associated with Breast Cancer Cell Migration and Metastasis. Proteomics 19, e1900073 (2019).

26. Kitajima, Y. New insights into desmosome regulation and pempigus blistering as a desmosome-remodeling disease. Kaohsiung J. Med. Sci. 29, 1–13 (2013).

27. Okada, T. et al. Carboplatin and Docetaxel in Patients With Salivary Gland Carcinoma: A Retrospective Study. Vivo 33, 843–853 (2019).

28. Wolff, A. C. et al. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/Clinical College of American Pathologists clinical practice guideline update. J. Clin. Oncol. 31, 3997–4013 (2013).

29. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760 (2009).

30. Kohsaka, S. et al. Identification of Novel CD74-NRG2alpha Fusion From Comprehensive Profiling of Lung Adenocarcinoma in Japanese Never or Light Smokers. J. Thorac. Oncol. 15, 948–961 (2020).

31. Merdel, C. H. et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. 12, R41 (2011).

32. Kohsaka, S. et al. Comprehensive molecular and clinicopathological profiling of desmoid tumours. Eur. J. Cancer 145, 109–120 (2021).
ACKNOWLEDGEMENTS
The authors would like to thank Y. Yamamoto and A. Maruyama-Shiino for technical assistance. This study was supported by the grants from the Project for Cancer Research And Therapeutic Evolution (P-CREATE) under grant number JP20cm0106502, and the Practical Research for Innovative Cancer Control under grant number JP20ck0106536 from the Japan Agency for Medical Research and Development, AMED. This work was also supported by the JSPS Grants-in-Aid for Scientific Research (C) under grant number 18K09386 and 20K07417.

AUTHOR CONTRIBUTIONS
S.K., Y.T., and H.M. conceived the project and designed the study. S.K., T.U., and H.M. developed the methodology. S.K. and Y.S. performed the experiments. S.K, T.U., and S.K. analyzed and interpreted the data. Y.T., M.A., M.N., T.N., and H.M. wrote and edited the paper with feedback from all authors.

COMPETING INTERESTS
The authors declare no competing interests.

ADDITIONAL INFORMATION
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41698-022-00324-1.

Correspondence and requests for materials should be addressed to Shinji Kohsaka, Yuichiro Tada or Hiroyuki Mano.

Reprints and permission information is available at http://www.nature.com/reprints

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

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

© The Author(s) 2022