

A single-nucleotide-polymorphism in the 5' - flanking region of MSX1 gene as a predictive marker candidate for platinum-based therapy of esophageal carcinoma

Mori, Takahiro

Departments of Clinical Oncology and Gastroenterological Surgery, National Hospital Organization Sagamihara National Hospital

Ueno, Kazuko

Genome Medical Science Project, National Center for Global Health and Medicine

Tokunaga, Katsushi

Genome Medical Science Project, National Center for Global Health and Medicine

Kawai, Yosuke

Genome Medical Science Project, National Center for Global Health and Medicine

他

<https://hdl.handle.net/2324/7177146>

出版情報 : Therapeutic Advances in Medical Oncology. 14, 2022-01. SAGE Publications

バージョン :

権利関係 : © The Author(s), 2022



A single-nucleotide-polymorphism in the 5'-flanking region of *MSX1* gene as a predictive marker candidate for platinum-based therapy of esophageal carcinoma

Takahiro Mori¹, Kazuko Ueno, Katsushi Tokunaga, Yosuke Kawai, Koichi Matsuda, Nao Nishida, Keigo Komine, Sakae Saito, Masao Nagasaki and The BioBank Japan Project

Abstract

Background: Platinum derivatives are important treatment options for patients with esophageal carcinoma (EC), and a predictive marker for platinum-based therapy is needed for precision medicine.

Patients and methods: This study contained two cohorts consisting of EC patients treated using platinum-based chemoradiation therapy (CRT) as the first-line and another external cohort of nationwide clinicogenomic data from the BioBank Japan (BBJ).

Results: Genome-wide association study (GWAS) of therapeutic outcomes, refractory disease or not, following platinum-based CRT as first-line in 94 patients in the first cohort suggested the association of 89 SNPs using $p < 0.0001$. The top 10 SNPs selected from each chromosomal region by odds ratio were evaluated for progression-free survival (PFS) and overall survival (OS) hazard ratios in the first cohort, resulting in four candidates ($p < 0.0025$). The four selected candidates were re-evaluated in another cohort of 24 EC patients, which included patients prospectively enrolled in this study to fulfill the sample size statistically suggested by the results of the first cohort, and of the four, only rs3815544 was replicated ($p < 0.0125$). Furthermore, this candidate genotype of rs3815544 proceeded to the re-evaluation study in an external cohort consisting of EC patients treated with platinum derivatives and/or by radiation therapy as the first-line treatment in BBJ, which confirmed that the alternative allele (G) of rs3815544 was statistically associated with non-response (SD or PD) to platinum-based therapy in EC patients (odds ratio = 1.801, $p = 0.048$). The methylation QTL database as well as online clinicogenomic databases suggested that the region including rs3815544 may regulate *MSX1* expression through CpG methylation, and this down-regulation was statistically associated with poor prognosis after platinum-based therapies for EC.

Conclusion: rs3815544 is a novel candidate predictive marker for platinum-based EC therapy.

Keywords: esophageal carcinoma, *MSX1*, platinum-based therapy, predictive marker, single nucleotide polymorphism

Received: 19 August 2021; revised manuscript accepted: 28 January 2022.

Introduction

Esophageal carcinoma (EC) is one of the leading causes of neoplastic death worldwide.^{1,2} Currently,

multimodalities consisting of neoadjuvant chemotherapy (ChT) or chemoradiation therapy (CRT) followed by surgery have become the gold

Ther Adv Med Oncol

2022, Vol. 14: 1–21

DOI: 10.1177/
17588359221080580

© The Author(s), 2022.
Article reuse guidelines:
sagepub.com/journals-
permissions

Correspondence to:

Takahiro Mori
Departments of
Clinical Oncology and
Gastroenterological
Surgery, National Hospital
Organization Sagamihara
National Hospital, 18-1
Sakuradai, Minami-ku,
Sagamihara 252-0392,
Kanagawa, Japan

Laboratory of Tumor
Immunology, Clinical
Research Center, National
Hospital Organization
Sagamihara National
Hospital, Sagamihara,
Japan

Genome Medical Science
Project, National Center
for Global Health and
Medicine, Tokyo, Japan
westendc7m@gmail.com

Kazuko Ueno
Katsushi Tokunaga
Yosuke Kawai
Genome Medical Science
Project, National Center
for Global Health and
Medicine, Tokyo, Japan

Department of Human
Genetics, Graduate School
of Medicine, The University
of Tokyo, Tokyo, Japan

Koichi Matsuda
Laboratory of Clinical
Genome Sequencing,
Department of
Computational Biology
and Medical Sciences,
Graduate School of
Frontier Sciences, The
University of Tokyo, Tokyo,
Japan

Nao Nishida
Genome Medical Science
Project, National Center
for Global Health and
Medicine, Ichikawa, Japan

Keigo Komine
Department of Medical
Oncology, Tohoku
University Hospital,
Sendai, Japan

Sakae Saito
Department of Integrative
Genomics, Tohoku Medical
Megabank Organization,
Tohoku University, Sendai,
Japan

Masao Nagasaki

Center for the Promotion of Interdisciplinary Education and Research, and Center for Genomic Medicine, Graduate School of Medicine, Kyoto University, Kyoto, Japan

The BioBank Japan

Project

The Institute of Medical Science, The University of Tokyo, Tokyo, Japan

standard treatment for EC.^{3,4} Alternatively, definitive CRT has become another powerful treatment for curability because some patients treated with neoadjuvant CRT followed by surgery achieved pathological complete remission (pCR).⁵ In fact, comparable prognoses after definitive CRT to surgery alone have been reported by a couple of independent prospective studies.^{6,7} Thus, CRT is now the most important treatment option for EC, regardless of whether it is combined with surgery or not. Platinum derivatives such as cisplatin have played important roles in CRT.^{5,8} Even for patients with distant metastases, who are commonly encountered clinically, although radiation therapy is often omitted in those, platinum derivatives are still important treatment options.⁹ Therefore, it is necessary to identify predictive markers for platinum derivatives.

Quite a few studies have attempted to identify factors predictive of platinum-based ChT/CRT efficiency, mostly by molecular-profiling endoscopic biopsies before treatment, but have not been as successful.¹⁰ Actors such as intratumoral heterogeneity cause difficulties in identifying predictive markers using biopsy samples.¹¹ The subclonal nature of genomic driver mutations in tumor contributes to important clinical implications that may not necessarily be observed in a single biopsy.¹² Therefore, some studies have suggested several genetic polymorphisms as predictive markers for platinum-based therapies using candidate-gene or candidate-polymorphism approaches.^{13–17} However, the whole genome approach may be better because responsible genes or precise molecular pathways with anti-tumor effects have not yet been completely unveiled.¹¹

Therefore, we assessed the effectiveness of genome-wide association studies (GWAS) to unveil noble single-nucleotide polymorphisms (SNPs) as predictive markers for platinum-based therapy. For this purpose, we recently developed a population-optimized SNP array ‘Japonica SNP-array’ for a GWAS using resequencing data from 2,000 Japanese individuals.¹⁸ This Japonica SNP array allowed us to catalog more comprehensive genetic polymorphisms in terms of predictive markers for platinum-based therapy, especially in the Japanese population. This study used the whole-genome approach in the first cohort, followed by candidate-SNP approach in

the second cohort including prospective recruitment of patients, before re-evaluating a candidate-polymorphism in an external cohort obtained from the nationwide clinicogenomic database, the BioBank Japan (BBJ).^{19,20} We selected EC patients treated with platinum-based CRT for the two independent cohorts to make the treatment background as uniform as possible; the second cohort was designed and prepared based on the survival hazard ratios of the first cohort. On the other hand, the cohorts from BBJ consisted of clinicogenomic data from a number of hospitals from all over Japan, in which EC patients were treated by various therapeutic combinations with platinum derivatives and/or radiation therapies and reflected real-world clinical practice. The candidate SNP suggested by the two aforementioned cohorts was re-evaluated for platinum-based therapies as well as for radiation therapy, using this nationwide clinicogenomic data. We also tried to unveil the molecular and pathogenic functions induced by this candidate SNP by referring to online published multiomics databases. The present study followed the guidelines of the reporting recommendations for tumor marker prognostic studies (REMARK).^{21,22}

Patients and methods

Ethical approval and consent to participate in two cohorts

This study contained two cohorts. The first cohort consisted of all of 94 patients who provided written informed consent to participate in this study between April 2003 and October 2005 at Tohoku University Hospital. The second cohort consisted of 24 patients, 12 of whom were enrolled at Tohoku University Hospital between December 2005 and June 2016, whereas the other half were enrolled at National Hospital Organization Mito Medical Center between May 2006 and February 2008. This study was approved by the Tohoku University Graduate School of Medicine IRB (permission number 2003-003, 2005-124, 2016-1-331), National Hospital Organization Mito Medical Center IRB (permission 20060529), National Hospital Organization Sagami National Hospital (permission 2018-051) and National Center for Global Health and Medicine (permission NCGM-A-003267). Written informed consent was obtained from all enrolled patients by a board-certified esophageal surgeon (TM).

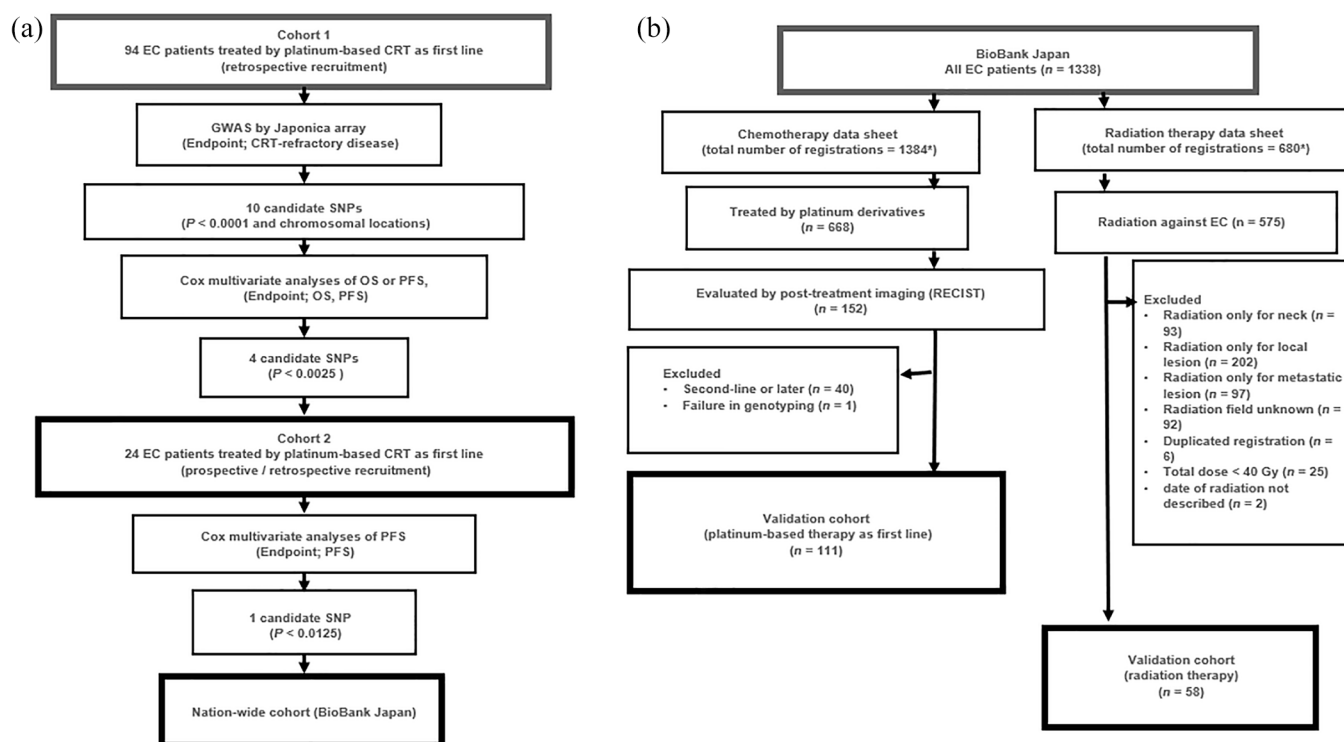


Figure 1. The flow diagrams of this study. (a) The flow diagram from two cohorts to the BBJ cohort. This study contained two cohorts. The first cohort consisted of 94 patients treated using platinum-based CRT as first-line. GWAS for screening of potential candidate SNPs for prognosis outcome, CRT-refractory disease or not, was applied to the first cohort, followed by OS/PFS analyses. The four candidate SNPs identified in the first cohort were examined in the second cohort recruited after the first cohort. Only rs3815544 was replicated by the candidate-SNP approach. (b) The flow diagram of the preparation of the external BBJ cohorts from nationwide clinicogenomic data. The clinicogenomic data from all EC patients enrolled in the BBJ were used. There were multiple registrations for chemotherapy data sheets or for radiation data sheets from a single patient because each registration was required for each chemotherapy regimen or for irradiation treatment for each lesion, as indicated by asterisks.

All the authors confirmed that this study was performed in accordance with the Declaration of Helsinki.

Patient characteristics of the first and the second cohorts

Both cohorts consisted of patients with EC who were primarily treated with definitive CRT (Figure 1(a), Table 1). We selected the second cohort after the whole genome SNP association study was conducted, which eventually enrolled 24 patients with EC primarily treated using CRT, including 6 patients who were prospectively recruited to fulfill the requirements of the study design and the sample size as calculated based on the results obtained by the first cohort; the number of enrolled patients, 24, in the second cohort was suggested by ‘Two Arm Survival’ of The

Cancer Research and Biostatistics (CRAB) statistic (<https://stattools.crab.org/Calculators/twoArmSurvival.html>) using the following: $\alpha = 0.05$, PFS hazard ratio = 5, proportion in standard group = 0.15, power = 0.8 for differences ($\beta = 0.2$), follow-up years = 10, and survival years of standard group = 8 years, based on the results for PFS by the first cohort. Both cohorts were Japanese and differed only in the dates of patient enrolment, as described above. The clinical and pathologic data of the enrolled patients are summarized in Table 1. We did not observe any familial relationships between the 118 patients.

All patients were followed up until the date of death or for as long as possible. Clinical stages were determined based on the International Union Against Cancer (UICC) TNM classification of malignant tumors, 7th edition.²³

Table 1. Patients' characteristics of the two cohorts in this study.

	First cohort	Second cohort
Number of enrolled patients	94	24
Recruitment period	April 2003 – October 2005	December 2005 – June 2016
Age (average \pm standard deviation; minimum–max)	63.4 \pm 9.0; 41 – 81	66.4 \pm 7.3; 56–79
Sex	male 87, female 7	male 21, female 3
Pathological type	squamous cell carcinoma 91, malignant melanoma 1, adenocarcinoma 1, carcinosarcoma 1	squamous cell carcinoma 23, neuroendocrine carcinoma 1
Clinical stage (UICC 7th edition)	Stage 1A/1B, 19 cases: Stage 2A/2B, 19 cases: Stage 3A/3B/3 C, 41 cases: Stage 4, 15 cases	Stage 1A/1B, 8 cases: Stage 2A/2B, 5 cases: Stage 3A/3B/3 C, 10 cases: Stage 4, 1 case
Radiation dose (Gy)	Mean 61.3 (30–70.2)*	Mean 59.2 (30–70)
Concurrent chemotherapy with platinum-derivatives	85**	22
Prognosis (survive/ EC-cause specific death/ death from other illness or unknown reasons)	36/ 37/ 21	8/ 13/ 3
CRT-refractory disease (no-refractory / refractory)	56/ 38	7/17

CRT, chemoradiation therapy; EC, esophageal carcinoma; UICC, International Union Against Cancer.
Data unavailable in four cases (*) or three cases (**) in the first cohort.

SNP genotyping and imputation

Genomic DNA was isolated from the peripheral white blood cells of each patient by using a publicly known procedure after obtaining written informed consent. All samples were genotyped using the Japonica Array v1 (Toshiba Corp, Tokyo, Japan). We conducted genotype calling using Affymetrix Power Tools (version 1.18.2; Thermo Fisher Scientific Inc., Waltham, MA). All samples met the manufacturer's quality control criteria (dish QC ≥ 0.82 and sample call rate $\geq 97\%$). SNPs that were categorized as 'Recommended' by the SNPfilter package (version 1.5.2; Thermo Fisher Scientific Inc.) were used for subsequent analyses. No cryptic relatives were found among the samples using the Maximum Unrelated Set Identification (IMUS) method implemented in PRIMUS v1.8.0²⁴ with default settings. Haplotype phasing was conducted using SHAPEIT (v2.r644)²⁵ as a pre-phasing for the genotype imputation after filtering SNPs with a call rate $< 97.0\%$, and a Hardy–Weinberg equilibrium test result of $p < 10^{-6}$, or minor allele frequency (MAF) $< 0.5\%$. Genotype imputation was performed using IMPUTE2 (ver. 2.3.1)²⁶ with the options -Ne 2000, -k_hap 1000, -k 120, -burnin 15, and -iter 50 using a phased

reference panel of 1,070 Japanese individuals.^{27,28} Each genotype was accepted if the highest genotype probability was higher or greater than 0.9. Otherwise, the genotype was treated as missing. Finally, SNPs in Hardy–Weinberg equilibrium with a p -value < 0.0001 , call rate < 0.990 , or MAFs < 0.01 were filtered out, and 4,095,500 SNPs were evaluated in the downstream analysis.

Statistical analysis

A genome-wide association test for the first screening of potential candidate SNP markers of prognosis outcome, CRT-refractory disease ($n = 38$) or not ($n = 56$), was applied to the imputed SNPs. We conducted the logistic regression analysis with sex as a covariate using Plink (version 1.9).²⁹ SNPs with p -values < 0.0001 were used in subsequent analyses as candidate SNPs to avoid losing potential candidates that might have an estimated P -value no less than GWAS significance ($5.0E-8$), since the sample size of this study was considered.

After a GWAS, we examined the hazard ratios (HRs) for PFS and OS by Cox univariate regression analysis including age at CRT, sex, radiation

doses and clinical stages in the first cohort, and only age at CRT was statistically associated with HRs for PFS and OS (data not shown). We then incorporated age at CRT and each SNP into a Cox proportional hazard model to determine HRs for PFS and OS.

Survival analysis using the Kaplan–Meier method and hazard ratio evaluation using a Cox proportional hazard model were performed using SPSS (version27) (IBM, Armonk NY, USA).

The EC cohorts from the BioBank Japan (BBJ)

We selected all EC patients that were treated with platinum-based ChT/CRT as first-line therapy and evaluated by the Response Evaluation Criteria In Solid Tumor (RECIST) version 1.1 of post-treatment imaging (platinum-based therapy cohort as first-line) or those that were treated by radiation therapy with radiation doses no less than 40 Gy (radiation therapy cohort as first-line), from all of 1,338 EC patients in the BBJ project, as shown in Figure 1(b). We adopted tumor response estimated by RECIST version 1.1 as the endpoint of clinical outcome for the platinum-based cohort because there are missing values in long-term prognosis; however, therapeutic efficiency was precisely evaluated by post-treatment imaging according to RECIST in this BBJ-EC cohort. In contrast, the therapeutic outcome of radiation therapy was estimated by the 1-year, 2-year and 3-year survivals due to unavailability of RECIST evaluations in clinical data sheets for radiation therapy in BBJ.

Association between rs3815544 and methylation status of MSX1 gene CpG sites

We used the iMETHYL database published online (<http://imethyl.iwate-megabank.org>) to evaluate the contribution of the candidate SNP, rs3815544, to the methylation status of CpG sites around the *MSX1* gene. The database provides whole-DNA methylation data (almost 24 million autosomal CpG sites) obtained from whole-genome bisulfite sequencing of CD4 + T lymphocytes and monocytes collected from a cohort of healthy subjects of Japanese ethnicity.³⁰

Evaluation of MSX1 gene features using online published databases

The data used for the analyses were obtained from the following four portals: Wanderer ([\[maplab.cat/wanderer\]\(http://www.maplab.cat/wanderer\)\),³¹ a tool that helps to visualize genomic and epigenomic data from The Cancer Genome Atlas \(TCGA\) Research Network \(<http://cancergenome.nih.gov/>\) was used for CpG methylation around the *MSX1* gene in ECs and normal esophageal mucosae. The Gene Expression Omnibus \(GEO\) portal site \(<https://www.ncbi.nlm.nih.gov/geo/>\) was used for H3K27ac peaks in EC cell-lines from results using ChIP sequencing assays \(GSE106433\) and expression data in ECs and normal esophageal mucosae before neo-adjuvant CRT \(GSE45670\) or neo-adjuvant chemotherapies \(GSE104958\) using microarray and pathologic diagnoses of therapeutic effects on EC tumors after esophagectomies. Shiny Methylation Analysis Resource Tool \(SMART\) \(<http://www.bioinfo-zs.com/smartapp/#tab-8724-4>\) was used to assess the correlation between CpG-site methylation and *MSX1* expression in ECs. The cBio Cancer Genomics Portal \(<https://www.cbioportal.org/>\),^{32,33} from which 17 cBioPortal.org Esophagus/Stomach studies representing 3,791 patients were identified and filtered in June 2021 and prognostic outcomes such as progression-free survival \(PFS\) and disease-free survival \(DFS\) were compared between patients with \(\$n = 50\$ \) and without \(\$n = 3,241\$ \) *MSX1* alterations among the 3,291 patients examined for *MSX1*.](http://www.</p>
</div>
<div data-bbox=)

Summary of clinical endpoints of the first, the second and BBJ cohorts and neoadjuvant CRT/ChT cohorts in online published clinicogenomic database (GSE studies)

The clinical endpoints of each analysis in this study were as follows: CRT-refractory disease for a GWAS in the first cohort, hazard ratios for OS and PFS by a Cox proportional hazard model for selected 10 SNPs in the first cohort, hazard ratios for PFS by a Cox proportional hazard model for four selected SNPs in the second cohort, tumor response estimated by RECIST version 1.1 for the BBJ platinum-based cohort, and pCR induction for neoadjuvant CRT (GSE45670 study) or ChT (GSE104958 study). OS was not included in the second cohort because the mean follow-up period was not sufficient in the second cohort compared to the first cohort (28 months in the second cohort vs 79 months in the first cohort). OS and PFS of the first and the second cohorts were re-evaluated by Kaplan–Meier method. The reason that we adopted tumor response estimated by RECIST as an endpoint of clinical outcome in the BBJ platinum-based cohort is described above.

Summary of biological functions of rs3815544 and MSX1 in ECs examined by online omics databases

We used five online-published omics databases as follows: iMETHYL database to assess the association between rs3815544 and CpG methylation, Wanderer tool to assess CpG methylation status around the *MSX1* gene in ECs and normal esophageal mucosae, GEO portal site to assess H3K27ac peaks in ECs (GSE106433) and to assess the association between pCR induction and *MSX1* expression status before platinum-based neoadjuvant CRT (GSE45670) or ChT (GSE104958), SMART portal site to assess the correlation between CpG-site methylation and *MSX1* expression in EC, and cBio Cancer Genomics Portal to assess *MSX1* genetic alterations and prognosis of the EC/Stomach cancer cohort. The details of each online multiomics database are described above in the 'Patients and Methods' section.

Results

The first and the second cohorts' characteristics

The characteristics of the first and the second cohorts of this study are summarized in Table 1. Mean follow-up months, from the initiation of the primary therapy with CRT to patients' death or the end of follow-up, were 79.3 months and 27.9 months in the first and the second cohorts, respectively.

Pathologic diagnosis of squamous cell carcinoma, male patients, and mean age at initiation of primary treatment accounted for 97% ($n = 91$), 93% ($n = 87$) and 63.4 years, respectively, in the first cohort, and for 96% ($n = 23$), 83% ($n = 21$), and 66.4 years, respectively, in the second cohort, which were generally representative of typical Japanese patients with EC. We included histological types other than squamous cell carcinoma, such as malignant melanoma or neuroendocrine carcinoma, as long as they had been treated with platinum-based CRT as first-line treatment.

The mean radiation doses were 61.3 Gy in the first cohort (data unavailable in four cases) and 59.2 Gy in the second cohort, which were almost equal to the mean radiation doses of definitive CRT previously reported.⁵ A total of 93 % ($n = 85$) of the first cohort (data unavailable in three cases) and 92 % ($n = 22$) of the second

cohort received CRT with concurrent use of platinum derivatives.

The overall survival (OS) and progression-free survival (PFS) in both cohorts were analyzed using the Kaplan–Meier method. In the first cohort, median OS and median PFS were 88.10 months (48.7–127.5 months; 95% confidence interval (CI)) and 88.10 months (40.1–136.1 months; 95% CI), respectively. In the second cohort, median OS and median PFS were 13.40 months (7.7–19.1 months; 95% CI) and 5.00 months (2.6–7.4 months; 95% CI).

Potential candidate SNPs for poor prognosis outcomes

All SNPs identified in the first cohort are summarized as potential candidates for prognosis-outcome prediction after CRT for EC in Figure 2(a) and Table 2. Eighty-nine SNPs were identified as risk factor for CRT-refractory diseases (Table 2).

We selected the leading 10 SNPs filtered by chromosomal location (Table 2), odds ratio (OR), and the number of genotyped patients, and examined the hazard ratios for PFS and OS using a Cox proportional hazard model incorporating age and each SNP, because only age at CRT was statistically associated with prognosis by Cox univariate regression analysis among age, sex, radiation dose, and clinical stage (data not shown). The results of the multivariate analyses are summarized in Table 3 (upper panel). Four SNPs were significant in PFS or OS, with a p -value < 0.0025 after Bonferroni correction. Each of these four was included in the candidate-polymorphism approach in the second cohort in terms of PFS hazard ratio by multivariate analysis using a Cox proportional hazard model (Table 3, lower). Using this candidate-polymorphism approach, only rs3815544, located close to *MSX1*, was found to be statistically significant ($p < 0.0125$, after Bonferroni correction) among the four SNPs in the second cohort.

In the BBJ cohort for platinum-based therapy, we adopted tumor response estimated by RECIST version 1.1³⁴ as an endpoint of clinical outcome because there are missing values in long-term prognosis and follow-up; however, therapeutic efficiency was precisely evaluated by the post-treatment imaging according to RECIST in this BBJ-EC cohort (Figure 1(b)). The alternative

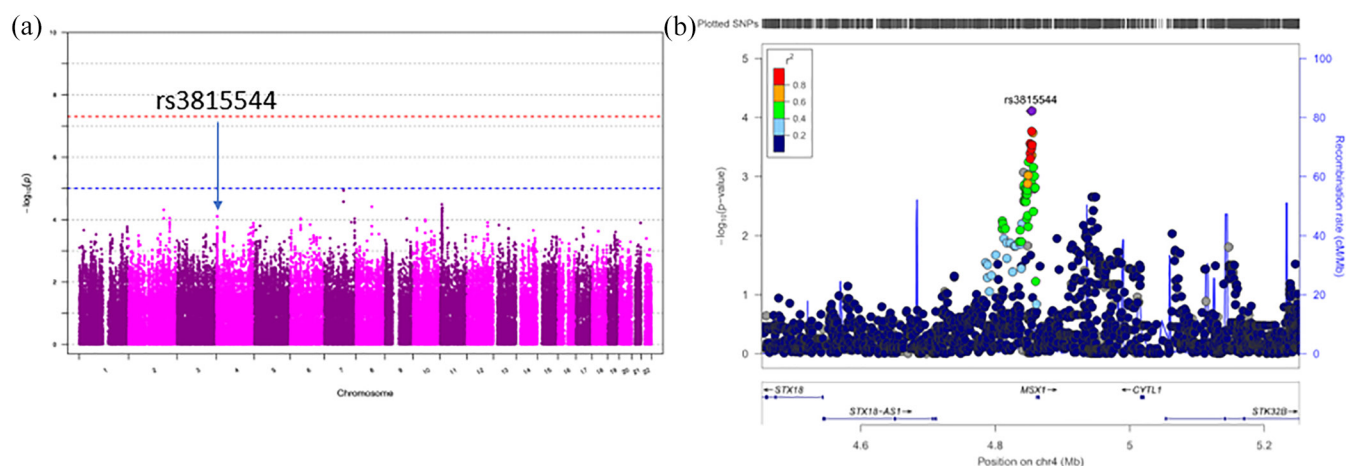


Figure 2. (a) Manhattan plot of association of SNPs with CRT-refractory disease in 94 EC patients; the results from 38 EC patients with CRT-refractory disease and 56 EC patients of control (non-CRT-refractory). (b) An example of the detailed view of GWAS; results around rs3815544 in the region of chromosome 4. rs3815544 was selected as a candidate predictive marker from this chromosomal region.

allele (G) of rs3815544 was significantly associated with the risk of non-response (SD or PD) to platinum-based ChT/CRT (OR = 1.801, $p = 0.0480$) (Table 4, right), despite the fact that there were no statistical differences among the patients with each genotype of rs3815544 in terms of age, sex, clinical stage and details of treatment procedures (Table 4, left). In contrast, we could not find any statistical significance between rs3815544 genotype and the therapeutic effect of radiation therapy, although the number of patients might not be enough and the therapeutic effect by radiation therapy was only estimated by the 1-year, 2-year, and 3-year survivals due to unavailability of RECIST evaluations in clinical data-sheets for radiation therapy in BBJ (Figure 1(b)); the ORs of the 1-year, 2-year, and 3-year survivals were 1.71 ($p = 0.2339$), 1.28 ($p = 0.5617$) and 1.42 ($p = 0.4331$), respectively.

Attribution of rs3815544 to methylation of CpG sites within the 5'-flanking region of the MSX1 gene

We used the Wanderer portal site³¹ containing original methylation data from TCGA to examine CpG methylation levels around *MSX1* gene in ECs and normal esophageal mucosae. As shown in Figure 3 (upper), the CpG methylation probe near rs3815544 was mostly methylated in both the normal mucosa and EC. In addition, there were two regions that were more methylated in ECs than in normal mucosae, including the CpG site that was located at a transcription start site

(TSS) and contained an H3K27 acetylation peak. Meanwhile, the contribution of rs3815544 to epigenetic modification of the promoter or enhancer regions of *MSX1* gene was examined using iME-THYL, which revealed that rs3815544 contributed to the methylation of CpG sites not only around rs3815544 but also other CpG sites, including a CpG (chr4: 4861443), which was close to the TSS of *MSX1* and H3K27 acetylation peak (Figure 3, middle). Specifically, the CpG site at chr4: 4861443 is located within the CpG island that contains other CpG sites, such as chr4: 4861101, chr4: 4861330, and chr4: 4862240. Among them, methylation of chr4: 4861330 and the region between chr4: 4861675 and chr4: 4861880 was associated with *MSX1* suppression by previous reports,^{35,36} and the methylation of 13 CpGs between chr4: 4861101 and chr4: 4862240 was statistically correlated with the suppression of *MSX1* within ECs by the SMART (Figure 3, lower).

MSX1 gene expression and EC prognoses after platinum-based chemotherapy or chemoradiation therapy in online published data

Comprehensive gene expression data from pre-treatment biopsies and pathologic diagnoses for neoadjuvant platinum-based ChT or CRT efficiencies are available in GSE45670 and GSE104958.^{37,38} We downloaded the data from the GEO portal site. The total number of ECs examined by microarray was 68 (28 in GSE45670

Table 2. SNPs selected in the first cohort by GWAS with P -value < 0.0001 .

Chromosome	SNP	BP (GRCh37)	Allele, tested	patients	OR	P
2	rs7426104*	176928633	T	94	4.257	0.00004786
2	rs113240312	176942991	A	91	4.082	0.00009020
2	rs10197152*	206790969	A	93	4.516	0.00008981
4	rs3815544*	4852841	G	94	4.775	0.00007755
6	rs70977374	52545656	GGT	93	0.2161	0.00009029
6	rs3823030*	52547463	A	94	0.2201	0.00009681
6	rs6923338	52549322	C	94	0.2201	0.00009681
7	rs151269855	97215572	G	92	8.201	0.00002633
7	rs74637678*	97240028	G	93	9.157	0.00001160
7	rs10952440	153346868	C	94	0.1655	0.00009169
7	rs11771591	153347062	C	94	0.1655	0.00009169
7	rs13230250	153348384	T	94	0.1655	0.00009169
7	rs13243274*	153348436	G	94	0.1655	0.00009169
7	rs6464356	153349518	A	94	0.1655	0.00009169
7	rs4410821	153350882	T	94	0.1655	0.00009169
7	rs4345479	153351015	T	94	0.1655	0.00009169
8	rs140503424	80742182	TC	94	6.64	0.00003814
8	rs1396858*	80743270	T	94	6.64	0.00003814
9	rs1962035*	110291122	A	94	0.1765	0.00009125
11	rs11042558	9961294	C	93	4.454	0.00009550
11	rs10770077	9968354	T	94	4.589	0.00005826
11	rs140277070	9968568	G	94	4.589	0.00005826
11	rs10840330	9972362	C	94	4.589	0.00005826
11	rs10500715	9973062	G	94	4.589	0.00005826
11	rs11482152	9975293	GT	94	4.589	0.00005826
11	rs11454412	9977418	TA	94	4.589	0.00005826
11	rs7114044	9977527	C	94	4.589	0.00005826
11	rs7106914	9977971	C	94	4.589	0.00005826
11	rs12222953	9978864	T	94	4.589	0.00005826
11	rs7951950	9979332	T	94	4.589	0.00005826
11	rs4909917	9981165	G	94	4.599	0.00004411
11	rs10840333	9981768	G	94	4.589	0.00005826

(Continued)

Table 2. (Continued)

Chromosome	SNP	BP (GRCh37)	Allele, tested	patients	OR	P
11	rs139854573	9982203	AAAT	93	4.466	0.00007194
11	rs78035723	9982516	TA	94	4.589	0.00005826
11	rs2403226	9983737	T	94	4.589	0.00005826
11	rs2403227	9984138	G	94	4.589	0.00005826
11	rs4910086	9985980	C	94	4.599	0.00004411
11	rs4576815	9987228	C	94	4.589	0.00005826
11	rs2403229	9987251	A	94	4.589	0.00005826
11	rs12290996	9987559	T	93	4.766	0.00005033
11	rs77182089	9990387	G	93	4.766	0.00005033
11	rs58084362	9991467	T	94	4.589	0.00005826
11	rs11042574	9994795	C	94	4.589	0.00005826
11	rs12283105	9999434	A	94	4.589	0.00005826
11	rs10770079	10002515	G	93	4.681	0.00004964
11	rs11526057	10007445	G	94	4.589	0.00005826
11	rs117290776	10009176	T	94	4.589	0.00005826
11	rs78411681	10010629	G	94	4.589	0.00005826
11	rs57420335	10020485	C	94	4.589	0.00005826
11	rs16907368	10025713	G	94	4.353	0.00008241
11	rs78858991	10028689	T	93	4.454	0.00009550
11	rs11042589	10036166	A	94	4.589	0.00005826
11	rs74403655	10039085	C	94	4.589	0.00005826
11	rs12221716	10040089	T	93	4.466	0.00007194
11	rs11042590	10040135	A	94	4.589	0.00005826
11	rs78622859	10041931	C	94	4.589	0.00005826
11	rs11042591	10043808	G	94	4.589	0.00005826
11	rs11042592	10046834	C	94	4.589	0.00005826
11	rs10500717	10069486	T	94	4.589	0.00005826
11	rs10840347	10074278	G	94	4.589	0.00005826
11	rs10840348	10075500	C	94	4.589	0.00005826
11	rs10840352	10076895	A	94	4.589	0.00005826
11	rs79044582	10086668	G	93	4.454	0.00009550
11	rs11042604	10089027	A	93	4.454	0.00009550

(Continued)

Table 2. (Continued)

Chromosome	SNP	BP (GRCh37)	Allele, tested	patients	OR	P
11	rs11042605	10089886	C	94	4.589	0.00005826
11	rs10840359	10091024	C	94	4.589	0.00005826
11	rs11042606	10091525	C	93	4.766	0.00005033
11	rs7946995	10093760	C	94	4.589	0.00005826
11	rs7114303	10095687	C	94	4.589	0.00005826
11	rs10458929	10099110	A	94	4.589	0.00005826
11	rs11042610	10099720	T	94	4.589	0.00005826
11	rs11042613	10103782	C	94	4.589	0.00005826
11	rs10840361	10106833	G	94	4.589	0.00005826
11	rs10840362	10113835	T	94	4.589	0.00005826
11	rs11042624	10117388	A	94	4.589	0.00005826
11	rs4529876*	10118993	G	94	5.033	0.00003233
11	rs11530489	10121828	G	94	4.589	0.00005826
11	rs7938788	10132712	C	94	4.589	0.00005826
11	rs1442732	10141219	C	94	4.589	0.00005826
11	rs11042635	10142675	C	94	4.589	0.00005826
11	rs10840370	10148856	T	94	4.589	0.00005826
11	rs11042639	10149372	T	94	4.589	0.00005826
11	rs10770092	10154095	C	93	4.505	0.00007105
11	rs11042645	10171400	T	94	4.589	0.00005826
11	rs11042647	10175364	A	94	4.589	0.00005826
11	rs11042669	10235812	G	94	4.599	0.00004411
11	rs79668791	10272208	A	94	4.349	0.00006808
11	rs7944869	12933329	A	93	0.1231	0.00005358
11	rs11022546*	12934243	A	94	0.1193	0.00004150

BP, base pair in physical position; GWAS, genome-wide association study; OR, odds ratio; OS, overall survival; PFS, progression-free survival; SNP, single-nucleotide polymorphism.
Top 10 SNPs indicated by asterisks proceeded to the evaluation of OS and PFS.

and 40 in GSE104958), accompanied by 15 normal mucosae (10 in GSE45670 and five in GSE104958). We determined the standard value of *MSX1* expression using the average and standard deviation (STDV) calculated from all values in normal mucosae in each GSE study. We summarized the suppression of *MSX1*, which we

determined below one STDV from the average in each GSE study, compared to pathologic diagnoses after platinum-based ChT or CRT (Table 5). The results showed that none of the cases with suppression of *MSX1* showed pathological complete remission (pCR), and the relative risk of negative *MSX1* for non-pCR was estimated as

Table 3. PFS and OS hazard ratios by Cox proportional hazard analyses, for each of 10 selected SNPs in the first cohort (upper) and four selected SNPs in the second cohort (lower).

Cohort	1	PFS	OS								
10 SNPs selected in each chromosomal localization		PFS hazard by each SNP			Hazards of co-variables		OS hazard by each SNP			Hazards of co-variables	
		Hazard	p-value	95% C.I.	Age at CRT	Hazard	p-value	95% C.I.	Age at CRT		
rs7426104	GT/GG vs TT	2.772	0.003868	1.388-5.537	1.028 [0.996-1.062]	2.836	0.002892	1.428-5.632	1.029 [0.997-1.062]		
rs10197152*	AG/GG vs AA	4.495	0.000195	2.038-9.913	1.041 [1.007-1.076]	4.944	0.000034	2.321-10.530	1.045 [1.011-1.080]		
rs3815544*	GA/AA vs GG	5.607	0.000055	2.426-12.958	1.045 [1.010-1.081]	5.095	0.000076	2.275-11.411	1.048 [1.013-1.083]		
rs3823030*	GA/AA vs GG	4.220	0.001050	1.784-9.984	1.027 [0.995-1.061]	3.735	0.002514	1.589-8.782	1.032 [1.000-1.065]		
rs74637678	GA/AA vs GG	12.629	0.022517	1.430-111.560	1.042 [1.008-1.078]	2.895	0.153294	0.673-12.458	1.041 [1.008-1.076]		
rs13243274	GA/GG vs AA	1.717	0.057714	0.982-3.001	1.030 [0.998-1.063]	1.784	0.037875	1.033-3.082	1.032 [1.000-1.064]		
rs1396858	CT/CC vs TT	5.311	0.007644	1.557-18.117	1.041 [1.009-1.074]	5.610	0.006087	1.636-19.238	1.044 [1.012-1.078]		
rs1962035	AC/AA vs CC	2.629	0.111450	0.800-8.647	1.038 [1.006-1.071]	1.638	0.424677	0.488-5.502	1.039[1.008-1.072]		
rs4529876*	GT/TT vs GG	7.031	0.000002	3.170-15.593	1.038 [1.003-1.074]	5.631	0.000031	2.496-12.700	1.038 [1.005-1.073]		
rs11022546	GA/AA vs GG	4.876	0.124166	0.647-36.746	1.025 [0.993-1.057]	4.620	0.136884	0.615-34.713	1.028 [0.996-1.060]		
Cohort	2	PFS									
4 SNPs, selected by the first cohort		PFS hazard by each SNP			Age at CRT						
		Hazard	p-value	95% C.I.	Age at CRT						
rs10197152	AG/GG vs AA	0.871	0.829357	0.248-3.062	0.937 [0.864-1.016]						
rs3815544	GA/AA vs GG	6.420	0.011190	1.526-27.009	0.931 [0.850-1.019]						
rs3823030	GA/AA vs GG	0.186	0.012531	0.050-0.697	1.022 [0.937-1.114]						
rs4529876	GT/TT vs GG	0.787	0.820501	0.099-6.228	0.974 [0.903-1.051]						
CRT, chemoradiation therapy; OS, overall survival; PFS, progression-free survival; SNP, single-nucleotide polymorphism. Four SNPs with significance, which proceeded to analysis by the second cohort, are marked with asterisks in the first cohort.											

CRT, chemoradiation therapy; OS, overall survival; PFS, progression-free survival; SNP, single-nucleotide polymorphism. Four SNPs with significance, which proceeded to analysis by the second cohort, are marked with asterisks in the first cohort.

Table 4. The association between genotypes of rs3815544 and RECIST-based evaluations by post-treatment imaging in the BBJ-EC cohort obtained from BioBank Japan.

Patients' characteristics'					rs3815544 alleles vs RECIST evaluations							
RECIST	rs3815544	Number of patients	sex	Age	Stage	Concurrent RT	Platinum derivative			Total number of allele		RECIST
				(average ± st. dev.)	(Japan Esophageal Society)	(no RT: conc. RT)	CDGP	CBDCA	CDDP	A	G	
CR/PR	AA	31	M = 28 F = 3	64.0 ± 7.5	stage0 = 1 stage1 = 1 stage2 = 3 stage3 = 11 stage4a = 5 stage4b = 2 (NA = 8)	21: 7 (NA = 3)	2	0	29	62	0	CR/PR (response)
GA		26	M = 25 F = 1	63.6 ± 8.5	stage0 = 0 stage1 = 2 stage2 = 2 stage3 = 6 stage4a = 9 stage4b = 1 (NA = 6)	14: 12	2	0	24	26	26	
GG		6	M = 6 F = 0	62.3 ± 4.0	stage0 = 0 stage1 = 1 stage2 = 2 stage3 = 0 stage4a = 1 (NA = 2)	5: 1	2	1	3	0	12	
total alleles												
SD/PD	AA	13	M = 10 F = 3	62.0 ± 9.1	stage0 = 0 stage1 = 1 stage2 = 3 stage3 = 2 stage4a = 2 stage4b = 0 (NA = 5)	9: 3 (NA = 1)	0	0	13	26	0	SD/PD (non-response)
GA		28	M = 24 F = 4	63.4 ± 9.1	stage0 = 0 stage1 = 2 stage2 = 3 stage3 = 8 stage4a = 5 stage4b = 2 (NA = 8)	17: 10 (NA = 1)	3	1	24	28	28	

(Continued)

Table 4. (Continued)

Patients' characteristics'							rs3815544 alleles vs RECIST evaluations					
RECIST	rs3815544	Number of patients	sex	Age	Stage	Concurrent RT	Platinum derivative			Total number of allele		RECIST
				(average ± st. dev.)	(Japan Esophageal Society)	(no RT: conc. RT)	CDGP	CBDCA	CDDP	A	G	
	GG	7	M = 6 F = 1	67 ± 2.1	stage0 = 0 stage1 = 3 stage2 = 0 stage3 = 3 stage4a = 0 stage4b = 0 (NA = 1)	4: 3	1	0	6	0	14	
										total alleles	A = 54 G = 42	
total	AA	44	M = 38 F = 6	63.4 ± 8.0	stage0-3 = 22 stage4a/4b = 9 (NA = 13)	30: 10 (NA = 4)	2	0	42	P = 0.0480 (Pearson's chi-square / Fisher exact test)		
	GA	54	M = 49 F = 5	63.5 ± 8.8	stage0-3 = 23 stage4a/4b = 17 (NA = 14)	31: 22 (NA = 1)	5	1	48	Relative risk = 1.801 [95% CI: 1.035-3.136]		
	GG	13	M = 12 F = 1	64.8 ± 7.5	stage0-3 = 9 stage4a/4b = 1 (NA = 3)	9: 4	3	1	9			
Pearson's chi-square test							p = 0.691	(ANOVA) p = 0.858	p = 0.242	p = 0.067		
ANOVA, analysis of variance; BBJ, BioBank Japan; CBDCA, carboplatin; CDDP, cisplatin; CDGP, nedaplatin; CI, confidence interval; CR, complete response; EC, esophageal carcinoma; NA: not available; RECIST: Response Evaluation Criteria In Solid Tumor; RT, radiation therapy; PD, progressive disease; PR, partial response; SD, stable disease.												

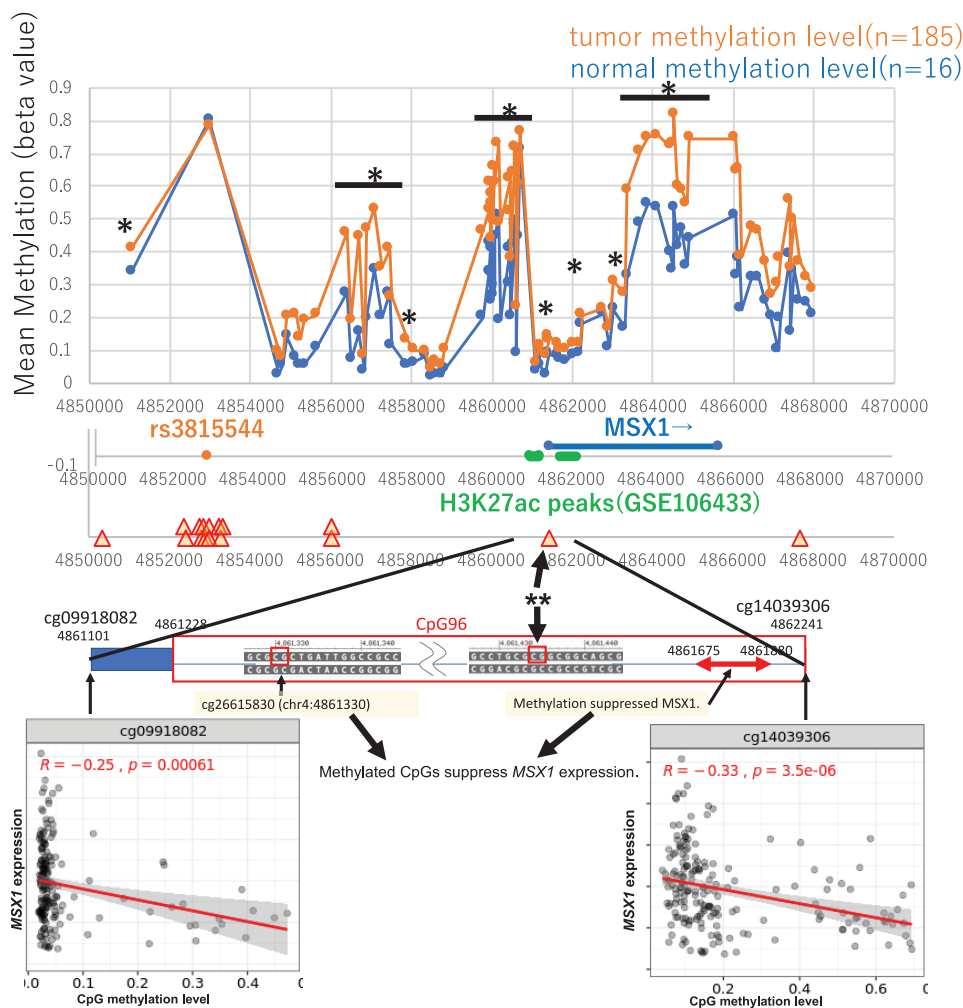


Figure 3. The mean methylation of *MSX1* and its promoter region in normal esophageal mucosa (blue) and tumor tissues of esophageal carcinoma (EC, red) were adopted from 'Wanderer' (<http://www.maplab.cat/wanderer>) in the upper row. Mean methylation as the beta-value is indicated along the Y-axis, and chromosomal locations are indicated on the X-axis (GRCh37). Each asterisk indicates a CpG site with a statistically significant methylation status difference between normal mucosae and tumor tissues ($p < 0.05$). An asterisk with an under-bar indicates a region significantly more methylated by all probes within the region. Attributions of rs3815544 to CpG sites are based on the iMETHYL database, and H3K27ac peaks in EC cell lines are derived from ChIP assay results (GSE106433), as shown in the middle of this figure. rs3815544 is indicated by an orange dot. Triangles indicate rs3815544-attributed CpG methylations. H3K27ac peaks around the *MSX1* gene in the EC cell-lines KYSE510, KTSE70, and TT, suggested by ChIP sequencing, are indicated by green bold lines. The online database, SMART, suggested that CpG site indicated by an arrow with double asterisks locates within CpG island (indicated as CpG96) containing other CpG sites (named as cg09918082 and cg14039306) that suppress *MSX1* within ECs through methylation, as shown at the bottom. The other two CpG sites were also reportedly associated with *MSX1* suppression.

1.488 with 95% confidence interval 1.1254—1.766 ($p < 0.05$).

Inactivating alterations such as homozygous deletions of MSX1 as prognostic-marker candidates of EC suggested by online published data

DFS and PFS were compared between the two groups, with or without *MSX1* alterations, and

Kaplan–Meier curves were generated using the cBio Cancer Genomics Portal. EC patients, including gastric cancer patients in some studies, with altered *MSX1* had significantly worse prognoses both in DFS and PFS, with p -values of 1.082×10^{-5} and 0.0225, respectively (Figure 4(a) and (b)). Of EC patients with altered *MSX1*, 50% ($n = 25$) were homozygous deletions (Figure 4(c)). Although there were many missing values in the details of treatment procedures in this

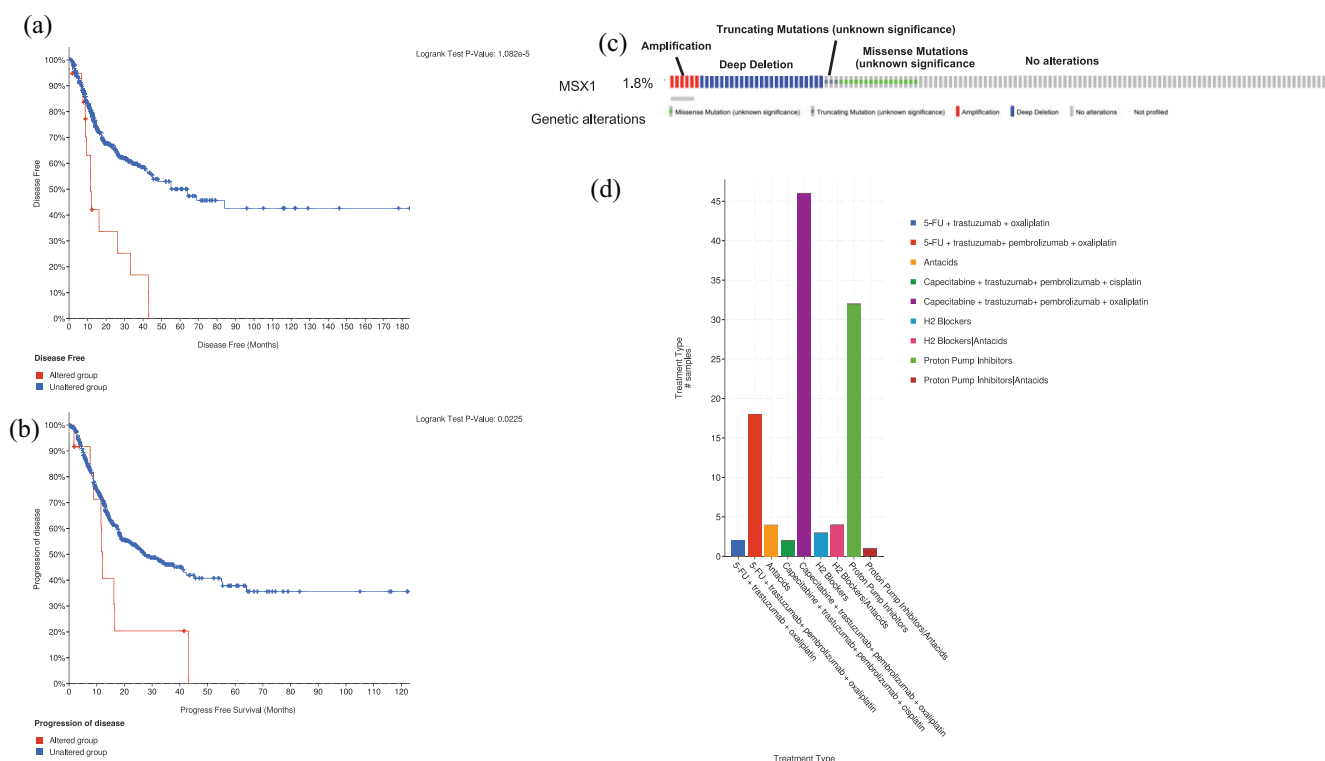


Figure 4. In the current study, 17 cBioPortal.org Esophagus/Stomach studies representing 3,791 patients were identified and filtered in June 2021, from the cBio Cancer Genomics portal (<https://www.cbioportal.org>). Prognostic outcomes such as PFS and DFS between the two subgroups, that is, EC/GC patients with or without *MSX1* alterations, were compared by Kaplan–Meier method. EC/GC patients with altered *MSX1* significantly had worse prognoses both in DFS (a) and PFS (b) with p -values of 1.082×10^{-5} and 0.0225, respectively. (c) Of those with altered *MSX1*, 50% ($n = 25$) were homozygous deletions. (d) Although there were many missing values in details of treatment procedures in this clinicogenomic data, platinum derivatives such as oxaliplatin or cisplatin were used in all cases available for treatment information, excluding gastric acid suppressants. Each figure above was created and adopted from cBio Cancer Genomics (<https://www.cbioportal.org>).

clinicogenomic data, platinum derivatives such as oxaliplatin and cisplatin were used in all cases for whom treatment data were available, excluding gastric acid suppressants (Figure 4(d)).

Discussion

This study involved GWAS screening for CRT-refractory diseases, followed by OS and PFS hazard ratio evaluation using a Cox proportional hazard model in the first cohort. This implicated the four candidate SNPs, as well as the sample size and follow-up periods of the second cohort for validation. Among these four, only rs3815544 was statistically confirmed in the validation study by the second cohort. This result was also re-evaluated using other external nationwide clinicogenomic data from the BioBank Japan (BBJ). Furthermore, the molecular and pathogenic functioning of this SNP was assessed using online published multiomics databases. The Wonderer

portal site was used to study the methylation status around *MSX1* gene within ECs and corresponding normal mucosae; the iMETHYL portal was used to investigate the relation between candidate SNP and CpG methylations around *MSX1* gene. Finally, the SMART portal was used to assess the relation between CpG methylations and expression status of *MSX1* within ECs; in addition, data from two GSE studies were analyzed to identify the relations between *MSX1* suppression and pCR induction in ECs.

To our knowledge, this study is the first to suggest a predictive SNP candidate utilizing the population-optimized whole-genome SNP array for patients with EC that had been treated primarily using platinum-based therapy. For patients with EC primarily treated using platinum-based therapy as first-line treatment, the number of enrolled patients expected from previous studies and the incidence of this disease suggested that the

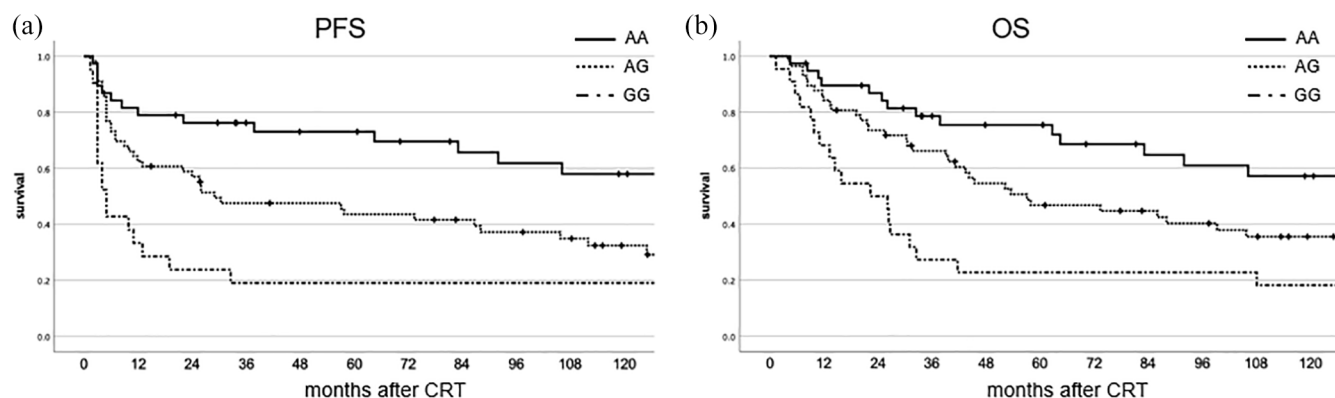


Figure 5. Kaplan–Meier survival curves of 118 EC patients from the two cohorts in this study by the genotypes of rs3815544. Survival ratios were significantly dependent on the genotypes of rs3815544 by Kaplan–Meier analyses, both in PFS with $p = 0.000058$ (logrank) (a) and OS with $p = 0.000151$ (log rank) (b); median survival time in each genotype was 5.0 months (GG), 26.0 months (GA), 162.6 months (AA) in PFS, and 22.4 months (GG), 53.6 months (GA), 162.6 months (AA) in OS.

current study design must be similar to that of rare and lethal cancer studies,³⁹ rather than previous GWASs that suggested predictive markers of more common malignant diseases such as breast cancer, lung cancer, and cervical cancer.^{40–42} In fact, the total number of enrolled patients, 118 in this study, and the length of observation after CRT – an average of 80 months of the first cohort – have been rarely observed in previous studies. For example, two representative large-scale long-term cohort studies on CRT for ECs consisted of 55 EC patients with a median survival time of 16 months and 56 EC patients with 14 months of median survival time.^{5,43}

We re-evaluated GWAS results using a Cox hazard proportional model for OS and PFS, incorporating age at CRT and each of 10 SNPs as potential candidates from GWAS filtered by chromosomal location, ORs and the number of genotyped patients (Table 2). From the above, we selected four SNPs (Table 3), which were used in the candidate-polymorphism study in the second cohort consisting of 24 patients that were suggested by CRAB statistics, including patients who were prospectively recruited to fulfill the study design. Eventually, only one SNP, rs3815544, was found to be significant hazard ratios for PFS (Table 3, lower).

Kaplan–Meier analyses showed that PFS and OS were significantly associated with each of alleles of rs3815544, suggesting the alternative allele (G) as a risk allele following first-line platinum-based CRT (Figure 5). For the first and the second cohorts, we selected EC patients treated with

platinum-based CRT to make the treatment background uniform, but we predicted that rs3815544 might be predictive not only for CRT but also for platinum-based chemotherapy for the reasons suggested by preceding studies, as described later. Therefore, we decided that this result should be validated in another external cohort of EC patients treated with platinum-based therapy obtained from BBJ.

BBJ is one of the largest disease biobanks in the world and included 1,338 patients with EC.^{19,20} We prepared the external cohort to investigate the association between platinum derivatives and objective evaluations of therapeutic effects such as RECIST evaluation (Figure 1(b)). The reasons that we adopted RECIST evaluation of post-treatment imaging as an endpoint of the prognosis outcome were as follows: first, patients with EC of the BBJ cohort included those who were not available for medium- or long-term follow-up data for survival analyses; second, precise, objective, and widely shared methods to evaluate therapeutic effects were needed because this nationwide cohort consisted of clinicogenomic data from a number of hospitals all over Japan. The results indicated that the alternative allele (G) of rs3815544 was statistically associated with tumor non-response (SD or PD) (Table 4, right). It should be noted that the result was replicated even in an external cohort that consisted of patients treated with various combinations of platinum derivatives with or without concurrent radiation therapy. There were no statistical differences in patient background, such as age, sex, and clinical factors such as clinical stages and the use

Table 5. *MSX1* expression levels before CRT or neoadjuvant chemotherapy and pathologic assessment of EC after neoadjuvant chemotherapy or CRT, by the online published data.

		GSE45670		GSE104958			GSE45670 + GSE104958	
<i>MSX1</i> expressions adjusted by normal mucosae		Non-pCR	pCR	Non-pCR	pCR	Negative <i>MSX1</i>	non-pCR	pCR
<i>MSX1</i> expression	< average - 1SD	2	0	3	0	Negative (< average - 1SD)	5	0
	average ± 1SD	12	11	6	2	Non-negative (≥ average -1SD)	43	21
	> average + 1SD	3	0	22	8			
						Relative risk (95% Confidence Interval)	1.488 (1.254-1.766)	
CRT, chemoradiation therapy; EC, esophageal carcinoma; pCR, pathological complete remission; SD, standard deviation.								

of concurrent radiation therapy, between genotypes of rs3815544 in this BBJ-EC cohort (Table 4, left). In fact, patients with relatively early stage ECs might account for a substantial portion of those with homozygous alternative alleles (GG) ($p = 0.121$), which might rather support the view that the alternative allele (G) of rs3815544 is associated with poor prognosis. We also investigated the association of the rs3815544 genotype and the therapeutic effect of radiation therapy but did not find any statistical significance.

MSX1 reportedly plays an important role as a functional tumor suppressor in the genesis of various types of human cancers.^{44–48} A germline variant in *MSX1* was identified in a Dutch family with clustering of Barrett's esophagus and esophageal adenocarcinoma, suggesting that the loss of *MSX1* plays an important role in esophageal carcinogenesis.⁴⁹ In fact, *MSX1* was listed among candidate driver genes for esophageal carcinogenesis, as suggested by copy number variants from the TCGA database.⁵⁰ Furthermore, a promoter hypermethylation signature of *MSX1* has been reported in various types of human malignancies by genome-wide screening for promoter methylation,⁵¹ which has also been confirmed in EC (Figure 3). Consistently, *MSX1* alterations in EC were mostly homozygous deletions, as suggested by the cBio Cancer Genomics Portal database (Figure 4(c)). Taking these factors into consideration, *MSX1* was suspected to play an important role in esophageal carcinogenesis, suggesting the presence of an EC subpopulation driven by

deleterious *MSX1* either from homozygous deletion or promoter methylation.

The suppression of *MSX1* has previously been reported to be associated with platinum-resistant diseases in high-grade epithelial ovarian cancers, in which the expression of *MSX1* was dependent on CpG-methylation status.⁵² High *MSX1* expression was listed among the genes highly expressed in the intestinal type of gastric carcinoma, which is more sensitive to platinum-based chemotherapy than the diffuse type.⁵³ These previous studies strongly suggest that the expression/suppression of *MSX1* contributes to sensitivity/resistance to platinum derivatives. We examined the association between *MSX1* expression and pathological CR (pCR) induction in EC patients using online published data; comprehensive gene expression data by microarray before neo-adjuvant CRT/ChT and pathologic diagnoses after treatments have been published online as GSE45670 study (platinum-based CRT) and GSE104958 study (platinum-based ChT: a combination of cisplatin, docetaxel, and 5-fluorouracil).^{37,38} This result was compatible with the current study; all cases with reduced *MSX1* showed non-pCR. The estimated relative risk of reduced *MSX1* for non-pCR was 1.488 ($p < 0.05$, 95% confidence interval 1.1254–1.766) (Table 5). In addition, the cBio Cancer Genomics Portal database suggested that EC patients with *MSX1* alterations that mostly consisted of homozygous deletions had poor prognoses in both DFS ($p = 1.082 \times 10^{-5}$) and PFS ($p = 0.0225$) compared to those without

MSX1 alterations (Figure 4(a) and (b)). Platinum derivatives, such as oxaliplatin and cisplatin, were used in the all patients for whom treatment data were available. Taken together, it is plausible that the suppression of *MSX1* might confer resistance to platinum-based chemotherapy in patients with EC.

To date, there is no direct evidence of an association between rs3815544 and *MSX1* expression in EC. However, rs3815544 could be attributed to CpG methylation around the 5'-flanking region of *MSX1* (Figure 3, upper panel). Using online multiomics databases such as Wonderer, iME-THYL and SMART, we found that rs3815544 contributed to the methylation of some CpG sites, including one at chr4: 4861443. This CpG site at chr4: 4861443 was close to the transcription start site (TSS) of *MSX1* and the H3K27 acetylation peak and located within the CpG island that contains other CpG sites, such as chr4: 4861101, chr4: 4861330, and chr4: 486224 (Figure 3, middle). These three CpG sites were all reportedly associated with *MSX1* suppression in ECs^{35,36} (Figure 3, bottom). Therefore, the G allele of rs3815544 induces the methylation of CpGs around the H3K27 acetylation sites and downregulates *MSX1*, which was supported by a phenomenon of gene expression control by DNA methylation and histone modification.⁵⁴

In conclusion, the current study suggests a novel predictive marker candidate of platinum derivatives for EC patients, but further studies are needed to validate the findings in another large-scale clinical trial as well as by molecular research to unveil the underlying mechanisms of *MSX1* suppression through CpG methylation around the 5'-promoter/enhancer region.

Acknowledgements

The authors are grateful to Dr. Akira Ono (National Center for Global Health and Medicine) for his technical support and helpful advice. The authors also express sincere thanks to all physicians and laboratory members who have contributed to the foundation and development of the BBJ, which has been added to the coauthors' list as a banner.

Author contributions

Takahiro Mori: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Project administration; Resources; Software; Supervision; Validation; Visualization;

Writing – original draft; Writing – review & editing

Kazuko Ueno: Data curation; Formal analysis; Investigation; Methodology; Software; Validation; Visualization; Writing – original draft; Writing – review & editing

Katsushi Tokunaga: Investigation; Resources; Validation; Writing – review & editing

Yosuke Kawai: Data curation; Formal analysis; Investigation; Methodology; Software; Writing – original draft; Writing – review & editing

Koichi Matsuda: Investigation; Resources; Validation; Writing – review & editing

Nao Nishida: Formal analysis; Resources; Writing – original draft; Writing – review & editing

Keigo Komine: Data curation; Investigation; Writing – review & editing

Sakae Saito: Investigation; Resources; Writing – review & editing

Masao Nagasaki: Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Software; Supervision; Validation; Writing – original draft; Writing – review & editing

The BioBank Japan Project: Resources

Conflict of interest statement

The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: TM received personal fees from OncoTherapy Science outside the submitted work in 2017 and 2018. MN holds a concurrent post at the Department of Cohort Genome Information Analysis endowed by Toshiba Corporation. MN received research funding from Toshiba Corporation.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by a Grant-in-Aid from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan [grant numbers 18591451 and 16K10488], and was partially supported by the Japan Agency for Medical Research and Development (AMED) [grant number JP18km0405205], the Center of Innovation Program from the Japan Science and Technology

Agency, JST, and the research fund from Repertoire Genesis.

ORCID iD

Takahiro Mori  <https://orcid.org/0000-0001-7324-8574>

References

1. Global Burden of Disease Cancer Collaboration, Fitzmaurice C, Allen C, *et al.* Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 32 cancer groups, 1990 to 2015: a systematic analysis for the Global Burden of Disease Study. *JAMA Oncol* 2017; 3(4): 524–548.
2. Tachimori Y, Ozawa S, Numasaki H, *et al.* Comprehensive registry of esophageal cancer in Japan, 2011. *Esophagus* 2018; 15: 127–152.
3. Urba S. Esophageal cancer: preoperative or definitive chemoradiation. *Ann Oncol* 2004; 15(Suppl. 4): iv93–iv96.
4. Swisher SG, Moughan J, Komaki RU, *et al.* Final results of NRG oncology RTOG 0246: an organ-preserving selective resection strategy in esophageal cancer patients treated with definitive chemoradiation. *J Thorac Oncol* 2017; 12: 368–374.
5. Gkika E, Gauler T, Eberhardt W, *et al.* Long-term results of definitive radiochemotherapy in locally advanced cancers of the cervical esophagus. *Dis Esophagus* 2014; 27: 678–684.
6. Rawat S, Kumar G, Kakria A, *et al.* Chemoradiotherapy in the management of locally advanced squamous cell carcinoma esophagus: is surgical resection required? *J Gastrointest Cancer* 2013; 44: 277–284.
7. Teoh AY, Chiu PW, Yeung WK, *et al.* Long-term survival outcomes after definitive chemoradiation versus surgery in patients with resectable squamous carcinoma of the esophagus: results from a randomized controlled trial. *Ann Oncol* 2013; 24: 165–171.
8. Crosby TD, Brewster AE, Borley A, *et al.* Definitive chemoradiation in patients with inoperable oesophageal carcinoma. *Br J Cancer* 2004; 90: 70–75.
9. Ku GY. Systemic therapy for esophageal cancer: chemotherapy. *Chin Clin Oncol* 2017; 6: 49.
10. Wong C and Law S. Predictive factors in the evaluation of treatment response to neoadjuvant chemoradiotherapy in patients with advanced esophageal squamous cell cancer. *J Thorac Dis* 2017; 9(Suppl. 8): S773–S780.
11. Bronson NW, Diggs BS, Bakis G, *et al.* Molecular marker expression is highly heterogeneous in esophageal adenocarcinoma and does not predict a response to neoadjuvant therapy. *J Gastrointest Surg* 2015; 19: 2105–2110.
12. McGranahan N and Swanton C. Clonal heterogeneity and tumor evolution: past, present, and the future. *Cell* 2017; 168: 613–628.
13. Wu X, Gu J, Wu TT, *et al.* Genetic variations in radiation and chemotherapy drug action pathways predict clinical outcomes in esophageal cancer. *J Clin Oncol* 2006; 24: 3789–3798.
14. Kuwahara A, Yamamori M, Fujita M, *et al.* TNFRSF1B A1466G genotype is predictive of clinical efficacy after treatment with a definitive 5-fluorouracil/cisplatin-based chemoradiotherapy in Japanese patients with esophageal squamous cell carcinoma. *J Exp Clin Cancer Res* 2010; 29: 100.
15. Du Z, Zhang W, Zhou Y, *et al.* Associations of ATM polymorphisms with survival in advanced esophageal squamous cell carcinoma patients receiving radiation therapy. *Int J Radiat Oncol Biol Phys* 2015; 93: 181–189.
16. Warnecke-Eberz U, Vallbohmer D, Alakus H, *et al.* ERCC1 and XRCC1 gene polymorphisms predict response to neoadjuvant radiochemotherapy in esophageal cancer. *J Gastrointest Surg* 2009; 13: 1411–1421.
17. Duran G, Aguin S, Cruz R, *et al.* Association of GSTP1 and ERCC1 polymorphisms with toxicity in locally advanced head and neck cancer platinum-based chemoradiotherapy treatment. *Head Neck* 2019; 41: 2704–2715.
18. Kawai Y, Mimori T, Kojima K, *et al.* Japonica array: improved genotype imputation by designing a population-specific SNP array with 1070 Japanese individuals. *J Hum Genet* 2015; 60: 581–587.
19. Hirata M, Kamatani Y, Nagai A, *et al.* Cross-sectional analysis of BioBank Japan clinical data: a large cohort of 200,000 patients with 47 common diseases. *J Epidemiol* 2017; 27: S9–S21.
20. Nagai A, Hirata M, Kamatani Y, *et al.* Overview of the BioBank Japan Project: study design and profile. *J Epidemiol* 2017; 27: S2–S8.
21. McShane LM, Altman DG, Sauerbrei W, *et al.* REporting recommendations for tumour MARKer prognostic studies (REMARK). *Eur J Cancer* 2005; 41: 1690–1696.

22. Altman DG, McShane LM, Sauerbrei W, *et al.* Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): explanation and elaboration. *PLoS Med* 2012; 9: e1001216.
23. Rice TW, Rusch VW, Ishwaran H, *et al.* Cancer of the esophagus and esophagogastric junction: data-driven staging for the seventh edition of the American Joint Committee on Cancer/International Union Against Cancer Staging Manuals. *Cancer* 2010; 116: 3763–3773.
24. Staples J, Qiao D, Cho MH, *et al.* PRIMUS: rapid reconstruction of pedigrees from genome-wide estimates of identity by descent. *Am J Hum Genet* 2014; 95: 553–564.
25. Delaneau O, Howie B, Cox AJ, *et al.* Haplotype estimation using sequencing reads. *Am J Hum Genet* 2013; 93: 687–696.
26. Howie BN, Donnelly P and Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009; 5: e1000529.
27. Nagasaki M, Yasuda J, Katsuoka F, *et al.* Rare variant discovery by deep whole-genome sequencing of 1,070 Japanese individuals. *Nat Commun* 2015; 6: 8018.
28. Yamaguchi-Kabata Y, Nariiai N, Kawai Y, *et al.* iJGVD: an integrative Japanese genome variation database based on whole-genome sequencing. *Hum Genome Var* 2015; 2: 15050.
29. Chang CC, Chow CC, Tellier LC, *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015; 4: 7.
30. Komaki S, Shiwa Y, Furukawa R, *et al.* iMETHYL: an integrative database of human DNA methylation, gene expression, and genomic variation. *Hum Genome Var* 2018; 5: 18008.
31. Díez-Villanueva A, Mallona I and Peinado MA. Wanderer, an interactive viewer to explore DNA methylation and gene expression data in human cancer. *Epigenetics Chromatin* 2015; 8: 22.
32. Cerami E, Gao J, Dogrusoz U, *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2012; 2: 401–404.
33. Gao J, Aksoy BA, Dogrusoz U, *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 2013; 6: pl1.
34. Eisenhauer EA, Therasse P, Bogaerts J, *et al.* New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009; 45: 228–247.
35. Turan N, Ghalwash MF, Katari S, *et al.* DNA methylation differences at growth related genes correlate with birth weight: a molecular signature linked to developmental origins of adult disease? *BMC Med Genomics* 2012; 5: 10.
36. Dunwell TL, Hesson LB, Pavlova T, *et al.* Epigenetic analysis of childhood acute lymphoblastic leukemia. *Epigenetics* 2009; 4: 185–193.
37. Wen J, Yang H, Liu MZ, *et al.* Gene expression analysis of pretreatment biopsies predicts the pathological response of esophageal squamous cell carcinomas to neo-chemoradiotherapy. *Ann Oncol* 2014; 25: 1769–1774.
38. Fujishima H, Fumoto S, Shibata T, *et al.* A 17-molecule set as a predictor of complete response to neoadjuvant chemotherapy with docetaxel, cisplatin, and 5-fluorouracil in esophageal cancer. *PLoS One* 2017; 12: e0188098.
39. Andersson U, McKean-Cowdin R, Hjalmarsson U, *et al.* Genetic variants in association studies – review of strengths and weaknesses in study design and current knowledge of impact on cancer risk. *Acta Oncol* 2009; 48: 948–954.
40. Cao S, Wang S, Ma H, *et al.* Genome-wide association study of myelosuppression in non-small-cell lung cancer patients with platinum-based chemotherapy. *Pharmacogenomics J* 2016; 16: 41–46.
41. Li X, Huang K, Zhang Q, *et al.* Genome-wide association study identifies four SNPs associated with response to platinum-based neoadjuvant chemotherapy for cervical cancer. *Sci Rep* 2017; 7: 41103.
42. Escala-Garcia M, Abraham J, Andrulis IL, *et al.* A network analysis to identify mediators of germline-driven differences in breast cancer prognosis. *Nat Commun* 2020; 11: 312.
43. Gillham CM, Aherne N, Rowley S, *et al.* Quality of life and survival in patients treated with radical chemoradiation alone for oesophageal cancer. *Clin Oncol* 2008; 20: 227–233.
44. Park K, Kim K, Rho SB, *et al.* Homeobox Msx1 interacts with p53 tumor suppressor and inhibits tumor growth by inducing apoptosis. *Cancer Res* 2005; 65: 749–757.
45. Horazna M, Janeckova L, Svec J, *et al.* Msx1 loss suppresses formation of the ectopic crypts developed in the Apc-deficient small intestinal epithelium. *Sci Rep* 2019; 9: 1629.
46. Sliwinski T, Synowiec E, Czarny P, *et al.* The c.469+46_56del mutation in the homeobox MSX1 gene – a novel risk factor in breast cancer? *Cancer Epidemiol* 2010; 34: 652–655.

47. Yue Y, Yuan Y, Li L, *et al.* Homeobox protein MSX1 inhibits the growth and metastasis of breast cancer cells and is frequently silenced by promoter methylation. *Int J Mol Med* 2018; 41: 2986–2996.
48. Yue Y, Zhou K, Li J, *et al.* MSX1 induces G0/G1 arrest and apoptosis by suppressing Notch signaling and is frequently methylated in cervical cancer. *Onco Targets Ther* 2018; 11: 4769–4780.
49. Van Nistelrooij AMJ, Van Marion R, Van Ijcken WFJ, *et al.* Germline variant in MSX1 identified in a Dutch family with clustering of Barrett's esophagus and esophageal adenocarcinoma. *Fam Cancer* 2018; 17: 435–440.
50. Dong G, Mao Q, Yu D, *et al.* Integrative analysis of copy number and transcriptional expression profiles in esophageal cancer to identify a novel driver gene for therapy. *Sci Rep* 2017; 7: 42060.
51. Shames DS, Girard L, Gao B, *et al.* A genome-wide screen for promoter methylation in lung cancer identifies novel methylation markers for multiple malignancies. *PLoS Med* 2006; 3: e486.
52. Bonito NA, Borley J, Wilhelm-Benartzi CS, *et al.* Epigenetic regulation of the homeobox gene MSX1 associates with platinum-resistant disease in high-grade serous epithelial ovarian cancer. *Clin Cancer Res* 2016; 22: 3097–3104.
53. Tan IB, Ivanova T, Lim KH, *et al.* Intrinsic subtypes of gastric cancer, based on gene expression pattern, predict survival and respond differently to chemotherapy. *Gastroenterology* 2011; 141: 476–485, 485.e1–11.
54. Lee HT, Oh S, Ro DH, *et al.* The key role of DNA methylation and histone acetylation in epigenetics of atherosclerosis. *J Lipid Atheroscler* 2020; 9: 419–434.

Visit SAGE journals online
[journals.sagepub.com/
 home/tam](https://journals.sagepub.com/home/tam)

 SAGE journals