CANCER GENOMICS & PROTEOMICS *19*: 556-569 (2022)
doi:10.21873/cgp.20340

Mutational Signatures Associate With Survival in Gastrointestinal Carcinomas

PEETER KARIHTALA[1], KATJA PORVARI[2] and OUTI KILPIVAARA[3,4,5]

[1]Department of Oncology, Helsinki University Hospital Comprehensive Cancer Center and University of Helsinki, Helsinki, Finland;
[2]Department of Pathology, Medical Research Center Oulu, Oulu University Hospital and University of Oulu, Oulu, Finland;
[3]Applied Tumor Genomics Research Program, Faculty of Medicine, University of Helsinki, Helsinki, Finland;
[4]Department of Medical and Clinical Genetics, Medicum, Faculty of Medicine, University of Helsinki, Helsinki, Finland;
[5]HUSLAB Laboratory of Genetics, HUS Diagnostic Center, Helsinki University Hospital, Helsinki, Finland

Abstract. *Background/Aim: Mutational signatures reflect common patterns based on the counts of mutations and their sequence context. The prognostic value of these signatures, mirroring various carcinogenetic processes of cancers, are unexplored in gastrointestinal cancers. Our aim was to evaluate possible prognostic relevance of mutational signatures in gastrointestinal carcinomas after adjusting with the traditional prognostic factors. Materials and Methods: We used publicly available data from The Cancer Genome Atlas and Pan-Cancer Analysis of Whole Genomes to evaluate the associations between survival endpoints and activity of mutational signatures in seven types of gastrointestinal cancers. Results: Most strikingly, the high activity of age-related single-base substitution 5 (SBS5) and SBS40 signatures were in rectal adenocarcinomas associated with both improved overall survival (OS) [for SBS5 hazard ratio (HR) 0.130; 95% CI=0.03-0.56, for SBS40 HR=0.072; 95% CI=0.012-0.44, respectively] and similarly also to rectal cancer-specific survival. In patients with left-sided (but not right-sided) colon adenocarcinoma, the high activity of SBS2 signatures, formed due to APOBEC activity, predicted shortened OS. In pancreatic cancer, the high activity of SBS10b, caused by polymerase epsilon exonuclease proofreading defects, was associated both with longer OS (HR=0.44; 95% CI=0.205-0.96) and pancreatic cancer-specific survival (HR=0.32; 95% CI=0.112-0.91). Conclusion: Several mutational signatures seem to have clinically meaningful, cancer-specific associations with prognosis among gastrointestinal cancers.*

Correspondence to: Prof. Peeter Karihtala, MD, Ph.D., Department of Oncology, Helsinki University Hospital Comprehensive Cancer Center, P.O. Box 180, FI-00029 Helsinki, Finland. Tel: +358 94711, Fax: +358 947174205, e-mail: peeter.karihtala@hus.fi

Key Words: Colorectal cancer, COSMIC, mutations, pancreatic cancer, prognosis.

Globally, six out of 14 most common cancer types are gastrointestinal (GI) cancers, and they cover more than one quarter of all cancer deaths (1). GI cancers are very diverse in terms of their risk factors, geographical incidence, and prognosis. As a unifying factor, there are still huge gaps in their etiological research and reliable prognostic factors beyond TNM classification are rare. The 5-year survival of GI cancers varies from up to 70% of rectal adenocarcinomas to very few long-time survivors with pancreatic adenocarcinoma (1). Survival rates within the patients with cancer in same anatomical location may still vary a lot, and more precise and reproducible prognostic factors are needed to optimize surgical and oncological treatments, and surveillance.

Both endogenous and environmental sources of mutagenesis cause consistently identifiable patterns of mutations, mutational signatures, which reflect different carcinogenetic pathways (2). By examining the frequency of these signatures, insights to *e.g.*, past exposure to carcinogens and DNA repair mechanism defects can be achieved. Single-base substitution (SBS) mutational signatures consist currently of 53 distinct SBS in the Catalogue of Somatic Mutations in Cancer (COSMIC) database (3). While the etiology of some SBSs is still unknown, some signatures are caused by a specific DNA proofreading defects (*e.g.*, SBS10), others are related to exposure to specific chemotherapies (*e.g.*, SBS17) or may be secondary to tobacco chewing (*e.g.*, SBS29) or smoking (*e.g.*,

SBS4) (3, 4). Doublet base substitutions (DBS), which are generated after the concurrent modification of two consecutive nucleotide bases, and signatures of small insertions or deletions, known as indels, were only recently introduced to COSMIC and are still quite unexplored (3, 4).

As targeted therapeutics advance in GI cancers, whole-genome and whole-exome sequencing data are likely to become more common in the future. Although single mutations may offer prognostic value, exemplified by *BRAF* and *RAS* mutations in colorectal cancer, mutational signatures could provide reproducible and much more comprehensive insight to the aggressiveness of GI carcinomas by reflecting their carcinogenetic processes.

The study of mutational signatures is a rapidly emerging field of cancer research, but the association between mutational signatures and survival in the patients with GI carcinomas has not been thus far assessed. To elucidate this, we used publicly available The Cancer Genome Atlas (TCGA) whole-exome data to evaluate possible prognostic relevance of COSMIC signatures after adjusting with the traditional prognostic factors. The analyses were complemented with Pan-Cancer Analysis of Whole Genomes (PCAWG) whole-genome data, where appropriate.

Materials and Methods

Data. Mutational signature activity data (3) were accessed from the ICGC data portal (5). The data comprised of whole-genome sequenced tumors from the PCAWG consortium and whole-exome sequenced tumors from the TCGA. SBS, DBS and ID signatures were available for the PCAWG samples, and for the TCGA samples only SBS and ID signatures were available. Metadata on the PCAWG tumor samples were accessed from the ICGC data portal (6) and curated metadata for TCGA samples were accessed from supplementary Table I of (7). The data files were read and subjected to all further analysis using R, v. 4.0.2 (8). All statistical analyses of this study were performed by biomedical statisticians.

Colon cancers were divided for two groups for our analysis. Right-sided colon cancers included caecum, ascending colon, hepatic flexure of colon and transverse colon while left-sided colon cancers consisted of splenic flexure of colon, descending colon, and sigmoid colon.

Univariate survival analysis. The association between mutational signatures and overall survival (OS) was first tested in a univariate approach utilizing the R packages *survival*, v. 3.2-7 (9) and *survminer*, v. 0.4.8 (10). Only primary tumor samples and patients with available vital status and survival/follow-up time were included. TCGA cancer types and PCAWG cancer types were analyzed separately and, additionally, all TCGA samples were analyzed as one group and all PCAWG samples as one group. The association between the signature activity and in each cancer type was analyzed if the following criteria were met: 1) at least 20 samples had both signature and survival data 2) there were at least five death events among the patients and 3) there were at least five samples with non-zero signature activity. The association to survival was then tested using the log-rank test between low activity and high-activity tumors for the given signature. Low-activity tumors were defined as those with a median or lower activity of the signature within the cancer type, and high-activity tumors as those with above-median activity. For each analysis, a Kaplan-Meier curve was plotted using the function *ggsurvplot*.

Multivariate survival analysis. After the univariate survival analysis, a multivariate survival analysis was performed for seven selected GI cancer types using the TCGA data (CHOL, COAD, ESCA, LIHC, PAAD, READ and STAD), again using the packages *survival* and *survminer*. This analysis was carried out as a cancer type specific Cox proportional hazards regression with multiple variables: all signatures with a non-zero activity in at least 5% of the patients within the cancer type and selected clinical variables. Four alternative endpoints were used, generating separate regression models: OS, disease-specific survival (DSS), progression-free interval (PFI) and disease-free interval (DFI). The function *cox.zph* was used to test the proportional hazards assumptions and plot the Schoenfeld residuals for each variable and the combined model. Function *coxph* was used to run the Cox regression. Forest plots illustrating the hazard ratios of each variable were generated using the function *ggforest* of *survminer* package. This Cox regression was performed on seven cancer types of the GI tract treating the signature activities as binary high *vs*. low variables thresholded at their median as in the univariate analysis. In some cases, there were no non-censored patients (*i.e*., patients with a qualifying progression event) with a particular value of a clinical variable, leading to the failure to estimate their coefficients. In such cases, problematic variables were either left out or patients with particular values of that variable were left out or merged. Cross-tables for each of the seven GI tumors indicating the used clinical variables, mutational signatures, and the number of patients with a 1) low-activity and 2) high-activity status by each clinical variable value were produced using the R package *gtsummary* v.1.4.1 (11).

Signature-mutation count correlations. For the seven GI tumor types, the correlation between the activity of the SBS signatures included in the Cox regression and the total count of small somatic mutations was calculated across all patients of a given cancer type. The correlation coefficients and *p*-values for both Pearson's linear correlation and Spearman's rank correlation were calculated, and the association between signature activity and mutation count was also visualized using scatterplots generated by *ggscatter* function of R package *ggpubr*, v.0.4.0 (10).

Signature versus clinical variable associations. For the seven GI tumors, the association between the activity of the SBS signatures included in the Cox regression and the clinical variables was tested. For this, the patients were binarized into low- and high-activity groups as previously, and a *p*-value was calculated using Fisher's exact test. A significant *p*-value indicates that the low- and high-activity patients fall into the categories of the clinical variable non-randomly. The statistical testing and result table generation was performed using the package *gtsummary*.

The definition of the endpoints. TCGA data has four different endpoints: OS, DSS, DFI and PFI. These endpoints are explained in detail in Liu et al. 2018 (7). Briefly, OS is the period from the date of diagnosis until the date of death from any cause, and DSS until the date of death from the specific cancer. PFI is the period

from the date of diagnosis until the date of the first occurrence of any new tumor event. DFI is the period from the date of diagnosis until the date of the first new tumor progression after achieving disease-free status. PCAWG data includes only OS as an endpoint.

Results

Several mutational signatures had associations with prognosis among gastrointestinal cancers. These performed Cox proportional hazard analyses are reported in more detail in Table I and the results are also summarized below and in Figure 1, Figure 2, Figure 3, Figure 4, Figure 5, and Figure 6.

High prevalence of SBS17b signatures associated with shorter OS in esophageal carcinomas (HR=11.47; 95% CI=1.187-110.8) (Figure 1). Again, the high number of SBS18 signature was a predictor of improved disease-free interval in esophageal carcinomas (HR=0.18; 95% CI=0.038-0.86). Likewise, SBS18 associated with longer OS in PCAWG data in univariate analysis (p=0.043) (Figure 2). This was the only statistically significant survival association observed in the PCAWG dataset.

In stomach adenocarcinomas, SBS1 associated with shorter DFI (HR=2.55; 95% CI=1.132-5.8) while SBS40 predicted improved stomach adenocarcinoma-specific survival (HR=0.50; 95% CI=0.25-0.98) (Figure 3).

In pancreatic cancer, SBS10b signature was found to be associated both with longer OS (HR=0.44; 95% CI=0.205-0.96) and pancreatic cancer-specific survival (HR=0.32; 95% CI=0.112-0.91) (Figure 4).

In patients with colon adenocarcinoma, high number of SBS2 signatures predicted shortened OS (HR=2.42; 95% CI=1.04-5.6), but this was observed only in the patients with left-sided colon adenocarcinoma (HR=4.97; 95% CI=1.005-24.6), when right-sided and left-sided colon adenocarcinomas were studied separately (Figure 5). In addition, the patients with left-sided colon adenocarcinoma with high SBS5 (HR=6.9; 95% CI=1.039-45.8) or high SBS40 (HR=9.80; 95% CI=1.444-66.5) signature activity in their tumors had decreased OS.

In rectal cancer, SBS5 and SBS40 signatures were associated with both notably improved OS and rectal cancer-specific survival (for SBS5 and OS HR=0.130; 95% CI=0.03-0.56, for SBS5 and DSS HR=0.0025; 95% CI=0.000042-0.15, for SBS40 and OS HR=0.072; 95% CI=0.012-0.44 and for SBS40 and DSS HR=0.00098; 95% CI=0.000013-0.076, respectively) (Figure 6). In addition, the high number of SBS17b signature (present only in nine patients) associated with worse rectal cancer-specific survival (HR=1800; 95% CI=2.9-1.1x10^7).

In the cohorts consisting of only right-sided colon adenocarcinomas, hepatocellular carcinomas or cholangiocarcinomas, no statistically significant associations between mutational signatures and any of the survival endpoints were observed in Cox regression analysis.

Signature-mutation count correlations. The total mutation count had a strong positive correlation with most of the studied mutational signatures (Figure 7). The number of SBS5 mutational signatures and mutation count had a very clear correlation in the all studied tumor types, with the exception of rectal adenocarcinoma and left-sided colon adenocarcinoma. Also, the number of SBS1 signatures had strong correlations with total mutation counts in all but the smallest cohorts of hepatocellular carcinomas and cholangiocarcinomas.

Furthermore, several inverse correlations between mutation count and mutational signatures were recorded. In esophageal carcinoma SBS16 (p=0.037; R=–0.15) and SBS18 (p=0.024; R=–0.17) correlated inversely with mutation count, in hepatocellular carcinoma SBS30 (p=0.0073; R=–0.15), in pancreatic adenocarcinoma SBS15 (p=0.0054; R=–0.21) and in stomach adenocarcinoma SBS2 had such a correlation (p=0.0049; R=–0.14). In right-sided colon adenocarcinoma, high SBS40 activity correlated with lower number of mutation count (p=0.00023; R=–0.28).

The associations between the traditional prognostic factors and signature activity in each of the studied cancer types are presented in Supplementary Tables I-VII.

Discussion

The impact of mutational signatures on patient survival has not been assessed previously. The herein presented data suggest that COSMIC mutational signatures have cancer-specific associations with diverse prognostic groups among the major GI cancers.

From all GI cancers, SBS1, SBS5 and SBS40 signatures have been most frequently found in colorectal adenocarcinomas and SBS5 is ubiquitous also in benign GI tissues (3, 12). Both SBS5 and SBS40 are flat signatures, and their misattribution has not been excluded (3, 13). The activity of both SBS5 and SBS40 correlate with age, but the etiology especially behind SBS40 is still unknown (3). According to our analysis, the high activity of both SBS5 and SBS40 associated with improved OS in patients with rectal adenocarcinoma. In addition, both SBS5 and SBS40 associated very convincingly with long DSS, the upper limit of 95%CI of the HR being as low as 0.076 for SBS5 and 0.15 for SBS40, respectively. These estimates exceed by far the traditional colorectal cancer prognostic factors, including stage. Nevertheless, rectal cancer-specific survival should be considered as an approximation in TCGA data, while OS is more strongly recommended for use (7). Intriguingly, SBS5 activity did not reflect the total number of mutations in rectal adenocarcinomas, although SBS40 had such a correlation. Neither SBS5 and SBS40 associated with stage, sex, or the site of primary carcinoma in rectum. This may emphasize their occurrence possibly in the early phases of carcinogenesis, but also suggests their independent roles as novel prognostic factors

Table I. *Summary of the Cox proportional hazard (PH) analyses from the TCGA data.*

Cancer type	End-point	Signatures eligible for the analysis	Excluded variables/ variable levels	Variables with rejected proportional hazards assumptions	Signatures with p-value <0.05 from the CoxPH test
Esophageal carcinoma	OS	SBS1, SBS2, SBS3, SBS5, SBS13, SBS16, SBS17a, SBS17b, SBS18	Tissue "Upper third of esophagus" and SBS16 excluded	Age, primary diagnosis	SBS17b
	DSS	SBS1, SBS2, SBS3, SBS5, SBS13, SBS16, SBS17a, SBS17b, SBS18	Squamous cell carcinoma, keratinizing, NOS included in Other category, Tissue or organ of origin excluded, SBS16 excluded	Age	
	DFI	SBS1, SBS2, SBS3, SBS5, SBS13, SBS16, SBS17a, SBS17b, SBS18	Stage not included in the model, Primary diagnosis categories Squamous cell carcinoma, keratinizing, NOS and Other excluded, signature SBS17b excluded	Primary diagnosis, SBS3 and SBS5	SBS18
	PFI		Squamous cell carcinoma, keratinizing, NOS included in Other category	SBS5	
Stomach adeno-carcinoma	OS	SBS1, SBS2, SBS5, SBS13, SBS15, SBS17a, SBS17b, SBS20, SBS40, SBS54	Tissue or organ categories "Lesser curvature of stomach, NOS", "Pylorus" excluded		
	DSS	SBS1, SBS2, SBS5, SBS13, SBS15, SBS17a, SBS17b, SBS20, SBS40, SBS54	Tissue or organ categories "Lesser curvature of stomach, NOS", "Pylorus" excluded	SBS40	
	DFI	SBS1, SBS2, SBS5, SBS13, SBS15, SBS17a, SBS17b, SBS20, SBS40, SBS54	Tissue or organ categories "Lesser curvature of stomach, NOS", "Pylorus" excluded, Primary diagnosis "Papillary adenocarcinoma, NOS" samples and stage IV samples excluded	SBS20	SBS1
	PFI	SBS1, SBS2, SBS5, SBS13, SBS15, SBS17a, SBS17b, SBS20, SBS40, SBS54	Tissue or organ categories "Lesser curvature of stomach, NOS", "Pylorus" excluded	SBS13	
Cholangio-carcinoma	OS	SBS1, SBS2, SBS5, SBS13, SBS40	Tissue or organ categories other than "Intrahepatic bile duct" were combined into "Other" category, stage III samples excluded, Primary diagnosis not included in the model as only one category	SBS5	
	DSS	SBS1, SBS2, SBS5, SBS13, SBS40	Tissue or organ categories other than "Intrahepatic bile duct" were combined into diagnosis "Other" category, stage III excluded, Primary not included in the model as only one category		
	DFI	SBS1, SBS2, SBS5, SBS13, SBS40	Tissue or organ, stage and Primary diagnosis not included in the model	Age, SBS2	
	PFI				
Pancreatic adeno-carcinoma	OS	SBS1, SBS5, SBS10b, SBS15	Primary diagnosis categories "Adenocarcinoma with mixed subtypes", "Carcinoma, undifferentiated, NOS", "Neuroendocrine carcinoma, NOS" and "Mucinous adenocarcinoma" were combined into "Other" category	Stage	SBS10b
	DSS	SBS1, SBS5, SBS10b, SBS15	Same as above		SBS10b

Table I. *Continued*

Table I. *Continued*

Cancer type	End-point	Signatures eligible for the analysis	Excluded variables/ variable levels	Variables with rejected proportional hazards assumptions	Signatures with p-value <0.05 from the CoxPH test
	DFI	SBS1, SBS5, SBS10b, SBS15	Same as above	Sex	
	PFI	SBS1, SBS5, SBS10b, SBS15	Same as above		
Hepatocellular carcinoma	OS	SBS1, SBS5, SBS9, SBS22, SBS30, SBS40			
	DSS	SBS1, SBS5, SBS9, SBS22, SBS30, SBS40		Stage	
	DFI	SBS1, SBS5, SBS9, SBS22, SBS30, SBS40	Stage IV excluded		
	PFI	SBS1, SBS5, SBS9, SBS22, SBS30, SBS40		SBS1	
Rectal adeno-carcinoma	OS	SBS1, SBS2, SBS5, SBS17b, SBS40	Primary diagnosis categories other than "Adenocarcinoma, NOS" were combined into "Other" category as only few samples, Tissue categories other than "Rectum, NOS" and "Rectosigmoid junction" were combined into "Other" category		SBS5, SBS17b, SBS40
	DSS	SBS1, SBS2, SBS5, SBS17b, SBS40	Stage excluded as 3 categories had inflated coefficients		
	DFI	SBS1, SBS2, SBS5, SBS17b, SBS40	Primary diagnosis and Tissue or organ of origin excluded	Age, Sex	
	PFI	SBS1, SBS2, SBS5, SBS17b, SBS40	Primary diagnosis categories other than "Adenocarcinoma, NOS" were combined into "Other" category as only few samples, Tissue categories other than "Rectum, NOS" and "Rectosigmoid junction" were combined into "Other" category		
Colon adeno-carcinoma	OS	SBS1, SBS2, SBS5, SBS15, SBS40, SBS44	Tissue or organ of origin excluded from the model	SBS5	SBS2
	DSS	SBS1, SBS2, SBS5, SBS15, SBS40, SBS44	Tissue or organ of origin excluded from the model	SBS5	
	DFI	SBS1, SBS2, SBS5, SBS15, SBS40, SBS44	Tissue or organ of origin and Primary diagnosis not included in the model, stage IV and SBS15 excluded		
	PFI	SBS1, SBS2, SBS5, SBS15, SBS40, SBS44	Tissue or organ of origin excluded from the model		
Colon adeno-carcinoma, right-side	OS	SBS1, SBS2, SBS5, SBS15, SBS40, SBS44	Tissue Hepatic flexure of colon excluded		
	DSS	SBS1, SBS2, SBS5, SBS15, SBS40, SBS44	Tissue Hepatic flexure of colon excluded		
	DFI	SBS1, SBS2, SBS5, SBS15, SBS40, SBS44	Stage and Primary diagnosis not included in the model, Tissues Transverse colon and Hepatic flexure of colon excluded, SBS2, SBS15 and SBS44 excluded		
		SBS1, SBS2, SBS5, SBS15, SBS40, SBS44	Tissue Hepatic flexure of colon excluded		
Colon adeno-carcinoma, left-side	OS	SBS1, SBS2, SBS5, SBS15, SBS40, SBS44	SBS44 excluded	Sex	SBS2, SBS5, SBS40
	DSS	SBS1, SBS2, SBS5, SBS15, SBS40, SBS44	Primary diagnosis not included in the model, Tissue Splenic flexure of colon and AJCC pathologic stage I excluded, SBS15 and SBS44 excluded	Stage	
	DFI	SBS1, SBS2, SBS5, SBS15, SBS40, SBS44	Primary diagnosis and Tissue or organ of origin not included in the model, stage IV excluded, SBS2, SBS15 and SBS44 excluded	Sex, SBS5	
	PFI	SBS1, SBS2, SBS5, SBS15, SBS40, SBS44	SBS15 excluded	Tissue or organ of origin, SBS5, SBS40	

DFI: Disease-free interval; DSS: disease-specific death; OS: overall survival; PFI: progression-free interval; SBS: single-base substitution.

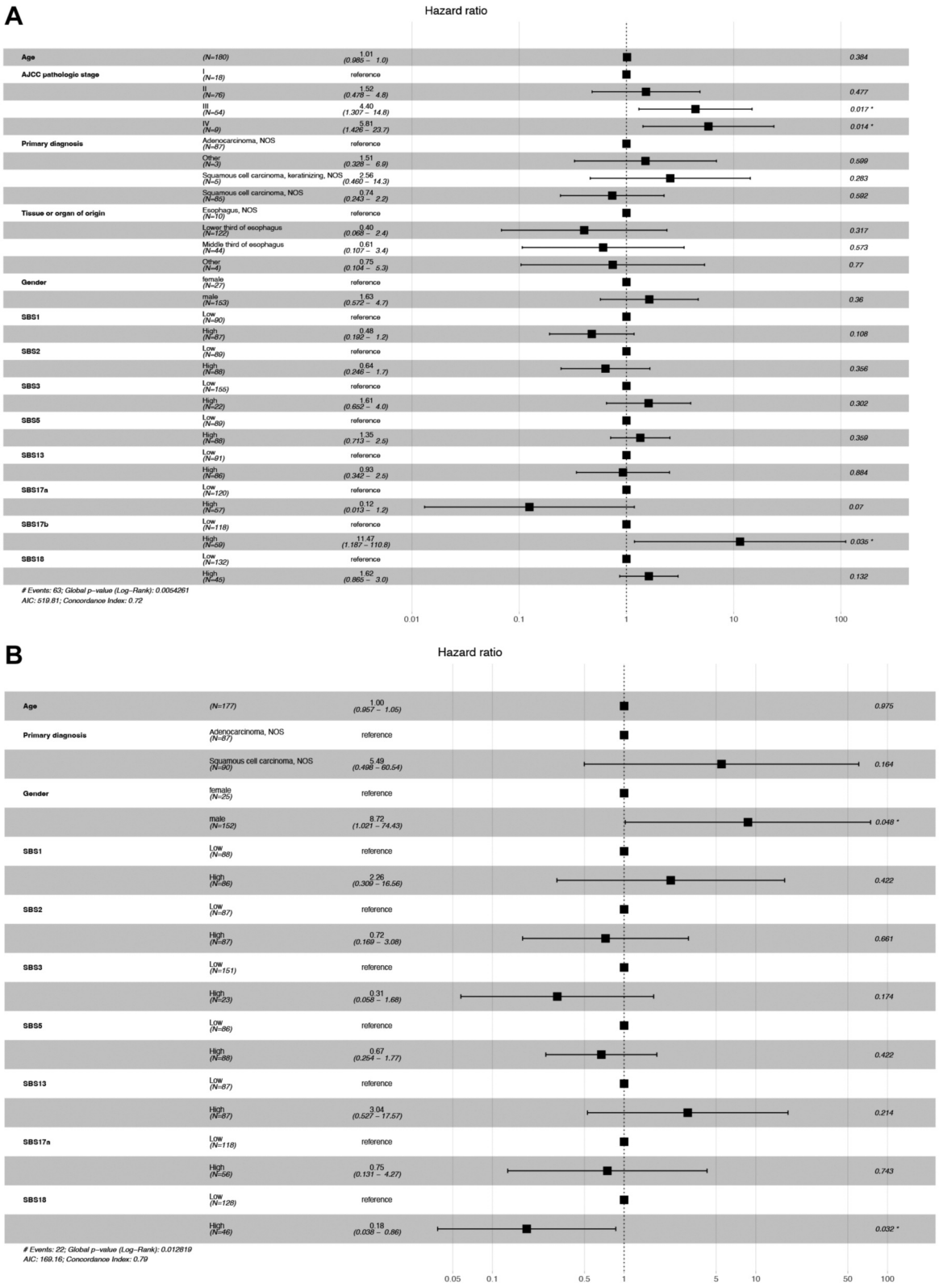

Figure 1. *Cox proportional hazards regression analysis results for esophageal carcinoma in TCGA. Higher stage and SBS17b prevalence were associated with worse overall survival (A). Male sex predicted worse and the low activity of SBS18 improved disease-free interval (B). AJCC: American Joint Committee on Cancer; NOS: not otherwise specified; SBS: single-base substitution.*

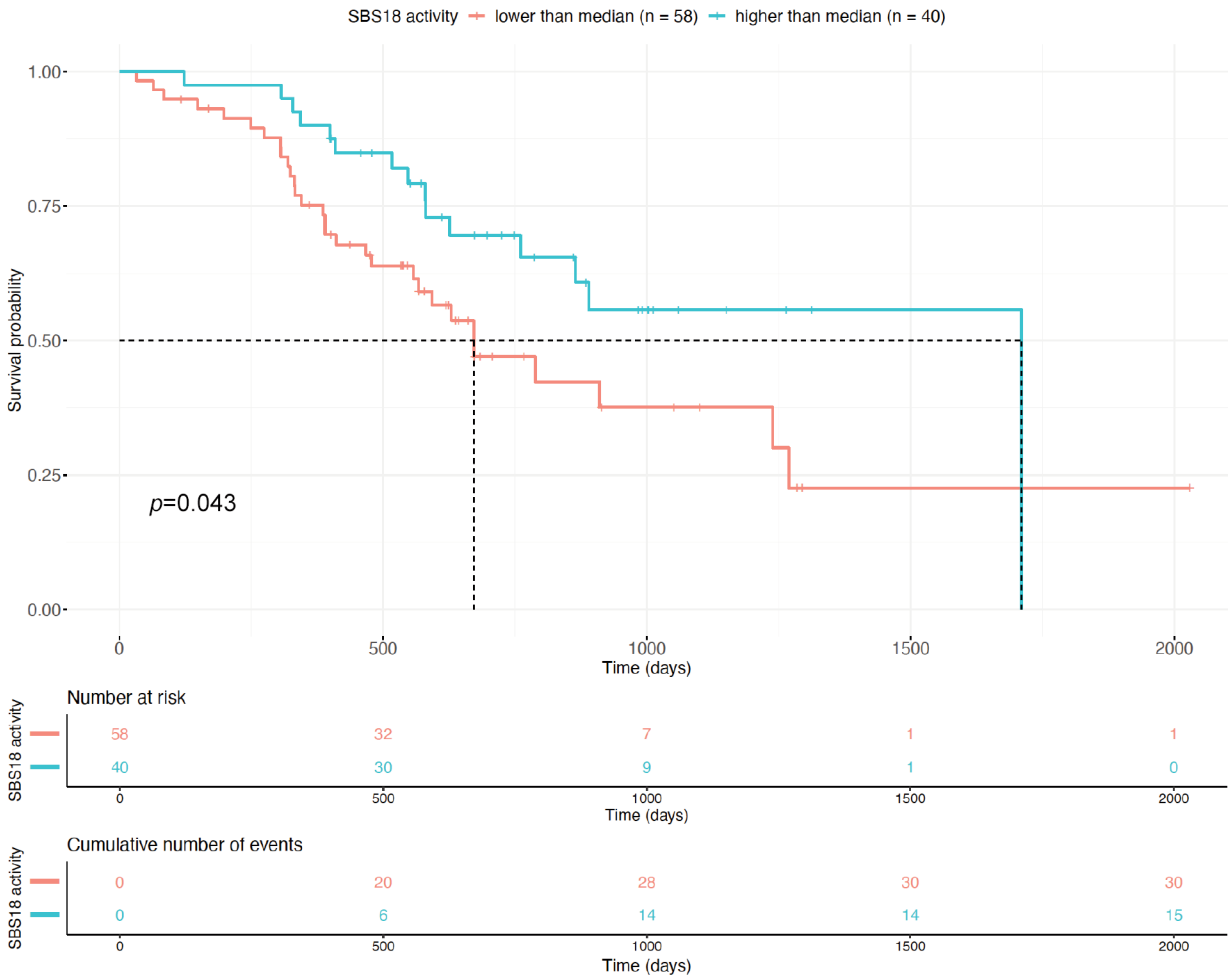

Figure 2. *SBS18 activity higher than median associated with improved overall survival in the cohort of The Pan-Cancer Analysis of Whole Genomes esophageal carcinoma patients (log-rank p-value=0.043).*

in rectal adenocarcinomas. These results are in contradiction with the inherent clock-like nature of these SBS5 and SBS40 and suggest that their association to non-malignant behavior of rectal adenocarcinomas should be assessed in more mechanistical studies.

SBS17b is considered an easily trackable signature and it has a marked characteristic of T>G substitutions, possibly caused by oxidative damage in the nucleotide pool (4). Although high SBS17b activity was observed only in 7% of the evaluable patients with rectal adenocarcinoma, it associated very strongly with poor rectal adenocarcinoma-specific survival. Again, these DSS rates should be interpreted with caution due to relatively short follow-up (7). There is an enrichment of high SBS17b activity in tumors treated with 5-fluorouracil or capecitabine, which are widely used agents in stage III rectal adenocarcinomas requiring perioperative therapy (3, 14). The association between high SBS17b activity and poor survival was still independent from stage. Recent evidence suggests that high SBS17b activity is also connected to anti-EGFR antibody resistance in colorectal cancer (15).

Sidedness affects a wide spectrum of CRC features, including consensus molecular subtypes and microbiome, but has also a clinically meaningful impact on prognosis and treatment in metastatic setting (16). Contradictory to what was observed in patients with rectal adenocarcinoma, we observed dismal OS outcome in patients with left-sided colon adenocarcinoma, who had either high SBS5 or SBS40 activity. Such associations were not observed in patients with adenocarcinoma originating from the right side of colon and thus the clock-like nature of these signatures does not explain these results. Whether preoperative radiotherapy, used in rectal adenocarcinomas but not in colon adenocarcinomas, could induce SBS5 and SBS40 is currently

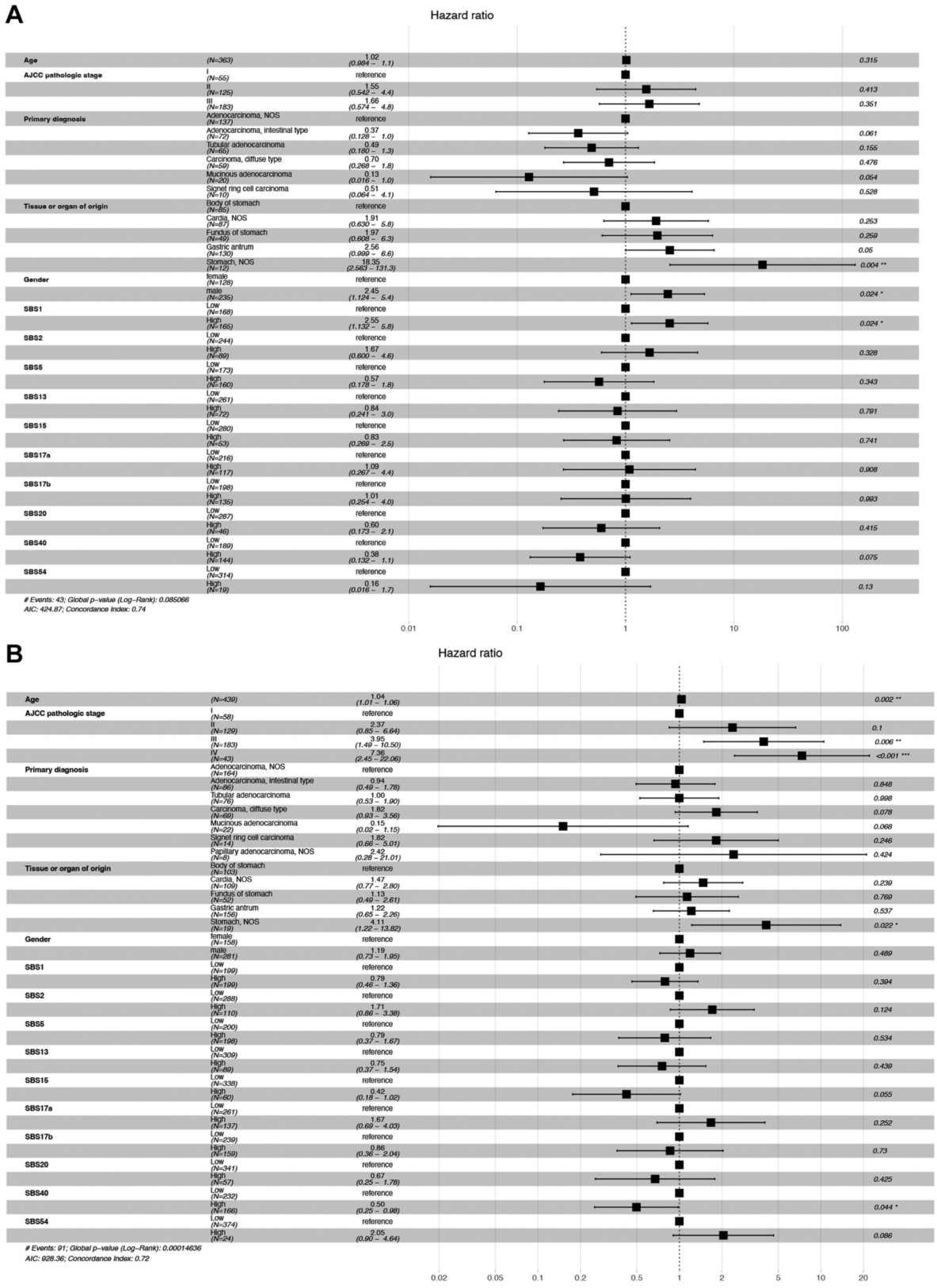

Figure 3. *In TCGA material of stomach adenocarcinoma patients, the high activity of SBS1 associated with worse disease-free interval (A), while the higher activity of SBS40 associated with improved disease-specific survival (B). NOS: Not otherwise specified; SBS: single-base substitution.*

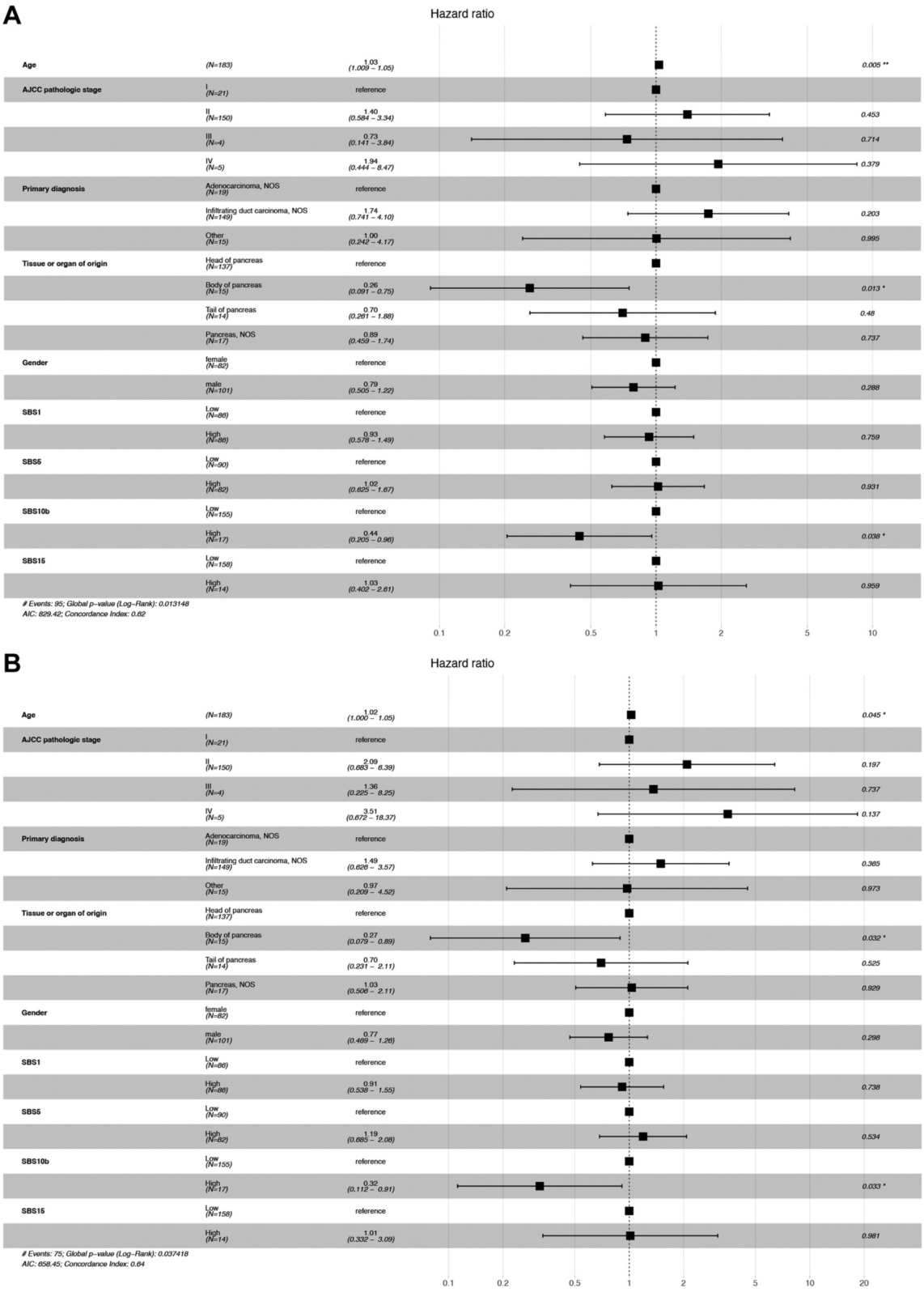

Figure 4. *Cox proportional hazards regression analysis results for pancreatic adenocarcinoma in TCGA cohort. The high activity of SBS10b signature was found to be associated both with longer overall survival (A) and pancreatic cancer-specific survival (B). NOS: Not otherwise specified; SBS: single-base substitution.*

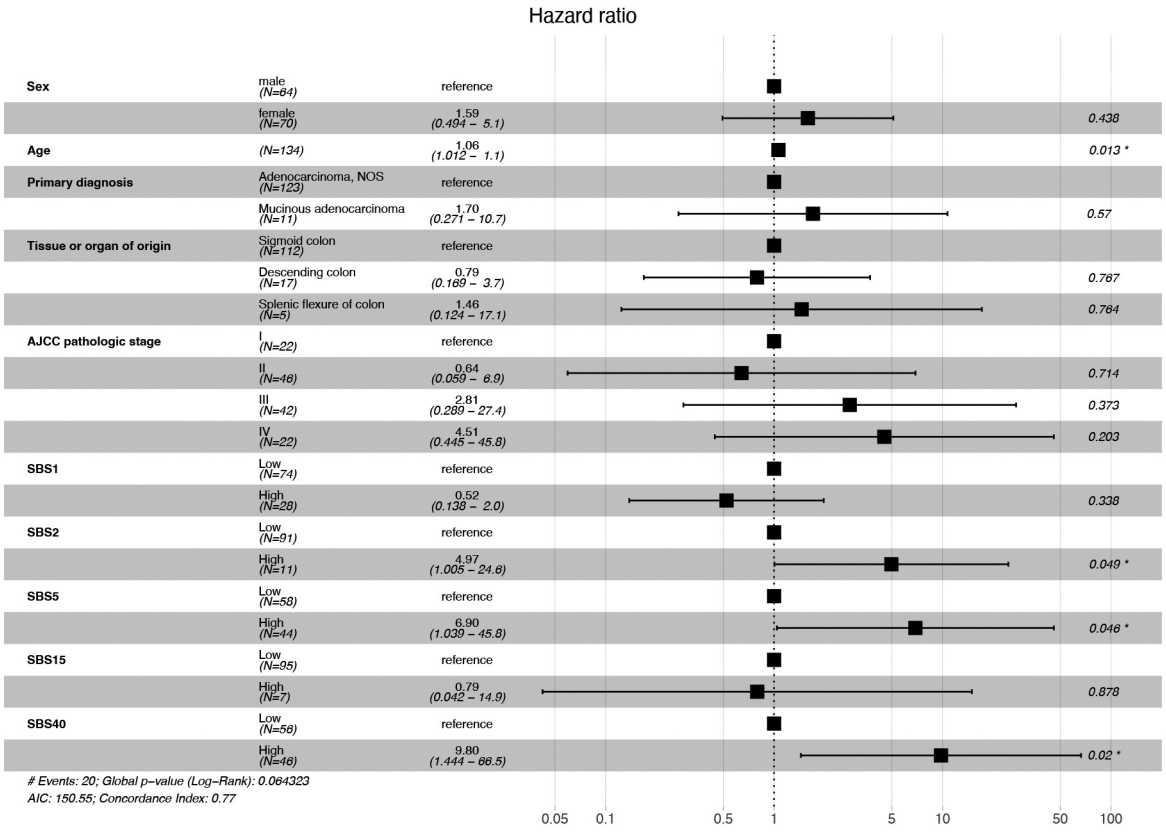

Figure 5. *Cox proportional hazards regression analysis results for left-sided colon adenocarcinoma in TCGA, using overall survival as the endpoint. The high number of SBS2, SBS5 and SBS40 predicted shortened overall survival. AJCC: American Joint Committee on Cancer; NOS: not otherwise specified; SBS: single-base substitution.*

unknown. High SBS2 activity predicted poor OS in patients with left-sided colon adenocarcinoma, but again not in those whose primary tumor was right-sided. SBS2 signature is formed due to APOBEC activity and is one of the most well-defined mutational signatures (4, 17). Although the prognostic value of mutational signatures has been previously described hardly in any malignancies, APOBEC signatures seem to associate with high mutational load and worse OS also in the patients with multiple myeloma (18).

Pancreatic adenocarcinoma has a dismal prognosis and lacks reproducible genetic or molecular prognostic biomarkers. SBS10 signature was recently split into distinct SBS10a and SBS10b signatures, which are both caused by polymerase epsilon exonuclease (POLE) proofreading defects (3, 19, 20). Specific *POLE* mutations define in endometrial carcinomas a specific ultramutated subtype, with improved prognosis (21). In line with this, the high activity of SBS10b signature in the patients with pancreatic adenocarcinoma was present in 10% of patients and it associated with longer DSS and OS, with a similar magnitude of effect for both endpoints. Both OS and DSS are considered as reliable endpoints in the TCGA data of pancreatic adenocarcinomas (7). SBS10b was not associated with any of the studied traditional prognostic factors or total mutation load. It is worth emphasizing that the stage distribution of pancreatic adenocarcinomas in TCGA does not represent the distribution seen in usual clinical practice, as 95% cases in the TCGA dataset are stage I-II tumors. Thus, these results suggest that the high activity of SBS10b, indicating POLE repair deficiency and hypermutator phenotype, may be a novel prognostic factor in early pancreatic adenocarcinoma. This could offer a possibility to guide risk-based adjuvant therapy and surveillance after surgery with curative intention. In advanced endometrial cancers, high SBS10 signature activity has been connected to improved response for checkpoint inhibitors, probably due to increased neoepitope formation (4, 19, 20).

In contrast to other GI carcinomas, there are some preliminary data regarding the COSMIC signatures and survival in patients with esophageal and gastroesophageal junction (GEJ) carcinoma. High SBS17 activity (without a separation to SBS17a and SBS17b) predicted worse survival

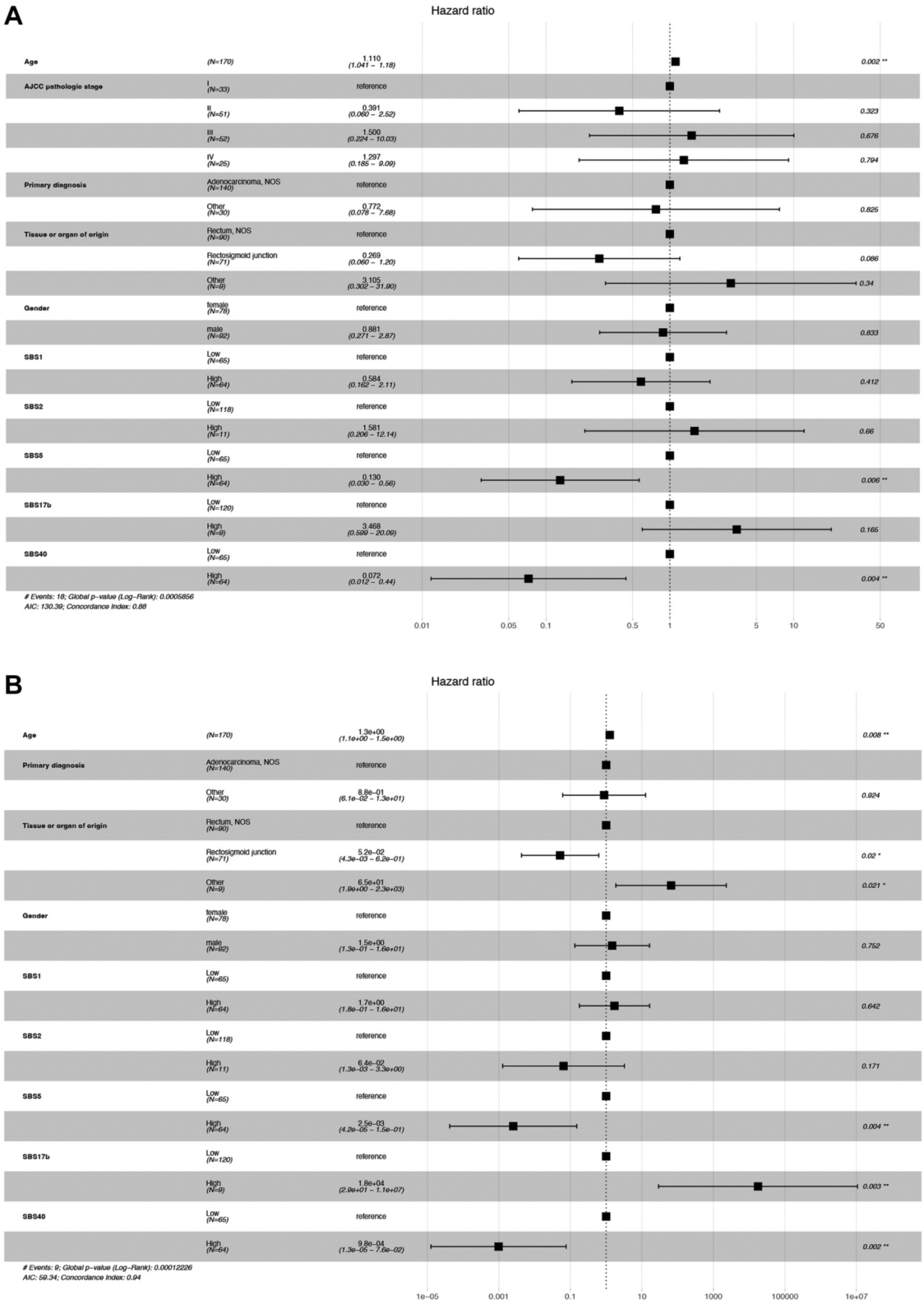

Figure 6. *Cox proportional hazards regression analysis results for rectal adenocarcinoma in TCGA cohort. Both SBS5 and SBS40 signatures were associated with both notably improved overall survival (A) and rectal cancer-specific survival (B), while the high number of SBS17b was associated with worse rectal cancer-specific survival. AJCC: American Joint Committee on Cancer; NOS: not otherwise specified; SBS: single-base substitution.*

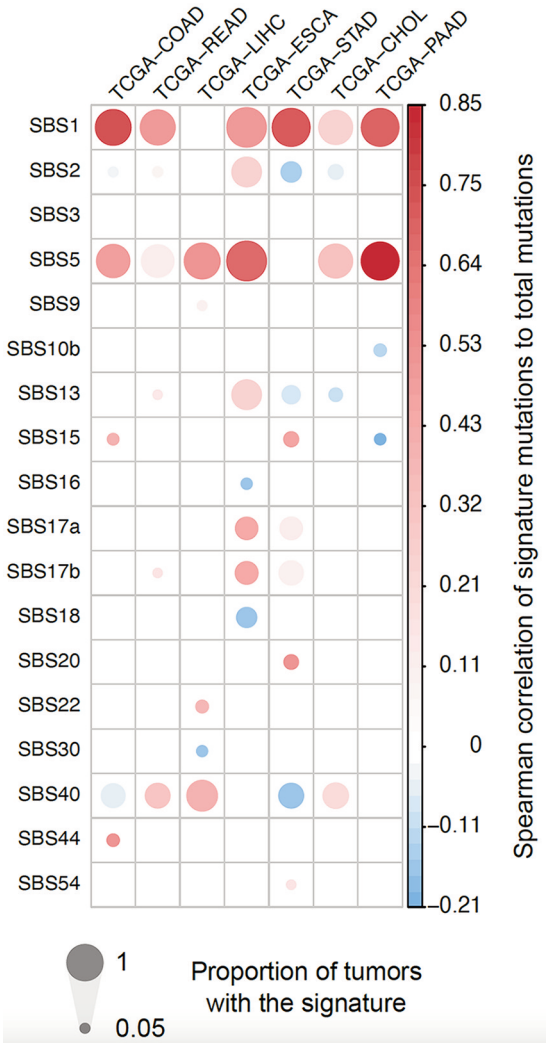

Figure 7. *Correlations of mutational signatures in relation to the total mutation rate in the studied tumor types in TCGA material. Red indicates higher and blue lower values, respectively. The size of the dots indicates the proportion of tumors with each signature. COAD: Colon adenocarcinoma; CHOL: cholangiocarcinoma; ESCA: esophageal carcinoma; LIHC: hepatocellular carcinoma; PAAD: pancreatic adenocarcinoma; READ: rectal adenocarcinoma; STAD: stomach adenocarcinoma; TCGA: The Cancer Genome Atlas.*

in the material of 83 Chinese GEJ adenocarcinoma patients (22). Also in a small set of esophageal adenocarcinomas, the characteristic of SBS signatures, SBS, 5′-C[T>G]T-3′, predicted worse OS in univariate analysis (23). In line with these data, and with the current results from rectal adenocarcinomas, SBS17b activity predicted poor OS outcomes in esophageal carcinomas, consisting almost solely of adenocarcinomas.

SBS1 associated in stomach adenocarcinomas with short DFI, which is considered as a reliable endpoint in TCGA stomach adenocarcinoma cohort (7). SBS1 correlates tightly with age and mutation load in most cancers, also in stomach adenocarcinomas (3, 24) and age is one of the most important prognostic factors in stomach cancer (25).

Both SBS17b and SBS18 signatures arise after cellular stress, especially due to reactive oxygen species (4). Particularly, SBS17 signatures may be a consequence of exposure for gastric acid or 5-fluorouracil/capecitabine (4), which are one of the most applied chemotherapeutic agents also in esophageal carcinomas. It is possible, that the linkage between SBS17b and shorter survival in patients with esophageal carcinomas may just reflect the increased use chemotherapy or (chemo)radiotherapy in the most aggressive esophageal carcinomas, which would have consequently led to increased number of SBS17b signatures. From all GI carcinomas, SBS18 is the most prevalent in both esophageal carcinomas (3, 26). High SBS18 activity associated with improved DFI in TCGA data and also with prolonged OS in PCAWG dataset. This was the only result from whole-exome TCGA dataset, which could be confirmed in the whole-genome PCAWG data. Taken together, it seems that there is a different origin of signature between SBS17b and SBS18 and consequently diverse contribution to survival in esophageal carcinomas.

There are several limitations in our study, which mostly relate to the nature of TCGA data. We did not have the treatment data available, although we were able to use various other clinically important prognostic factors as co-variates in multivariate analysis. Using PFI and DFI may be criticized for them being surrogates rather than clinically hard endpoints. On the other hand, in TCGA cohorts PFI and DFI are considered the most reliable endpoints due to generally sufficient follow-up (7). Again, DSS results should be interpreted with caution in most cancer types as discussed above (7). As TCGA is based on exome data, only a small proportion of mutations in human genome footprint are covered. This is also a likely reason for our final analyses including only SBSs, but not other types of signatures. The results from PCAWG data were mainly not in concordance with TCGA results. This is not surprising since PCAWG whole-genome data consists of only one endpoint, OS, it lacks the largely clinical variables, and above all, has a limited sample size.

Conclusion

We conclude that several COSMIC mutational signatures seem to have an independent prognostic role among GI cancers. This is highlighted by tremendously improved survival in rectal adenocarcinoma patients with high SBS5 and SBS40 activity. Both SBS5 and SBS40 are rather poorly characterized signatures in terms of their activities at molecular level and more experimental studies are required to resolve this. Some mutational signatures had similar prognostic impact in different

sites of cancer origin, as exemplified by poor outcome in the patients with esophageal or rectal cancer and high SBS17b activity. It should be still kept in mind that the observed results are associations, not causations. In following studies, the carcinogenetic processes behind of certain signatures should be clarified and the current results should be confirmed in a material with longer follow-up.

Supplementary Material

Available at: https://www.dropbox.com/sh/9lyfv5s0bo8u0sk/AAAncv9tMsXBigA74sr_cl3Ya?dl=0

Conflicts of Interest

The Authors have no conflicts of interest related to the manuscript.

Authors' Contributions

PK drafted the manuscript. All Authors participated in the planning the study, evaluating the results, and writing the final versions of the manuscript.

Acknowledgements

We thank Genevia Technologies for their help in statistical analyses and in composing the Figures and Tables.

References

1 Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A and Bray F: Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin *71(3)*: 209-249, 2021. PMID: 33538338. DOI: 10.3322/caac.21660

2 Alexandrov LB and Stratton MR: Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. Curr Opin Genet Dev *24*: 52-60, 2014. PMID: 24657537. DOI: 10.1016/j.gde.2013.11.014

3 Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, Boot A, Covington KR, Gordenin DA, Bergstrom EN, Islam SMA, Lopez-Bigas N, Klimczak LJ, McPherson JR, Morganella S, Sabarinathan R, Wheeler DA, Mustonen V, PCAWG Mutational Signatures Working Group, Getz G, Rozen SG, Stratton MR and PCAWG Consortium: The repertoire of mutational signatures in human cancer. Nature *578(7793)*: 94-101, 2020. PMID: 32025018. DOI: 10.1038/s41586-020-1943-3

4 Koh G, Degasperi A, Zou X, Momen S and Nik-Zainal S: Mutational signatures: emerging concepts, caveats and clinical applications. Nat Rev Cancer *21(10)*: 619-637, 2021. PMID: 34316057. DOI: 10.1038/s41568-021-00377-7

5 ICGC data portal. Available at: https://dcc.icgc.org/releases/PCAWG/mutationalsignatures/SignaturesinSamples/SPSignaturesinSamples [Last accessed on November 1, 2021]

6 ICGC data portal. Available at: https://dcc.icgc.org/releases/PCAWG/clinicalandhistology [Last accessed on November 1, 2021]

7 Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, Kovatich AJ, Benz CC, Levine DA, Lee AV, Omberg L, Wolf DM, Shriver CD, Thorsson V, Cancer Genome Atlas Research Network and Hu H: An integrated TCGA Pan-Cancer clinical data resource to drive high-quality survival outcome analytics. Cell *173(2)*: 400-416.e11, 2018. PMID: 29625055. DOI: 10.1016/j.cell.2018.02.052

8 R Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: http://www.R-project.org/ [Last accessed on November 1, 2021]

9 Therneau T: A package for survival analysis in R. R package version 3.2-7. Available at: https://CRAN.R-project.org/package=survival [Last accessed on November 1, 2021]

10 Kassambara A, Kosinski M and Biecek P: survminer: Drawing survival curves using 'ggplot2'. R package version 0.4.8. Available at: https://CRAN.R-project.org/package=survminer [Last accessed on November 1, 2021]

11 Sjoberg DD, Whiting K, Curry M, Lavery JA and Larmarange J: Reproducible summary tables with the gtsummary package. The R Journal *13(1)*: 570-580, 2021.

12 Li R, Di L, Li J, Fan W, Liu Y, Guo W, Liu W, Liu L, Li Q, Chen L, Chen Y, Miao C, Liu H, Wang Y, Ma Y, Xu D, Lin D, Huang Y, Wang J, Bai F and Wu C: A body map of somatic mutagenesis in morphologically normal human tissues. Nature *597(7876)*: 398-403, 2021. PMID: 34433965. DOI: 10.1038/s41586-021-03836-1

13 Moore L, Cagan A, Coorens THH, Neville MDC, Sanghvi R, Sanders MA, Oliver TRW, Leongamornlert D, Ellis P, Noorani A, Mitchell TJ, Butler TM, Hooks Y, Warren AY, Jorgensen M, Dawson KJ, Menzies A, O'Neill L, Latimer C, Teng M, van Boxtel R, Iacobuzio-Donahue CA, Martincorena I, Heer R, Campbell PJ, Fitzgerald RC, Stratton MR and Rahbari R: The mutational landscape of human somatic and germline cells. Nature *597(7876)*: 381-386, 2021. PMID: 34433962. DOI: 10.1038/s41586-021-03822-7

14 Ishaque N, Abba ML, Hauser C, Patil N, Paramasivam N, Huebschmann D, Leupold JH, Balasubramanian GP, Kleinheinz K, Toprak UH, Hutter B, Benner A, Shavinskaya A, Zhou C, Gu Z, Kerssemakers J, Marx A, Moniuszko M, Kozlowski M, Reszec J, Niklinski J, Eils J, Schlesner M, Eils R, Brors B and Allgayer H: Whole genome sequencing puts forward hypotheses on metastasis evolution and therapy in colorectal cancer. Nat Commun *9(1)*: 4782, 2018. PMID: 30429477. DOI: 10.1038/s41467-018-07041-z

15 Woolston A, Barber LJ, Griffiths B, Pich O, Lopez-Bigas N, Matthews N, Rao S, Watkins D, Chau I, Starling N, Cunningham D and Gerlinger M: Mutational signatures impact the evolution of anti-EGFR antibody resistance in colorectal cancer. Nat Ecol Evol *5(7)*: 1024-1032, 2021. PMID: 34017094. DOI: 10.1038/s41559-021-01470-8

16 Molina-Cerrillo J, San Román M, Pozas J, Alonso-Gordoa T, Pozas M, Conde E, Rosas M, Grande E, García-Bermejo ML and Carrato A: BRAF mutated colorectal cancer: New treatment approaches. Cancers (Basel) *12(6)*: 1571, 2020. PMID: 32545884. DOI: 10.3390/cancers12061571

17 Roberts SA, Lawrence MS, Klimczak LJ, Grimm SA, Fargo D, Stojanov P, Kiezun A, Kryukov GV, Carter SL, Saksena G, Harris S, Shah RR, Resnick MA, Getz G and Gordenin DA: An APOBEC cytidine deaminase mutagenesis pattern is widespread

in human cancers. Nat Genet *45(9)*: 970-976, 2013. PMID: 23852170. DOI: 10.1038/ng.2702

18 Walker BA, Wardell CP, Murison A, Boyle EM, Begum DB, Dahir NM, Proszek PZ, Melchor L, Pawlyn C, Kaiser MF, Johnson DC, Qiang YW, Jones JR, Cairns DA, Gregory WM, Owen RG, Cook G, Drayson MT, Jackson GH, Davies FE and Morgan GJ: APOBEC family mutational signatures are associated with poor prognosis translocations in multiple myeloma. Nat Commun *6*: 6997, 2015. PMID: 25904160. DOI: 10.1038/ncomms7997

19 Mehnert JM, Panda A, Zhong H, Hirshfield K, Damare S, Lane K, Sokol L, Stein MN, Rodriguez-Rodriquez L, Kaufman HL, Ali S, Ross JS, Pavlick DC, Bhanot G, White EP, DiPaola RS, Lovell A, Cheng J and Ganesan S: Immune activation and response to pembrolizumab in POLE-mutant endometrial cancer. J Clin Invest *126(6)*: 2334-2340, 2016. PMID: 27159395. DOI: 10.1172/JCI84940

20 Howitt BE, Shukla SA, Sholl LM, Ritterhouse LL, Watkins JC, Rodig S, Stover E, Strickland KC, D'Andrea AD, Wu CJ, Matulonis UA and Konstantinopoulos PA: Association of polymerase e-mutated and microsatellite-instable endometrial cancers with neoantigen load, number of tumor-infiltrating lymphocytes, and expression of PD-1 and PD-L1. JAMA Oncol *1(9)*: 1319-1323, 2015. PMID: 26181000. DOI: 10.1001/jamaoncol.2015.2151

21 León-Castillo A, Britton H, McConechy MK, McAlpine JN, Nout R, Kommoss S, Brucker SY, Carlson JW, Epstein E, Rau TT, Bosse T, Church DN and Gilks CB: Interpretation of somatic POLE mutations in endometrial carcinoma. J Pathol *250(3)*: 323-335, 2020. PMID: 31829442. DOI: 10.1002/path.5372

22 Lin Y, Luo Y, Sun Y, Guo W, Zhao X, Xi Y, Ma Y, Shao M, Tan W, Gao G, Wu C and Lin D: Genomic and transcriptomic alterations associated with drug vulnerabilities and prognosis in adenocarcinoma at the gastroesophageal junction. Nat Commun *11(1)*: 6091, 2020. PMID: 33257699. DOI: 10.1038/s41467-020-19949-6

23 Nones K, Waddell N, Wayte N, Patch AM, Bailey P, Newell F, Holmes O, Fink JL, Quinn MCJ, Tang YH, Lampe G, Quek K, Loffler KA, Manning S, Idrisoglu S, Miller D, Xu Q, Waddell N, Wilson PJ, Bruxner TJC, Christ AN, Harliwong I, Nourse C, Nourbakhsh E, Anderson M, Kazakoff S, Leonard C, Wood S, Simpson PT, Reid LE, Krause L, Hussey DJ, Watson DI, Lord RV, Nancarrow D, Phillips WA, Gotley D, Smithers BM, Whiteman DC, Hayward NK, Campbell PJ, Pearson JV, Grimmond SM and Barbour AP: Genomic catastrophes frequently arise in esophageal adenocarcinoma and drive tumorigenesis. Nat Commun *5*: 5224, 2014. PMID: 25351503. DOI: 10.1038/ncomms6224

24 Alexandrov LB, Jones PH, Wedge DC, Sale JE, Campbell PJ, Nik-Zainal S and Stratton MR: Clock-like mutational processes in human somatic cells. Nat Genet *47(12)*: 1402-1407, 2015. PMID: 26551669. DOI: 10.1038/ng.3441

25 Joharatnam-Hogan N, Shiu KK and Khan K: Challenges in the treatment of gastric cancer in the older patient. Cancer Treat Rev *85*: 101980, 2020. PMID: 32065879. DOI: 10.1016/j.ctrv.2020.101980

26 Moody S, Senkin S, Islam SMA, Wang J, Nasrollahzadeh D, Cortez Cardoso Penha R, Fitzgerald S, Bergstrom EN, Atkins J, He Y, Khandekar A, Smith-Byrne K, Carreira C, Gaborieau V, Latimer C, Thomas E, Abnizova I, Bucciarelli PE, Jones D, Teague JW, Abedi-Ardekani B, Serra S, Scoazec JY, Saffar H, Azmoudeh-Ardalan F, Sotoudeh M, Nikmanesh A, Poustchi H, Niavarani A, Gharavi S, Eden M, Richman P, Campos LS, Fitzgerald RC, Ribeiro LF, Soares-Lima SC, Dzamalala C, Mmbaga BT, Shibata T, Menya D, Goldstein AM, Hu N, Malekzadeh R, Fazel A, McCormack V, McKay J, Perdomo S, Scelo G, Chanudet E, Humphreys L, Alexandrov LB, Brennan P and Stratton MR: Mutational signatures in esophageal squamous cell carcinoma from eight countries with varying incidence. Nat Genet *53(11)*: 1553-1563, 2021. PMID: 34663923. DOI: 10.1038/s41588-021-00928-6

Received April 21, 2022
Revised May 11, 2022
Accepted June 16, 2022