Mutational Signatures Associate With Survival in Gastrointestinal Carcinomas

PEETER KARIHTALA¹, KATJA PORVARI² and OUTI KILPIVAARA^{3,4,5}

¹Department of Oncology, Helsinki University Hospital Comprehensive Cancer Center and University of Helsinki, Helsinki, Finland; ²Department of Pathology, Medical Research Center Oulu, Oulu University Hospital and University of Oulu, Oulu, Finland;

³Applied Tumor Genomics Research Program, Faculty of Medicine, University of Helsinki, Helsinki, Finland;

⁴Department of Medical and Clinical Genetics, Medicum,

Faculty of Medicine, University of Helsinki, Helsinki, Finland;

⁵HUSLAB 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 agerelated 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

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.



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC-ND) 4.0 international license (https://creativecommons.org/licenses/by-nc-nd/4.0). 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.

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., 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, wholegenome 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 cancerspecific 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=2.9-1.1x10⁷).

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 cancerspecific 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

		eligible for	variables/	rejected proportional	<i>p</i> -value <0.05 from
		the analysis	variable levels	hazards assumptions	the CoxPH test
Esophageal	OS	SBS1, SBS2, SBS3, SBS5, SBS13, SBS16, SBS17a, SBS17b, SBS18	Tissue "Upper third of esophagus" and SRS16 excluded	Age, primary diagnosis	SBS17b
	DSS	SBS1, SBS2, SBS3, SBS5, SBS13,	Squamous carcinoma, keratinizing,	Age	
		SBS16, SBS17a, SBS17b, SBS18	NOS included in Other category, Tissue or organ of origin excluded, SBS16 excluded		
	DFI	SBS1, SBS2, SBS3, SBS5, SBS13,	Stage not included in the model, Primary	Primary diagnosis,	SBS18
		SBS16, SBS17a, SBS17b, SBS18	diagnosis categories Squamous cell	SBS3 and SBS5	
			carcinoma, keratinizing, NOS and Other excluded, signature SBS17b excluded		
	PFI		Squamous cell carcinoma, keratinizing,	SBS5	
Ctomoch	00	SDS1 SDS2 SDS13 SDS15	There are around a contraction of the category		
adeno-	20	SBS17a, SBS17b, SBS20, SBS40, SBS54	of stomach, NOS", "Pylorus" excluded		
carcinoma	DSS	SBS1, SBS2, SBS5, SBS13, SBS15,	Tissue or organ categories "Lesser curvature	SBS40	
		SBS17a, SBS17b, SBS20, SBS40, SBS54	of stomach, NOS", "Pylorus" excluded		
	DFI	SBS1, SBS2, SBS5, SBS13, SBS15,	Tissue or organ categories	SBS20	SBS1
		SBS17a, SBS17b, SBS20, SBS40, SBS54	"Lesser curvature of stomach,		
			"Domition: "Pylorus" excluded, Primary diagnosis		
			raphilary agenocarchionna, NOS samples and stage IV samples excluded		
	PFI	SBS1, SBS2, SBS5, SBS13, SBS15,	Tissue or organ categories "Lesser curvature	SBS13	
		SBS17a, SBS17b, SBS20, SBS40, SBS54	of stomach, NOS", "Pylorus" excluded		
Cholangio-	SO	SBS1, SBS2, SBS5, SBS13, SBS40	Tissue or organ categories other than	SBS5	
carcinoma			"Intrahepatic bile duct" were combined		
			into "Other" category, stage III samples		
			excluded, Primary diagnosis not included		
			in the model as only one category		
	DSS	SBS1, SBS2, SBS5, SBS13, SBS40	Tissue or organ categories other than		
			Intranepatic one duct were compilied 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	Age, SBS2	
	DEI		diagnosis not included in the model		
Pancreatic	SO	SBS1, SBS5, SBS10b, SBS15	Primary diagnosis categories	Stage	SBS10b
adeno-carcinoma			"Adenocarcinoma with mixed subtypes",)	
			"Carcinoma, undifferentiated, NOS",		
			"Neuroendocrine carcinoma, NOS" and		
			"Mucinous adenocarcinoma" were		
	220		compined into Other category		SD C 1 UP
	202	3D31, 3D32, 3D3100, 3D31	Dalite as above		001000

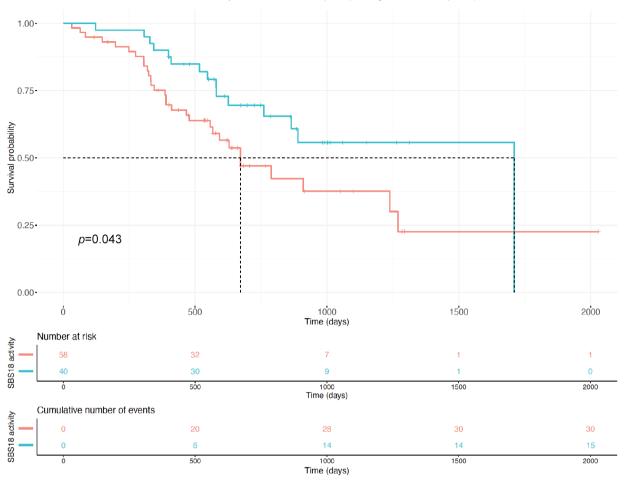
Table I. Summary of the Cox proportional hazard (PH) analyses from the TCGA data.

Table I. Continued

Calicel type I	Ellu-pullt	orgnatures	EXCINCE	variables with	orgliatures with
		eligible for	variables/	rejected proportional	<i>p</i> -value < 0.05 from
		the analysis	variable levels	hazards assumptions	the CoxPH test
	DFI	SBS1, SBS5, SBS10b, SBS15	Same as above	Sex	
Hepatocellular	DS OS	SBS1, SBS9, SBS90, SBS100, SBS10 SBS1, SBS5, SBS9, SBS22, SBS30, SBS40	Same as above		
carcinoma	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		SBS5, SBS17b, SBS40
			"Other" category as only few samples, Tissue "other" category as only few samples, Tissue		
			"Rectosigmoid junction" were combined		
			into "Other" category		
	250	3B31, 3B32, 3B33, 3B317b, 3B340	Stage excluded as 5 categories had inflated coefficients		
	DFI	SBS1, SBS2, SBS5, SBS17b, SBS40	Primary diagnosis and Tissue	Age, Sex	
			or organ of origin excluded		
	PFI	SBS1, SBS2, SBS5, SBS1/b, SBS40	Primary diagnosis categories other than		
			"Adenocarcinoma, NOS" were combined into "Other"		
			category as only few samples, Tissue categories other than "Rectum. NOS" and "Rectosismoid iunction"		
			were combined into "Other" category		
Colon adeno-	OS	SBS1, SBS2, SBS5, SBS15, SBS40, SBS44	Tissue or organ of origin excluded from the model	SBS5	SBS2
carcinoma	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-	SO	SBS1, SBS2, SBS5, SBS15, SBS40, SBS44	Tissue Hepatic flexure of colon excluded		
carcinoma,	DSS	SBS1, SBS2, SBS5, SBS15, SBS40, SBS44	Tissue Hepatic flexure of colon excluded		
right-side	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	SO	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	
			stage I excluded, SBS15 and SBS44 excluded		
	DFI	SBS1, SBS2, SBS5, SBS15, SBS40, SBS44	Primary diagnosis and Tissue or organ of origin not included in the model, stage IV excluded,	Sex, SBS5	
	PFI	SBS1. SBS2. SBS5. SBS15. SBS40. SBS44	SBS2, SBS15 and SBS44 excluded SBS15 excluded Tissue or	Tissue or organ of origin. SBS5. SBS40	BS40

Apple <th></th> <th></th> <th></th> <th></th> <th></th> <th></th>						
Reprint (Participants)Note (Part	10	(N=180)	1.01			0.384
	ICC pathologic stage		reference		i.	
Main		II (N=76)	(0.478 - 4.8)			0.477
image departsimage						
$ \left $	duran damasta	(N=0) Adenocarcinoma, NOS				0.014 *
Bigsmann amount weight wei	rimary diagnosis				-	0.599
Bigs of all all all all all all all all all al		Squamous cell carcinoma, keratinizing	(0.328 - 6.9) (0.460 - 14.3)	-	-	
Name of part of the set of t		Squamous cell carcinoma, NOS (N=85)				0.592
Mathematerial Mathematerial Mathematerial Mathematerial Mathematerial Mathematerial Mathematerial Mathematerial Mathematerial Mathematerial Mathematerial Mathematerial Mathematerial Mathematerial Mathematerial Mathematerial 	issue or organ of origin	Esophagus, NOS (N=10)			•	
		Lower third of ecophagus (N=122) Middle third of ecophagus	(0.068 - 2.4)		-	
imageNoteN			(0.107 - 3.4) 0.75			
Mage Base Base Base Base Base Base Base Base Base Base Base Base 	Bender					0.77
Bai Solution Solution<						0.36
SetNo </td <td>BBS1</td> <td></td> <td>reference</td> <td></td> <td>ė.</td> <td></td>	BBS1		reference		ė.	
						0.108
BandNon-open state<	BS2					
NoteN	2009					0.356
Bad And And </td <td></td> <td></td> <td></td> <td></td> <td></td> <td>0.302</td>						0.302
	BS5					
Bang			(0.713 - 2.5)			0.359
	BS13	Low (N=91)	reference		•	
Max Max <td></td> <td></td> <td></td> <td></td> <td></td> <td>0.884</td>						0.884
	BS17a			-		0.07
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	BS17b					0.07
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	BS18		reference			
Marry stagnosis Appropriations, NOS stature attenue $a_{0.25}^{0.25}$ ditenue $a_{0.25}^{0.25}$						
δρωσου oct acromm, NOS ρ. 49-40.5 μ μ <thμ< th=""> μ <thμ< th=""> <</thμ<></thμ<>			1.00			
barrow Reference Reference <threference< th=""> <threference< th=""></threference<></threference<>						0.6
Note:	rımary diagnosis	Adenocarcinoma, NOS (N=87)	reference			
BB31 Import		Adenocarcinoma, NOS (N=87) Squamous cell carcinoma, NOS (N=90)	reference 5.49 (0.498 - 60.54)	-	•	a.
المال المال المال المال المال المال المال المال المال المال المال المال المال <t< td=""><td></td><td>Adenocarcinoma, NOS (N=87) Squamous cell carcinoma, NOS (N=90)</td><td>reference 5.49 (0.498 - 60.54)</td><td>-</td><td></td><td></td></t<>		Adenocarcinoma, NOS (N=87) Squamous cell carcinoma, NOS (N=90)	reference 5.49 (0.498 - 60.54)	-		
1000 10000 1000		Adenacarcinoma, NOS (N=87) Sguagnaus cell carcinoma, NOS (N=80) (N=25)	reference (0.498 – 60.54) reference	-		
B822	Gender	Aderocarcinoma, NOS (N=87) Spuamous cell carcinoma, NOS (N=80) (N=26) (N=25) (N=152)	reference (0.488 - 60.54) reference (1.021 - 74.43)			
High (N=67) (0, 0, 0, 2, 00) 3880 (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0	Gender	Adenocarcinoma, NOS (N=47) Souamous cell carcinoma, NOS (N=40) Innale (N=752) Low (N=88)	reference (0.496 - 80.54) reference (1.021 - 74.43) reference			
BB33 Image: Series (Mm = Single) reference Image: Series (Mm = Single)	Gender 8861	Adenocarchioma, NOS (N=07) Spaamous cell carcinoma, NOS (N=25) main (N=152) Lov (N=86) High (N=86)	reterence (0.49 = -0.54) reterence (1.021 - 74.43) reterence (0.30 = 26.56)			• 0.1
الأولى المؤالى المؤلى المؤالى المؤلى ال المؤالى المؤالى المؤلى المؤلى المؤالى المؤالى المؤالى المؤلى المؤلى المؤالى المؤالى المؤالى المؤالى المؤالى المؤالى المؤلى المؤلى المؤلى المؤلى المؤلى ال	Gender 8861	Aderocarchoma, NOS (N-87) N-80 (N-82) (N-82) (N-82) (N-89) (N-89) (N-89) (N-89) (N-87) (N-87)	rotorence (0.492 + 0.54) roterence (1.021 + 74.43) roterence (0.302 + 0.56) roterence			• • • • • • • • • • • • • • • • • • •
BBS5 I/W ass roterence I/W ass 0.67 I/W ass 0.224 - 1.77 I/W ass 0.227 - 11.57 I/W ass 0.237 - 11.57 I/W ass 0.237 - 11.57 I/W ass 0.137 - 4.27 I/W ass 0.038 - 0.01 I/W ass 0.038 - 0.01 I/W ass 0.038 - 0.01	Gender 8861	Adenpearchoma, NOS (N=87) Spuamous ceil carcinoma, NOS (maile (N=28) maile (N=82) (N=80) (N=80) (N=87) High (N=87)	rotorence (0.492 + 0.54) roterence (1.021 + 74.43) roterence (0.302 + 0.56) roterence			• 0.1
High (h-88) (0.254 - 1.77) 88813 (0.007 - 1.757) (h-67) reference (h-67) (0.537 - 1.757)	Dender 9981 9952	Adenpearchoma, NOS (N=87) Spuamous ceil carcinoma, NOS (maile (N=28) maile (N=82) (N=80) (N=80) (N=87) High (N=87)	reference (0.465 44 60.457 480.54) reference (1.021 - 74.43) reference (0.302 - 66.56) reference (0.102 - 6.56) reference			• • • • • • • • • • • • • • • • • • •
BBB13 Low (N=07) reference 100 (N=07) (0.527-17.57) 100 (N=100) (0.527-17.57) 100 (N	Dender 9981 9952	Aderocarchoma, NOS (N=87) Spaamous cell carchoma, NOS fernale (N=28) (N=89) L(N=89) High (N=89) L(N=89) High (N=87) L(N=87) L(N=87) L(N=87)	reterence (0.49 = -0.54) reterence (1.221 - 74.43) reterence (0.302 - 76.56) reterence (0.169 - 3.00) reterence			• • • • • • • • • • • • • • • • • • •
BBB13 Low (N=07) reference 100 (N=07) (0.527-17.57) 100 (N=100) (0.527-17.57) 100 (N	Dender 3993 3993	Aderocarchoma, NOS (N-87) SQLamous cell carchoma, NOS (N-80) (N-80) (N-80) (N-80) (N-80) (N-80) (N-80) (N-87) (N-87) (N-87) (N-87) (N-87) (N-87)	$(0.40^{5} - 80.54)$ $(0.40^{5} - 80.54)$ $(1.20^{7} - 74.43)$ $reterence$ $(0.30^{5} - 8.65)$ $reterence$ $(0.160^{-7} - 3.69)$ $reterence$ $(0.20^{5} - 3.65)$			• 0.1 • 0.1
High (N=87) (0.527-17.57) Image: Constraint of the second	Dender 3993 3993	Advergezarchionna, NOS Spaanous cell carchionna, NOS (N=87) Immaio (N=72) Immaio (N=72) Immaio	roterence (0.40 - 40.54) reterence (1.02 - 74.43) reterence (0.30 - 76.56) reterence (0.169 - 3.06) reterence (0.060 - 1.66) reterence			• 0.1 • 0.1
SB817a L ^{WW} _(N=118) reference Image: Comparison of the compari	Dender 8891 8892 8893 8895	Aderocarchoma, NOS Spaamous cell carchoma, NOS Spaamous cell carchoma, NOS Immage Immagee	rolsrence $(0.496 + 0.54)$ reference $(1.221 - 74.43)$ reference $(0.302 + 16.56)$ reference $(0.107 + 3.06)$ reference $(0.202 + 16.56)$ reference $(0.202 + 16.56)$ reference $(0.202 + 16.56)$ reference $(0.202 + 1.57)$ reference $(0.202 + 1.57)$			
High (N=56) 0.75 (0.131 - 4.27)	Dender 8891 8892 8893 8895	Aderocarchoma, NOS Spaanous cell carchoma, NOS Spaanous cell carchoma, NOS Immain	rotorence (0.40 ⁶ = 60.54) rotorence (1.02 ⁸ - 74.43) rotorence (0.30 ² - 46.56) rotorence (0.10 ⁹ - 2.00) rotorence (0.00 ⁸ - 1.68) rotorence (0.00 ⁸ - 1.68) rotorence			
88818 L ^W _(N=128) roterance Here) (0.005 ^{-0.069}) → → → → → → → → → → → → → → → → → → →	Dender 9891 9892 9893 9893 9893 9893 9893 9893	Aftergocarcinoma, NOS Splanous cell carcinoma, NOS Image Image <tdimage< td=""> Image</tdimage<>	rotorence (0.40 ² - 80.54) reterence (1.02 ¹ - 74.43) reterence (0.30 ² - 16.56) reterence (0.160 ² - 16.56) reterence (0.050 ³ - 1.69) reterence (0.254 ⁴ - 1.77) reterence (0.25 ³ - 17.57)			
нар 0.18 (N-н6) (0.058 - 0.56)	Dender 9891 9892 9893 9893 9893 9893 9893 9893	Afergocarichome, NOS Spituarious cell carchoma, NOS Immaile	reference (0.48 ² - 80.54) reference (1.02 ¹ - 74.43) reference (0.30 ² - 86.56) reference (0.16 ² - 8.56) reference (0.05 ³ - 1.68) reference (0.05 ³ - 1.68) reference (0.25 ⁴ - 1.77) reference (0.52 ³ - 17.57)			
	Dender De01 De02 De02 De02 De02 De02 De02 De01 De02 De02 De01 De02 De02 De01 De02 De02 De02	Aftergocarchoma, NOS Aftergocarchoma, NOS Splaaphous cell carchoma, NOS Immail	товятелов (0.40 ⁶ ⁴ 0.54) тяветелов (1.02 ⁸ ⁷² .4.43) товятелов (0.30 ⁸ ²⁶ .6.69) товятелов (0.10 ⁹ ⁷³ .00) тяветелов (0.00 ⁹ ²¹ .68) товятелов (0.00 ⁹ ²¹ .68) товятелов (0.20 ⁹ ²¹ .77) товятелов (0.20 ⁹ ⁴⁷ .77) товятелов			
	Dender De01 De02 De02 De02 De02 De02 De02 De01 De02 De02 De01 De02 De02 De01 De02 De02 De02	Afterspectroment, NOS Splaamous cell carcinoma, NOS Immail Immail<	roterence (0.40 - 40.54) roterence (1.20 - 7.4.43) roterence (0.20 - 7.4.43) roterence (0.20 - 7.4.43) roterence (0.20 - 7.4.43) roterence (0.20 - 7.77) roterence (0.20 - 7.77) roterence (0.20 - 7.77) roterence (0.20 - 7.77) roterence			

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.



SBS18 activity + lower than median (n = 58) + higher than median (n = 40)

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 adenocarcinomaspecific 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

10	(N=363)	(0.984 - 1.1)		-	0.315
ge JCC pathologic stage	(N=363) I (N=55)	reference			0.315
	II (N=125)	$\begin{array}{r} 1.55\\ (0.542 - 4.4)\\ 1.66\\ (0.574 - 4.8)\end{array}$			0.413 0.351
rimary diagnosis	III (N=183) Adenocarcinoma, NOS (N=137)	(0.574 - 4.8) reference			0.351
	Adenocarcinoma, intestinal type	0.37			0.061
	(N=5) Carcinoma, diffuse type (N=59)	$\begin{array}{c} (0.128 - 1.0) \\ 0.49 \\ (0.180 - 1.3) \\ 0.70 \\ (0.268 - 1.8) \\ 0.13 \\ (0.016 - 1.0) \\ 0.51 \end{array}$			0.155 0.476
	(N=59) Mucinous adenocarcinoma (N=20)	(0.268 - 1.8) 0.13 (0.016 - 1.0)			0.054
	Signet ring cell carcinoma	(0.064 - 4.1)		-	0.528
issue or organ of origin	(N=10) Body of stomach (N=85) Carclia, NOS (N=87)	reference 1.91 (0.630 - 5.8)			0.253
	(N=87) Fundus of stomach (N=49) Gastric antrum (N=130)	(0.630 - 5.8) 1.97 (0.608 - 6.3)			0.259
	Gastric antrum (N=130)	(0.630 - 5.8) 1.97 (0.608 - 6.3) 2.56 (0.999 - 6.6) 18.35 (2.563 - 131.3)			0.05
lender	(N=130) Stomach, NOS (N=12) female (N=128)	(2.563 – 131.3) reference			0.004 **
ondor	male (N=235)	2.45 (1.124 - 5.4)			0.024 *
BS1	Low (N=168)	reference			
BS2	High (N=165) Low (N=244)	2.55 (1.132 - 5.8) reference			0.024 *
	(N=244) High (N=89)	1.67 (0.600 - 4.6)			0.328
BS5	Low (N=173)	reference		•	
B\$13	High (N=160) Low	0.57 (0.178 - 1.8) reference			0.343
	Low (N=261) High (N=72)	0.84 (0.241 - 3.0)			0.791
B\$15	Low (N=280)	reference			
BS17a	High (N=53) Low (N=216)	0.83 (0.269 – 2.5) reference			0.741
worrd	(N=216) High (N=117)	1.09 (0.267 - 4.4)			0.908
B\$17b	Low (N=198)	reference			
B\$20	High (N=135) Low	1.01 (0.254 - 4.0) reference	-		0.993
8520	Low (N=287) High (N=46)	0.60 (0.173 - 2.1)			0.415
BS40	Low (N=189)	reference			
	High (N=144)	0.38 (0.132 - 1.1)			0.075
BS54	Low (N=314) High (N=19)	0.16 (0.016 - 1.7)			0.13
/ Events: 43; Global p-value (Log-Rank): UC: 424.87; Concordance Index: 0.74		(0.016 - 1.7) 0.01	0.1 Hazard ratio		10 100
IC: 424.87; Concordance Index: 0.74	0.085068	0.01			10 100
IC: 424.87; Concordance Index: 0.74	(N=439)				
IC: 424.87; Concordance Index: 0.74	(N=439) (N=59) (N=59) (N=29)	0.01 (1.01 – 1.06) reference 2.37 (0.85 – 6.64)			10 100
IC: 424.87; Concordance Index: 0.74	(N=439) I (N=59) I N=129) III (N=183)	0.01 (1.01 - 1.06) reference (2.37 (0.85 - 6.64) (1.48 - 5.50)			10 100 a.core a.t a.core
IIC: 444.87; Concordance Index: 0.74 Ige IJCC pathologie stage	(N=439) I (N=59) I N=129) III (N=183)	0.07 (1.01 – 1.06) (1.06 – 1.07) (1.05 – 6.64) 3.95 (1.06 – 10.52) (2.45 – 22.06) (2.45 – 22.06)			10 100
IIC: 444.87; Concordance Index: 0.74 Ige IJCC pathologie stage	(N=430) (N=50) (N=50) (N=120) (N=120) (N=43) (N=43) (N=43) (N=43) (N=43)	0.01 (1.0.1 ± 1.00) reference (2.85 ± 6.64) (1.43 ± 9.5.59) (2.45 ± 9.2.06) reference (0.42 ± 7.78)			10 100 0.002 0.1 0.006 0.0000 0.00000 0.00000 0
IIC: 444.87; Concordance Index: 0.74 Ige IJCC pathologie stage	(N=439) (N=439) (N=58) (N=129) (N=1	(1.01 = 1.08) (1.01 = 1.08) (1.04 = 1.08) (1.04 = 1.08) (2.43 = 22.08) (2.43 = 22			10 100
IIC: 444.87; Concordance Index: 0.74 Ige IJCC pathologie stage	(N=439) (N=439) (N=59) (N=129) (N=129) (N=129) (N=49) ((1.01 = 1.08) (1.01 = 1.08) (1.04 = 1.08) (1.04 = 1.08) (2.43 = 22.08) (2.43 = 22			10 100 0.002 0.1 0.006 0.0000 0.00000 0.00000 0
IIC: 444.87; Concordance Index: 0.74 Ige IJCC pathologie stage	(N=439) (N=439) (N=58) (N=120) (N=120) (N=120) (N=120) (N=163) N=100 (N=164) Adenovarianoma, infestinal hype Tubular adanovarianoma (N=70) ma, difuse hype N_427 ross adanovarianoma M_277 ross adanovarianoma Sapart (mo cell carronna Sapart (mo cell carronna	0.07 (1,01 - 1.66) (160 - 1.67) (164 - 0.05) (164 - 0.05)	Hazard ratio		10 100
IC: 444.87; Concordance Index: 0.74 Age MGC pathologie stage Yimary diagnosis	(N=439) (N=439) (N=58) (N=120) (N=120) (N=120) (N=120) (N=163) N=100 (N=164) Adenovarianoma, infestinal hype Tubular adanovarianoma (N=70) ma, difuse hype N_427 ross adanovarianoma M_277 ross adanovarianoma Sapart (mo cell carronna Sapart (mo cell carronna	(1.01 - 10.68) treference (2.02 - 20.00) (1.04 - 10.68) (1.04 - 10.69) (2.45 - 20.69) (2.45 - 20.69)	Hazard ratio		10 100
IC: 444.87; Concordance Index: 0.74 Age MGC pathologie stage Yimary diagnosis	(N=439) (N=439) (N=58) (N=120) (N=120) (N=120) (N=120) (N=163) N=100 (N=164) Adenovarianoma, infestinal hype Tubular adanovarianoma (N=70) ma, difuse hype N_427 ross adanovarianoma M_277 ross adanovarianoma Sapart (mo cell carronna Sapart (mo cell carronna	0.07 (1,0,1 = 1,66) (1,64 = 1,66) (1,63 = 6,63) (1,63 = 6,63) (2,43 = 2,66) (2,43 = 2,66) (2,43 = 2,66) (2,43 = 2,66) (2,43 = 2,66) (2,43 = 2,66) (2,53 = 2,66) (2,53 = 2,67) (2,53 = 5,61) (2,54 = 5,61)(2,54 = 5,61) (2,54 = 5,	Hazard ratio		10 100 0.002 0.002 0.0000 0.00000 0.00000 0.0000 0.000
IC: 444.87; Concordance Index: 0.74 Age MGC pathologie stage Yimary diagnosis	(N-439) (N-439) (N-439) (N-439) (N-129) (N-129) (N-164) (N-	$(1,2) \stackrel{1=4}{=} 1,96)$ reference (2,3) $\frac{7}{3}, \frac{3}{2}, \frac{5}{2}, \frac{6}{2}, \frac{6}{$	Hazard ratio		10 100 0.002 0.002 0.002 0.002 0.002 0.0000 0.00000 0.0000 0.0000 0.0000 0.00000 0.00000 0.0000 0.000
IC: 444.87; Concordance Index: 0.74 Age MGC pathologie stage Yimary diagnosis	(N-439) (N-439) (N-439) (N-439) (N-129) (N-129) (N-164) (N-	$(1,2) \stackrel{1=4}{=} 1,96)$ reference (2,3) $\frac{7}{3}, \frac{3}{2}, \frac{5}{2}, \frac{6}{2}, \frac{6}{$	Hazard ratio		10 100 0.002 0.002 0.0000 0.00000 0.00000 0.0000 0.000
IC: 494.87; Concordance Index: 0.74	(N=439) (N=439) (n=59) (n=59) (n=63) (n=163) (n=164	$(1,2) \stackrel{1=4}{=} 1,96)$ reference (1,2) \stackrel{2=5}{=} 6,87) (1,43) $\stackrel{2=5}{=} 6,87)$ (1,43) $\stackrel{2=5}{=} 2,268)$ reference (2,45) $\stackrel{2=6}{=} 2,268)$ (2,45) $\stackrel{2=6}{=} 2,268$ (2,45) $2=$	Hazard ratio		10 100 10 100 10 100 10 100 10 100 10 100 10 0.002
Ic: 424.87; Concordance Index: 0.74	(N=430) (N=50) (N=50) (N=52) (N=120	$\begin{array}{c} (1,01-1,06)\\ (1,01-1,06)\\ (1,01-1,06)\\ (1,01-1,06)\\ (1,01-1,06)\\ (1,01-1,06)\\ (2,01-1,06)$	Hazard ratio		10 100
Ic: 424.87; Concordance Index: 0.74	0.085099 (N=459) (N=59) (N=59) (N=163) (N=164) (N=1	$\begin{array}{c} 0.07\\ (1.01-1.06)\\ reference\\ (1.01-1.06)\\ reference\\ (1.43-5.02)\\ (2.43-5.02)\\ reference\\ (2.43-5.02)$	Hazard ratio		10 100 10 100 10 100 10 100 10 100 10 100 10 0.002
Ilissue or organ of origin	2.085099 (N=439) (N=59) (N=59) (N=129) (N=1	(1,0) = 1,06) (1,0) = 1,06) (1,0) = 1,06) (1,0) = 1,06) (1,0) = 1,06) (1,0) = 1,06) (2,4) = 2,06) (1,0) = 1,00 (2,4) = 2,06) (1,0) = 1,00 (2,4)	Hazard ratio		10 100 0.002 0
Ite: 494.87; Concordance Index: 0.74	(N=430) (N=430) (N=430) (N=420) (N=120) (N=120) (N=420) (N=	$\begin{array}{c} 0.07\\ (1.0] = 1.06)\\ (reference\\ (1.0) = 6.67\\ (1.43) = 6.6$	Hazard ratio		10 100 0.002 0.002 0.002 0.0000 0.00000 0.00000 0.0000 0.0000 0.00
Ic: 424.87; Concordance Index: 0.74	2.085099 (N=430) (N=50) (N=50) (N=120) (N=1	$(1,2) = \frac{104}{100}, (1,2) =$	Hazard ratio		10 100 0.002 0
Ic: 424.87; Concordance Index: 0.74	0.085099 (N=439) (n=59) (n=59) (n=59) (n=63) (n=	$\begin{array}{c} 0.01\\ (1.01 \\ 1.06\\ (1.05 \\ -6.05\\ 0.$	Hazard ratio		10 100 0.002 0.002 0.002 0.002 0.0000 0.00000 0.0000 0.0000 0.00000 0.00000 0.00000 0.00000 00
Events: 43; Global p-value (Log-Rank): IC: 424.87; Concordance Index: 0.74 Age JuCC pathologic stage Primary diagnosis Resue or organ of origin Jander 3881 3852 3855	2.085099 (N=439) (N=439) (N=43	$(1,0) = 1,00, \\ (1,0) = 1,00, \\ (1,0) = 1,00, \\ (1,0) = 1,00, \\ (1,0) = 1,00, \\ (1,0) = 1,00, \\ (2,1) = 1,00$	Hazard ratio		10 100 10 100 10 100 10 0002 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.000
Lic: 424.87; Concordinge Index: 0.74 Age Age Age Primary diagnosis Tissue or organ of origin lender B851 B852 B853 B851 B852 B853 B8513 B8513 B8515 B8513 B8515 B8515 B8515 B8513 B8515 B851 B8515 B8515 B8515 B8515 B851 B851 B851	20.085099 (N=439) (N=439) (N=59) (N=639) (N=639) (N=639) (N=649) (N	$\begin{array}{c} 0.01\\ (1.01-1.06)\\ reference\\ (2.05-6.64)\\ (1.43-6.05)\\ (2.45-6.64)\\ (1.43-6.05)\\ (2.45-6.64)\\ (1.43-6.05)\\ (2.45-6$	Hazard ratio		10 100 0.002 0.002 0.002 0.002 0.0000 0.00000 0.0000 0.0000 0.00000 0.00000 0.00000 0.000000 0
Lic: 424.87; Concordinge Index: 0.74 Age Age Primary diagnosis Tissue or organ of origin lender B851 B852 B853 B8513 B8513 B514 B514 B514 B514 B514 B514 B515 B513 B515 B515 B51 B51	2.085099 (N=439) (N	$\begin{array}{c} 0.07\\ (1,0,1,0,1,0,0)\\ (1,0,3,5,-2,6,0)\\ (2,3,5,-2,6,0)\\ (2,4,3,-2,6,0)$	Hazard ratio		10 100 10 100 10 100 10 0002 0.00
Age LIGC 494.87; Concordinos Index: 0.74 Age Name Name of origin Sender SBS1 SBS5 SBS5 SBS13 SBS15 SBS17a	20.085099 (N=430) (N=50) (N=50) (N=120) (N=	$\begin{array}{c} 0.01\\ (1,0] = 1.06)\\ (1,0) = 5.67\\ (2,0) = 5.67\\ (3$	Hazard ratio		10 100 0.002 0
Ics: 424.87; Concordinge Index: 0.74 Age UGC pathologie stage Primary diagnosis Issue or organ of origin Issue or organ of origin Bender B851 B852 B853 B853 B853 B853 B853 B853 B853 B853	2.085099 (N=439) (N=439) (N=43) (N	$\begin{array}{c} 0.07\\ (1.0.1^{1.0.4}_{-1.06})\\ (1.0.3^{-5}_{-5.6}-6.4)\\ (2.0.3^{$	Hazard ratio		10 100 10 100 10 100 10 0002 0.00
Ics: 424.87; Concordinge Index: 0.74 Age UGC pathologie stage Primary diagnosis Issue or organ of origin Issue or organ of origin Bender B851 B852 B853 B853 B853 B853 B853 B853 B853 B853	2.085099 (N=439) (N=439) (n=59) (n=59) (n=59) (n=63) (n=63) (n=64) (n=64	$\begin{array}{c} 0.01\\ (1.0] = 1.06\\ (reference\\ (2.05 \pm 6.64)\\ (1.43 \pm 0.63)\\ (2.43 \pm 0.05)\\ (2.43 \pm 0.05$	Hazard ratio		10 100 0.002 0.002 0.002 0.002 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
Age	20.085099 (N=439) ((1,2) = 1,40) references $(2,3) = 4,40$ references $(3,3) = 4,40$ references $(3,4) = 1,40$ references $(4,4) = 1,40$ r	Hazard ratio		10 100 10 100 10 100 10 100 10 0000 10 000000 10 0000 10 000000 10 000000 10 0000 10 0000 10 0000 10 0000 10 0000 1
Ke: 44.87; Concordinos Inde: 0.74 kge kge rimary diagnosis rimary diagnosis lasue or organ of origin lasue or organ of origin lasue of organ of organ of origin lasue of organ of organ of organ lasue of organ of organ of organ of organ lasue of organ of organ of organ of organ lasue of organ of organ of organ of organ of organ lasue of organ of organ of organ of organ of organ of organ lasue of organ of organ of organ of organ	20.085099 (N=439) (N=439) (N=439) (N=459) (0.07 (1,1)=1.06) reference (2,2)=7,44 (1,4)=7,63 (2,3)=7,44 (1,4)=7,63 (2,3)=7,44 (1,4)=7,63 (2,3)=7,45 (2,3)=7,15	Hazard ratio		10 100 0.002 0.002 0.002 0.002 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
Lic: 424.87; Concordinge Index: 0.74 Age Age Primary diagnosis Tissue or organ of origin Tissue or organ of origin Bass1 Bass2 Bass3 Bass13 Bass14 Bass17 Bass17 Bass20 Bass2	2.085099 (N-439) (N-59) (N-	(1,2) = 1,40) references $(2,3) = 4,40$ references $(3,3) = 4,40$ references $(3,4) = 1,40$ references $(4,4) = 1,40$ r	Hazard ratio		10 100 0.002 0

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.

ge	(N=183)	1.03 (1.009 – 1.05)					0.005 **
JCC pathologic stage	(N=21)	(1.009 – 1.05) reference					0.000
ooo paaloogio sage	(N=21) II (N=150)	1.40 (0.584 - 3.34)				_	0.453
	(N=150)	(0.584 - 3.34) 0.73 (0.141 - 3.84)			-		0.714
	(N=4) IV (N=5)	(0.141 – 3.84) 1.94 (0.444 – 8.47)		-			4 0.379
rimary diagnosis	(N=5) Adenocarcinoma, NOS (N=19)	(0.444 – 8.47) reference					4 0.378
Innu y ungnour	(N=19) Infiltrating duct carcinoma, NOS (N=149)	1.74 (0.741 - 4.10)			-		0.203
	(N=149)" Other (N=15)	(0.741 - 4.10) 1.00 (0.242 - 4.17)			-		
	(N=15) Head of pancreas (N=137)	(0.242 – 4.17) reference					0.995
lissue or organ of origin	(N=137) Body of pancreas (N=15)		_				
		0.26 (0.091 - 0.75)	-				0.013 *
	Tail of pancreas (N=14)	0.70 (0.261 - 1.88)	-	-			0.48
	Pancreas, NOS (N=17)	0.89 (0.459 – 1.74)		-			0.737
Gender	female (N=82)	reference		-			
	male (N=101)	0.79 (0.505 – 1.22)		-	-		0.288
SBS1	Low (N=86)	reference		•			
	High (N=86)	0.93 (0.578 – 1.49)					0.759
6B\$5	Low (N=90)	reference		•			
	High (N=82)	(0.625 – 1.67)					0.931
SBS10b	Low (N=155)	reference					
	High (N=17)	0.44 (0.205 – 0.96)					0.038 *
GBS15	Low (N=158)	reference					
	High (N=14)	1.03 (0.402 - 2.61)					0.959
			0.1 0.2 Hazard ratio	0.5 1	2	5	10
						5	
1ge	(N=183)	(1.00 ¹ 02 (1.005 - 1.05)		•	L	5	10 0.045 *
Nge LLICC pathologic stage	l (N=21)	reference				5	0.045 *
	l (N=21) II (N=150)	reference (0.683 - 6.39)			•	-	0.045 * 0.197
	(N=21) (N=150) 	reference (0.883 - 8.39) (0.225 - 8.25)			•		0.045 * 0.197 0.737
LICC pathologic stage	$ _{(N=21)}$ $ _{(N=150)}$ $ _{(N=4)}$ $ _{(N=5)}$	reference (0.863 - 6.39) (0.252 - 8.25) (0.225 - 8.25) (0.872 - 18.37)			•		0.045 * 0.197
	(N-21) (N=150) III (N=4) (N=5) Adenocatrinoma, NOS	reference (0.883 - 0.39) (0.225 - 8.25) (0.275 - 18.37) reference			-		0.045 * 0.197 0.737 4 0.137
LICC pathologic stage	(N=21) (N=150) (N=4) V (N=5) Adenocarcinoma, NOS (N=16) (N=149) duct carcinoma, NOS	reference (0.663 - 0.39) (0.225 - 0.25) (0.672 - 10.37) reference (0.672 - 10.37)			-		0.045 * 0.197 0.737
UCC pathologie atage Yimary diagnosis	(N-21) (N-150) (N-4) (N-5) Adenocarcinoma, NOS (N-149) (N-149) (N-19)	reference (0.68 ³ - 0.39) (0.22 ³ - 0.25) (0.07 ² - 18.37) reference (0.62 ⁶ - 3.57) (0.62 ⁶ - 3.57) (0.62 ⁶ - 3.57)			-		0.045 * 0.197 0.737 4 0.137
LICC pathologic stage	(N-21) (N-150) (N-4) (N-5) (N-5) Metroparationoma, NOS (N-16) (N-16) (N-17) (N-16) (reference (2.65 ²⁰ 6.39) (2.225 ⁻⁰ 6.29) (2.67 ² - 18.37) reference (2.68 ⁴ - 3.57) (2.69 ⁹ - 4.57) (2.69 ⁹ - 4.59) reference			-		0.045 * 0.197 0.737 4 0.137 0.385
UCC pathologie atage Yimary diagnosis	(N-21) (N-150) (N-4) (N-5)	reference (0.05 ²⁰ 0.39) (0.225 ³⁰ 0.39) (0.07 ² - 8.29) (0.07 ² - 18.37) reference (0.05 ⁹⁷ - 4.59) reference (0.07 ⁹⁷ - 4.59)					0.045 * 0.197 0.737 4 0.137 0.385
UCC pathologie atage Yimary diagnosis	(N-21) (N-150) (N-4) (N-4) (N-5) Advancesamona, NOS (N-16) (N-1	$\begin{tabular}{lllllllllllllllllllllllllllllllllll$					0.045 * 0.197 0.737 4 0.137 0.305 0.973
UCC pathologie atage Yimary diagnosis	(N-21) (N-150) (N-4) (N-5)	reference (0.05 ²⁰ 0.39) (0.225 ³⁰ 0.39) (0.07 ² - 8.29) (0.07 ² - 18.37) reference (0.05 ⁹⁷ - 4.59) reference (0.07 ⁹⁷ - 4.59)					0.045 * 0.197 0.737 4 0.137 0.385 0.973
UCC pathologie atage Yimary diagnosis	(N-21) (N-150) (N-4) (N-4) (N-5) Advancesamona, NOS (N-16) (N-1	$\begin{tabular}{lllllllllllllllllllllllllllllllllll$					0.045 * 0.197 0.737 4 0.137 0.385 0.973 0.385
WCC pathologic stage Yimary diagnosis Tissue or organ of origin	(N-21) (N-150) (N-150) (N-5) Adesocarismoma, NOS (N-15) (N-150) (N-1	$\begin{array}{c} \text{reference} \\ (a, a_2^{2,00}, a, sy) \\ (a, 2b^{5,00}, a, sy) \\ (a, 2b^{5,00}, a, sy) \\ (a, 672^{-1} a, s7) \\ \text{reference} \\ (a, 2b^{0,01}, a, sy) \\ (a, 2b^{0,01}, a, $					0.045 * 0.197 0.737 4 0.137 0.385 0.973 0.385
WCC pathologic stage Yimary diagnosis Tissue or organ of origin	(N-21) (N-150) (N-150) (N-1) (N-1) (N-1) (N-1) (N-10)	$\begin{tabular}{lllllllllllllllllllllllllllllllllll$					0.045 * 0.197 0.737 4 0.137 0.385 0.973 0.385 0.929
LICC pathologic stage Primary diagnosis Tissue or organ of origin	N-21) (N-10) (N-10) (N-1)	$\begin{tabular}{lllllllllllllllllllllllllllllllllll$					0.045 * 0.197 0.737 4 0.137 0.385 0.973 0.385 0.929
LICC pathologic stage Primary diagnosis Tissue or organ of origin	N-21) (N-2) (N-10) (N-1) (N-3) Adenceancoma, NOS (N-10)	reference $(0.83^{-20}_{-0.39}, 0.39)$ $(0.22^{-30}_{-0.3}, 0.39)$ $(0.22^{-30}_{-0.3}, 0.39)$ $(0.63^{-2}_{-0.3}, 1.43, 0.39)$ $(0.63^{-9}_{-0.3}, 0.39)$ $(0.23^{-7}_{-0.39}, 0.39)$ $(0.23^{-70}_{-0.2}, 0.39)$ $(0.23^{-70}_{-0.2}, 0.39)$ $(0.23^{-70}_{-0.2}, 0.39)$ $(0.23^{-70}_{-0.2}, 0.39)$ $(0.469^{-7}_{-1.29}, 0.39)$ reference $(0.469^{-7}_{-1.29}, 0.39)$ reference					0.045 * 0.197 0.737 0.375 0.973 0.325 0.973 0.325 0.929
LICC pathologie stage Yimary diagnosis Issue or organ of origin Jender Jender	N-21) (N-10) (N-10) (N-1) (N-2) Adveragenceman, NOS (N-10) (N-	$\begin{tabular}{ c c c c c c c } \hline reference & \hline $(0,83^{-2},9^0,3.9)$ \\ \hline $(0,23^{-2},9^0,2.9)$ \\ \hline $(0,07^{-2},11,3,2.7)$ \\ \hline $(0,07^{-2},11,3,2.7)$ \\ \hline $(0,20^{-9},2.7,2.7)$ \\ \hline $(0,20^{-9},2.7,2.7)$ \\ \hline $(0,20^{-9},2.7,2.7)$ \\ \hline $(0,20^{-7},2.7)$ \\ \hline $(0,20^$					0.045 * 0.197 0.737 4 0.137 0.385 0.973 0.032 * 0.525 0.929
LICC pathologie stage Yimary diagnosis Issue or organ of origin Jender Jender	N-21) (N-150) (N-150) (N-15) (N-5) (N-5) (N-5) (N-5) (N-16)	reference $(a, a_2^{2})^{00} - a, 39$ $(a, 2b_3^{-30} - a, 29)$ $(a, 2b_3^{-30} - a, 29)$ $(a, 2b_3^{-30} - a, 29)$ $(a, 2b_3^{-9} - a, 39)$ $(a, 2b_3^{-9} - a, 39)$ $(a, 2b_3^{-9} - a, 29)$ $(a, 2b_3^{-7} - a, 29)$ $(a, 2b_3^{-7} - a, 29)$ $(a, 2b_3^{-7} - a, 21)$ $(a, 2b_3^{-7} - a, 21)$ $(a, 2b_3^{-7} - a, 21)$ reference $(a, 2b_3^{-7} - 1, 20)$ reference $(a, 2b_3^{-7} - 1, 2b)$ reference $(a, 2b_3^{-7} - 1, 2b)$ reference $(a, 2b_3^{-7} - 1, 2b)$ reference					0.045 * 0.197 0.737 4 0.137 0.365 0.973 0.032 * 0.525 0.929 0.298
WCC pathologic stage Yrimary diagnosis Issue or organ of origin Jender 1851	N-21) (N-150) (N-1	reference $(a, a_2^{23} \stackrel{0.9}{=} a, 39)$ $(a, 22^{53} \stackrel{0.9}{=} a, 39)$ $(a, 22^{53} \stackrel{0.9}{=} a, 37)$ $(a, 27^{53} \stackrel{0.9}{=} a, 37)$ reference $(a, 29^{57} \stackrel{0.9}{=} a, 39)$					0.045 * 0.197 0.737 4 0.137 0.385 0.973 0.032 * 0.525 0.929 0.298 0.738
WCC pathologic stage Yrimary diagnosis Issue or organ of origin Jender 1851	(N-21) (N-150) (N-150) (N-150) (N-150) (N-5) (N-5) (N-5) (N-5) (N-10) (reference $(a, a_2^{23})^0, a, sy)$ $(a, a_2^{23})^0, a, sy)$ $(a, a_2^{23})^0, a, sy)$ $(a, a_2^{23})^-, a, sy)$ reference $(a, a_2^{23})^0, a, sy)$ reference $(a, a_2^{23})^1, 1, sy)$ reference $(a, a_2^{23})^1, 2, ay)$					0.045 * 0.197 0.737 0.305 0.973 0.365 0.929 0.228 0.228 0.738

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.

Sex	male (N=64)	reference							
	female (N=70)	1.59 (0.494 – 5.1)				-			0.438
Age	(N=134)	1.06 (1.012 - 1.1)				i i i i i i i i i i i i i i i i i i i			0.013
Primary diagnosis	Adenocarcinoma, NOS (N=123)	reference							
	Mucinous adenocarcinoma (N=11)	1.70 (0.271 – 10.7)		F		-		-	0.57
Tissue or organ of origin	Sigmoid colon (N=112)	reference				•			
	Descending colon (N=17)	0.79 (0.169 – 3.7)			-	H	-		0.767
	Splenic flexure of colon (N=5)	1.46 (0.124 – 17.1)		-		-			0.764
AJCC pathologic stage	l (N=22)	reference				•			
	 (N=46)	0.64 (0.059 - 6.9)			-				0.714
	III (N=42)	2.81 (0.289 – 27.4)							0.373
	IV (N=22)	4.51 (0.445 – 45.8)			-		-		
SBS1	Low (N=74)	reference				ė.			
	High <i>(N=28)</i>	0.52 (0.138 – 2.0)			-				0.338
SBS2	Low (N=91)	reference				•			
	High (N=11)	4.97 (1.005 – 24.6)					-		0.049
SBS5	Low (N=58)	reference							
	High (N=44)	6.90 (1.039 – 45.8)					-		0.046
SBS15	Low (N=95)	reference				•			
	High (N=7)	0.79 (0.042 – 14.9)	-		-	H			0.878
SBS40	Low (N=56)	reference							
	High (N=46)	9.80 (1.444 – 66.5)							0.02 *
# Events: 20; Global p-value (Log-Rank AIC: 150.55; Concordance Index: 0.77			0.05 0.		0.5		5	10	50 100

Hazard ratio

1

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 welldefined 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

Age	(N=170)	1.110 (1.041 - 1.18)						0.002 **
AJCC pathologie stage	(N=33)	(1.041 – 1.18) reference						0.002
Acco planologio slago	(N=33) II (N=51)	0.391 (0.060 - 2.52)			-			0.323
	(N=51) III (N=52)	(0.060 - 2.52) 1.500 (0.224 - 10.03)				•		0.676
	(N=52) IV (N=25)	(0.224 - 10.03) 1.297 (0.185 - 9.09)			_		_	0.794
Primary diagnosis	(N=25) Adenocarcinoma, NOS (N=140)	(0.185 - 9.09) reference					•	0.784
Finaly diagnosis	(N=140) Other (N=30)	0.772 (0.078 - 7.68)	-				_	0.825
Tissue or organ of origin	(N=30) Rectum, NOS (N=90)	(0.078 - 7.68) reference			-			0.020
rissue or organ of organ	(N=90) Rectosigmoid junction (N=71)	0.269 (0.060 - 1.20)		-				0.086
	(N=71) Other (N=9)	(0.060 - 1.20) 3.105 (0.302 - 31.90)	-	-		-		- 0.34
Gender	(N=9) female (N=78)	(0.302 - 31.90) reference						10.04
	(N=78) male (N=92)	0.881 (0.271 - 2.87)			_			0.833
SBS1	(N=92) Low (N=65)	(0.271 – 2.87) reference			-			0.033
6861	(N=65) High (N=64)	0.584 (0.162 - 2.11)			_			0.412
SBS2	(N=64) Low (N=118)	(0.162 - 2.11) reference			-			0.412
	(N=118) High (N=11)	1.581 (0.206 - 12.14)						0.66
SBS5	(N=11) Low (N=65)	(0.206 - 12.14) reference						0.08
	(N=65) High (N=64)	0.130 (0.030 - 0.56)						0.006 **
SBS17b	(N=64) Low (N=120)	(0.030 - 0.56) reference						0.000
383179	(N=120) High (N=9)	3.468 (0.599 - 20.09)				-		0.165
SBS40	(N=9) Low (N=65)	(0.599 - 20.09) reference						0.765
	(N=65) High (N=64)	0.072 (0.012 - 0.44)			-			0.004 **
# Events: 18: Global p-wake (Lop-Rank): 0.0055656 AIC: 130.39; Concordance Index: 0.88			Hazard ratio	0.1	0.5 1	5	10	50
# Evente: 18: Citobel p-value (Log-Ramk): 0.0005866 AIC: 130.39; Concordance Index: 0.88				0.1	0.5 1	5	10	50
AIC: 130.39; Concordance Index: 0.88	(N=170)			0.1	0.5 1	5	10	50
AIC: 130.39; Concordance Index: 0.88		c			0.5 1	5	10	
AIC: 130.39; Concordance Index: 0.88	(N=170)	(1.130+00 (1.10+00 - 1.50+00)			0.5 1	5	10	
AIC: 130.39; Concordance Index: 0.88	(N=170) Agengcarchoma, NOS (N=140)	с (1.19+00 (1.19+00 – 1.59+00) гейевенсе			0.5 1	5	10	0.008 **
AIC: 130.39; Concordance Index: 0.88 Age Primary diagnosis	(N=170) Ademocaronama, NOS (N=30) (N=30)	(1.19+00) (1.19+00 - 1.59+00) reterence (6.19-02 - 1.39+01)				5	10	0.008 **
AIC: 130.39; Concordance Index: 0.88 Age Primary diagnosis	(N=170) Agenocaronoma, NOS (N=40) Clines (N=30) Pectam, NOS Pectam, NOS Pectam, NOS (N=77)	(1.16+00) - 1.56+00) reference (5.16+20) - 1.56+00) reference (4.56+00) - 1.56+00) reference (4.56+00) - 5.56+00			0.5 1	5	10	0.008 **
AIC: 130.39; Concordance Index: 0.88 Age Primary diagnosis Tilesue or organ of origin	(N=170) Adergoardnoma, NOS (N=30) Recture, NOS Pectosigmoid junction (N=9) (N=9)	(1, 1, 30+00) (1, 1, 10+00 - 1, 50+00) reterence (6, 10-62 - 1, 50+00) reterence (4, 30-03 - 6, 20-01) (1, 20+00 - 2, 30+00)				5		0.008 ** 0.02 *
AIC: 130.39; Concordance Index: 0.88 Age Primary diagnosis	(N=170) Adorocaticitoma, NOS (N=76) Pacture, NOS Pacture, NOS Receiptignoid junction (N=9) Chere (N=9) brinale (N=7)	۲ (1.10+00 - 1.50+00) (1.10+00 - 1.50+00) (1.10+00 - 1.50+00) (1.20+00 - 2.30+00) (1.20+00 - 2.30+00) (1.20+00 - 2.30+00) (1.20+00 - 2.30+00) (1.20+00 - 2.30+00)				5	10	0.008 ** 0.024 0.021 *
AIC: 130.39; Concordance Index: 0.88 Age Primary diagnosis Tilesue or organ of origin	(N=170) Agencearchoma, NOS (N=50) Citient Peecham, NOS Peecham, NOS Peecham, NOS (N=60) Citient (N=7)	(1, 1, 30+00) (1, 1, 10+00 - 1, 50+00) reterence (6, 10-62 - 1, 50+00) reterence (4, 30-03 - 6, 20-01) (1, 20+00 - 2, 30+00)				-		0.008 ** 0.02 *
AIC: 130.39; Concordance Index: 0.88 Age Primary diagnosis Tilesue or organ of origin	(N=170) Adorocaticitoma, NOS (N=76) Pacture, NOS Pacture, NOS Receiptignoid junction (N=9) Chere (N=9) brinale (N=7)	۲ (1.10+00 - 1.50+00) (1.10+00 - 1.50+00) (1.10+00 - 1.50+00) (1.20+00 - 2.50+00) (1.20+00 - 2.50+00) (1.20+00 - 2.50+00) (1.20+00 - 2.50+00) (1.20+00 - 2.50+00) (1.20+00 - 2.50+00)				-		0.008 ** 0.024 0.021 *
Alc: 130.39; Concordance Index: 0.88 Age Primary diagnosis Tissue or organ of origin Gender	(N=170) Agencearchoma, NOS (N=50) Citient Peecham, NOS Peecham, NOS Peecham, NOS (N=60) Citient (N=7)	(1,1,0,0,0) rotarence (4,3,0,2,0,2,0,0) rotarence (4,3,0,2,0,2,0,0,0) (1,0,0,0,0,2,3,0,0,0) roterence (1,3,0,0,0,0,0,0) roterence (1,3,0,0,0,0,0,0,0)				-		0.008 ** 0.024 0.021 *
Alc: 130.39; Concordance Index: 0.88 Age Primary diagnosis Tissue or organ of origin Gender	(N=170) Adenocarionoma, NOS (N=30) (N=30) Rectosigmoid junction (N=9) Rectosigmoid junction (N=9) (N=9) Rectosigmoid junction (N=9) Rectosigmoid junction (N=9) Rectosigmo	۲ (1, 1-9-00 (1, 1-9-00 1-9-00 (1, 1-9-00 (1, 1-				-		0.008 ** 0.024 0.02 * 0.021 *
AIC: 130.39; Concordance Index: 0.88 Age Primary diagnosis Tissue or organ of origin Gender BB01	(N=170) (N=	۲ (1,1)-0,-0 (1				-		0.000 ** 0.024 0.027 * 0.027 * 0.052
Alc: 130.39; Concordance Index: 0.88 Age Primary diagnosis Tissue or organ of origin Gender B881 S882	(N=170) Adengalarianama, NOS (Adengalarianama, NOS (Adengalarianama, NOS (Adengalarianama) (Adengalaria	۲ (1, 1-0-0) (1, 1-0-0) (1				-		0.008 ** 0.024 0.02 * 0.021 *
AIC: 130.39; Concordance Index: 0.88 Age Primary diagnosis Tissue or organ of origin Gender BB01	(N=170) (N=170) (N=170) (N=10) (N=	۲ (1.1 - 1.3 - 0.0 - 1.5 - 0.0 - 0				-		0.000 ** 0.024 0.02* 0.021* 0.021* 0.052 0.0542
AIC: 130.39; Concordance Index: 0.88 Age Primary disgnosis Tissue or organ of origin Gender B801 B802 B805	(N=170) (N=	۲ (1,1),0),0),0) (1,1),0),0),0) (1,1),0),0),0) (1,1),0),0),0),0),0),0),0),0),0),0),0),0),0)						0.000 ** 0.024 0.027 * 0.027 * 0.052
Alc: 130.39; Concordance Index: 0.88 Age Primary diagnosis Tissue or organ of origin Gender B881 S882	(N=170) (Alericgaronoma, NOS (Alericgaronoma, NOS (Alericgarono	۲ (1,1-0,-0) (1				-		0.000 ** 0.824 0.02 * 0.02 * 0.02 * 0.02 * 0.0542 0.0542 0.0542
Ak: 132.39; Concordance Index: 0.88 Age Primary diagnosis Tissue or organ of origin B851 B852 B853 B853 B8517b	(N=170) (N=10)	(1.1-0-0) (1.1-0-0) (1.1-0-0) (1.0-0) (1.1-0-0) (1.0-0) (1.0-0)				-		0.000 ** 0.024 0.02* 0.021* 0.021* 0.052 0.0542
AIC: 130.39; Concordance Index: 0.88 Age Primary disgnosis Tissue or organ of origin Gender B801 B802 B805	(N=170) (N=170) (N=170) (N=100) (N=	(1,1,2,0,-0) (1,1,2,0,-0)				-		0.000 ** 0.02* 0.02* 0.027 * 0.0552 0.0552 0.0552 0.0552
Ak: 132.39; Concordance Index: 0.88 Age Primary diagnosis Tissue or organ of origin B851 B852 B853 B853 B8517b	(N=170) (N=10)	(1.1-0-0) (1.1-0-0) (1.1-0-0) (1.0-0) (1.1-0-0) (1.0-0) (1.0-0)						0.009 ** 0.024 0.021 * 0.021 * 0.0512 0.0542 0.0542

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.

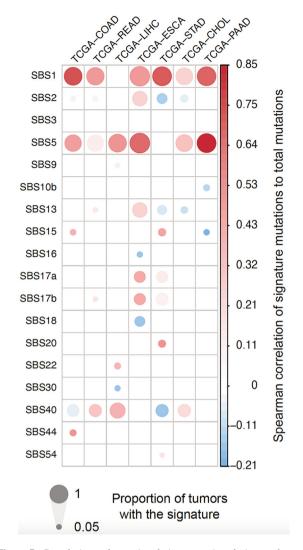


Figure 7. Correlations of mutational signatures in relation to the total mutation rate in the studied tumor types in TCGA material. Red indicates higher and blue lower values, respectively. The size of the dots indicates the proportion of tumors with each signature. COAD: Colon adenocarcinoma; CHOL: 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 wholegenome 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 covariates 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

- 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