nature medicine

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41591-019-0462-y

In the format provided by the authors and unedited.

Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer

Jakob Nikolas Kather[1,2,3,4,5]*, Alexander T. Pearson[4], Niels Halama[2,5,6], Dirk Jäger[2,3,5], Jeremias Krause[1], Sven H. Loosen[1], Alexander Marx[7], Peter Boor[8], Frank Tacke[9], Ulf Peter Neumann[10], Heike I. Grabsch[11,12], Takaki Yoshikawa[13,14], Hermann Brenner[2,15,16], Jenny Chang-Claude[17,18], Michael Hoffmeister[15], Christian Trautwein[1] and Tom Luedde[1]*

[1]Department of Medicine III, University Hospital RWTH Aachen, Aachen, Germany. [2]German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany. [3]Applied Tumor Immunity, German Cancer Research Center (DKFZ), Heidelberg, Germany. [4]Hematology/Oncology, Department of Medicine, University of Chicago, Chicago, IL, USA. [5]Medical Oncology, National Center for Tumor Diseases (NCT), Heidelberg, Germany. [6]Translational Immunotherapy, German Cancer Research Center (DKFZ), Heidelberg, Germany. [7]Institute of Pathology, University Medical Center Mannheim, Heidelberg University, Mannheim, Germany. [8]Institute of Pathology and Department of Nephrology, University Hospital RWTH Aachen, Aachen, Germany. [9]Hepatology and Gastroenterology, Charité University Medicine, Berlin, Germany. [10]Visceral and Transplant Surgery, University Hospital RWTH Aachen, Aachen, Germany. [11]Pathology & Data Analytics, Leeds Institute of Medical Research at St James's, University of Leeds, Leeds, UK. [12]Pathology and GROW School for Oncology and Developmental Biology, Maastricht University Medical Center+, Maastricht, the Netherlands. [13]Department of Gastrointestinal Surgery, Kanagawa Cancer Center, Yokohama, Japan. [14]Department of Gastric Surgery, National Cancer Center Hospital, Tokyo, Japan. [15]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany. [16]Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany. [17]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany. [18]Cancer Epidemiology Group, University Cancer Center Hamburg, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. *e-mail: jkather@ukaachen.de; tluedde@ukaachen.de

Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer

– Supplementary Tables 1-5–

	TCGA-STAD	**TCGA-CRC-KR**	**TCGA-CRC-DX**	**TCGA-UCEC**	**DACHS**	**KCCH**
Material	FFPE	snap frozen	FFPE	FFPE	FFPE	FFPE
Staining	HE	HE	HE	HE	HE	HE
N patients	315	387	360	327	378	185
Median age [years]	67	67	67	63	68	65
% UICC stage 1	13%	17%	17%	69%	20%	0%
% UICC stage 2	31%	37%	37%	6 %	33%	39%
% UICC stage 3	44%	29%	30%	19%	33%	55%
% UICC stage 4	10%	12%	13%	4%	14%	6%

Supplementary Table 1: Clinico-pathological variables of all patient cohorts. STAD = stomach adenocarcinoma, CRC = colorectal cancer, KR = snap-frozen slides, DX = diagnostic slides with FFPE processing, FFPE = formalin-fixed and paraffin-embedded, HE = hematoxylin and eosin, UICC = Union internationale contre le cancer, UCEC = uterine corpus endometrial carcinoma, KCCH = Yokohama gastric cancer cohort, DACHS = German colorectal cancer cohort, MSI = microsatellite instable, NA = not applicable.

	TCGA-STAD	TCGA-CRC-KR	TCGA-CRC-DX	TCGA-UCEC	DACHS	KCCH
Class MSI	MSI-H (sequencing, Ref.: (1)) Hypermutated with missing MSI status (2)			MSI-H (genetic) (3)	MSI found with genetic test* (4)	MMRd found with 4-IHC-panel (5)
Class MSS	MSS (1)			MSS (3)	Genetic test negative (4)	Normal IHC (5)
Exclude	MSI-L (1)			MSI-L (3)	None	None

Supplementary Table 2: Definition of MSI in all sets. In distinct cohorts, different methods to assign patients to "MSI" or "MSS" were used. This table is a summary of the patients that were assigned to the respective classes, explaining differences in MSI prevalence between the cohorts. MMRd = mismatch-repair deficiency, IHC = immunohistochemistry. *In DACHS, MSI status was determined genetically using a 3-marker panel (BAT25, BAT26, CAT25) as described in (6).

	TCGA-STAD Test set	**TCGA-CRC-KR Test set**	**TCGA-CRC-DX Test set**	**TCGA-UCEC Test set**	**DACHS**	**KCCH**
N patients in test set	98	109	100	110	378	185
% MSI in these patients	25.5 %	27.5 %	26.0 %	36.4 %	7.4 %	9.7 %
% MSI in stage 1	61.5 %	25.0 %	28.6 %	40.7 %	6.7 %	N/A
% MSI in stage 2	28.1 %	42.2 %	42.9 %	25.0 %	11.1 %	9.7 %
% MSI in stage 3	17.0 %	15.3 %	8.6 %	31.6 %	7.2 %	9.8 %
% MSI in stage 4	9.1 %	11.1 %	6.3 %	0.0 %	0.0 %	9.1 %
% MSI image tiles	23.6 %	22.5 %	28.6 %	35.9 %	7.7 %	13.9 %

Supplementary Table 3: Prevalence of microsatellite instability (MSI) in all test sets. For definition of "MSI" in these groups, see Supplementary Table 2.

Correlation of MSIness to …	**Ref.**	**TCGA-STAD**	**TCGA-CRC-KR**	**TCGA-CRC-DX**	**DACHS**
N patients	-	N=91 from test set	N=105 from test set	N=95 from test set	Subset N=134
CD8+ T-cells signature	(1)	**cor = 0.413 p = 5.16e-5 *****	cor = 0.0498 p = 0.617	cor = 0.088 p = 0.401	N/A
CD8+ T-cells (IHC for CD8)	(7)	N/A	N/A	N/A	cor = 0.050 p = 0.567
PD-L1 expression	(8)	cor = 0.044 p = 0.677	**cor = 0.292 p = 0.00247 ****	**cor = 0.408 p = 4.1e-5 *****	N/A
IFN-gamma signature	(1)	cor = 0.175 p = 0.099	**cor = 0.452 p = 1.63e-6 *****	**cor = 0.350 p = 0.0005 *****	N/A
Macrophages M1	(1)	**cor = 0.252 p = 0. 016 ***	cor = 0.032 p = 0.743	cor = 0.141 p = 0.173	N/A
Macrophages (IHC for CD163)	(7)	N/A	N/A	N/A	**cor = 0.207 p = 0.017 ***

Supplementary Table 4: Correlation of MSIness with immune gene expression signatures in test cohorts of gastrointestinal cancer. * < 0.05, ** <0.01, *** < 0.001; cor = Pearson correlation coefficient (calculated with R version 3.5.1 cor.test), IFN-gamma = interferon gamma, expression = gene expression signature, IHC = immunohistochemistry, N/A = not available. For the KCCH cohort, no genetic or IHC data were available. This table is related to Figure 2e in the main manuscript. Raw data are available through the respective references (column "Ref."). All statistical tests are two-sided.

covariates	Hazard ratio (HR)	HR low CI	HR high CI	p-value
only MSIness	1.647	1.074	2.524	0.0221 *
+ age	1.598	1.041	2.452	0.0320 *
+ age + sex	1.569	1.016	2.423	0.0420 *
+ age + sex + stage	1.373	0.883	2.137	0.1598

Supplementary Table 5: Prognostic value of high vs low MSIness in genetically determined MSS tumors in the DACHS cohort, multivariable Cox models. CI = 95% confidence interval, * < 0.05. MSIness was binarized at the mean value between true MSI and true MSS patients. These survival models were fitted using the 'survival' package in R (R-project.org). A Kaplan-Meier survival curve for patients stratified by MSIness is shown in Suppl. Figure 3b. This analysis was performed on all N=350 patients with genetic MSS status. All statistical tests are two-sided.

References for Supplementary Tables

1. Liu Y, Sethi NS, Hinoue T, Schneider BG, Cherniack AD, Sanchez-Vega F, et al. Comparative Molecular Analysis of Gastrointestinal Adenocarcinomas. Cancer Cell. 2018;33(4):721-35 e8.
2. Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, et al. Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell. 2018;174(4):1034-5.
3. Cancer Genome Atlas Research N, Kandoth C, Schultz N, Cherniack AD, Akbani R, Liu Y, et al. Integrated genomic characterization of endometrial carcinoma. Nature. 2013;497(7447):67-73.
4. Hoffmeister M, Blaker H, Kloor M, Roth W, Toth C, Herpel E, et al. Body mass index and microsatellite instability in colorectal cancer: a population-based study. Cancer Epidemiol Biomarkers Prev. 2013;22(12):2303-11.
5. Aoyama T, Hutchins G, Arai T, Sakamaki K, Miyagi Y, Tsuburaya A, et al. Identification of a high-risk subtype of intestinal-type Japanese gastric cancer by quantitative measurement of the luminal tumor proportion. Cancer Med. 2018;7(10):4914-23.
6. Brenner H, Chang-Claude J, Seiler CM, Hoffmeister M. Long-term risk of colorectal cancer after negative colonoscopy. J Clin Oncol. 2011;29(28):3761-7.
7. Kather JN, Suarez-Carmona M, Charoentong P, Weis CA, Hirsch D, Bankhead P, et al. Topography of cancer-associated immune cells in human solid tumors. Elife. 2018;7.
8. Anaya J. OncoLnc: linking TCGA survival data to mRNAs, miRNAs, and lncRNAs. PeerJ Computer Science. 2016;2:e67.