Supplementary Material

Fusobacterium nucleatum in situ hybridization and image analysis

RNA *in situ* hybridization (RNA-ISH) was conducted using the RNAscope® technology (Advanced Cell Diagnostics USA; Newark, CA) on the Ventana Discovery Ultra automated platform (Ventana Medical Systems; Tucson, AZ). B-Fusobacterium 23s RNA probe was obtained by ACD (cat no. 486411, accession no: CP003723). The sequence of the F. nucleatum-targeted probe was AATATGTCAAGGATACAGCCGTTGTATTTAAGGGGTTGAG GGACAAAGTAGTGAAGAACTGTAAGATATTCAATATAGTGTATTGATGAATTAGAATTG TATGGAAAGATAAACCGCAGAAGGTGAGAGTCCTGTATAAGTAAATCTTTACACATAT AACT. The RNAscope® 2.5 HD Reagent Kit-RED assay and Fast Red dye was used. Briefly, formalin fixed paraffin embedded (FFPE) samples were cut into 5 µm sections and deparaffinized for 32 minutes at 60ºC. After antigen retrieval, the probe was hybridized for 120 minutes at 43°C followed by signal amplification and visualization using a red chromogen. The counterstaining was performed with the Haematoxyln II before slides were finally mounted with an aqueous medium. Slides were digitalized using the NanoZoomer 2.0-HT scanner (Hamamatsu Photonics, Japan) and quality checked by a pathologist before digital image analysis (DIA). For signal quantification, we used an algorithm created through Visiopharm ® image analysis software (Author™ module; Hoersholm, Denmark) to automatically detect and count individual red signal corresponding to single bacteria mRNA molecule within a determined region of interest (ROI). For signals clustering together, the algorithm divided the total stained area by the median area of each individual dot. After automated tissue detection, tumour ROIs were manually refined by a board-certified pathologist to exclude normal tissue. Whenever possible, a ROI corresponding to adjacent normal rectal mucosa was drawn and included in the analysis. Results were expressed in density values (number of bacterial cells per mm^2 of tissue).

Assay specificity was determined using HCT116 cell lines infected with increasing dose of *F. nucleatum* (multiplicity of infection: 0, 1, 10, 100) previously used in [1]. Cell lines were fixed in formalin and embedded in paraffin and cut in 5 µm thick sections for RNA-ISH which was run in triplicate as described above. For comparative analysis between RNA-ISH and qPCR, we used qPCR data from 71 CRC samples of the “FFPE paired cohort” studied in [1]. Briefly, DNA was extracted with the QIAamp DNA FFPE Tissue Kit (Qiagen Inc., USA) according to manufactures instructions. A custom TaqMan primer/probe set was used to amplify *Fusobacterium* genus DNA (Integrated DNA technologies, CA) as previously described [2]. The primer and probe sequences for each TaqMan assay were as follows: Fusobacterium genus forward primer, 5’-AAGCGCGTCTAGGTGGTTATGT-3’; *Fusobacterium* genus reverse primer, 5’- TGTAGTTCCGCTTACCTCTCCAG-3’; *Fusobacterium* genus FAM probe, 5’- CAACGCAATACAGAGTTGAGCCCTGCATT-3’. PGT forward primer, 5’- A TCCCCAAAGCACCTGGTTT-3’; PGT reverse primer, 5’- AGAGGCCAAGA T AGTCCTGGTAA-3’; PGT FAM probe, 5’- CCATCCATGTCCTCATCTC-3’. The cycle threshold (Ct) values for *Fusobacterium* genus was normalized to the amounts of human genomic DNA in each reaction by using a primer and probe set for the prostaglandin transporter (PGT) reference gene, and the fold difference (2-ΔCt) of *Fusobacterium* load in tumour tissue was calculated as described before [3]. RNA-ISH was performed on non-consecutive sections from the same samples submitted to qPCR. To compare results obtained with the two methodologies (RNA-ISH and qPCR), we log10-transformed RNA-ISH continuous values. For categorical variables

comparisons, samples were divided in two qPCR groups based on arbitrary thresholds of expression: Negative (Ct ≥36) and Positive (Ct <36).

Immunohistochemistry

IHC was performed on consecutive sections using CD3, CD8, PD-L1 and Pan-Keratin antibodies. Briefly, the slides were heated and deparaffinized before heat-induced antigen retrieval. Primary antibodies were applied as indicated in supplementary Table S1. Reactions were detected using the UltraView Universal DAB Detection kit. Finally, the slides were counterstained with Haematoxylin and mounted with a Xylol based mounting medium. The entire process was performed in the Benchmark ULTRA system and all reagents were from Ventana Medical Systems. Slides were digitalized using a slide scanner (NanoZoomer 2.0-HT, Hamamatsu Photonics, Japan) and quality checked by a pathologist before DIA. For IHC staining quantification, we used different image analysis algorithms created through the Author™ module of Visiopharm®. CD3, CD8 and PD-L1 slides were aligned with the Pan-Keratin slide using Tissuealign® (Visiopharm®). In the case of CD3 and CD8, Pan-Keratin was used to select the tumour and the surrounding stroma (200µm from the tumour) areas where densities of stained immune cells were calculated. For PD-L1 analysis, the Composite Positive Score (CPS) was calculated dividing the number of PD-L1 positive cells by the total number of Pan-Keratin positive tumour cells multiplied by 100 [4].

References

1. Bullman S, Pedamallu CS, Sicinska E et al. Analysis of *Fusobacterium* persistence and antibiotic response in colorectal cancer. Science (80-.). 2017; 358(6369):1443–1448.
2. Martin FE, Nadkarni MA, Jacques NA, Hunter N. Quantitative Microbiological Study of Human Carious Dentine by Culture and Real-Time PCR: Association of Anaerobes with Histopathological Changes in Chronic Pulpitis. J. Clin. Microbiol. 2002; 40(5):1698–1704.
3. Castellarin M, Warren RL, Freeman JD et al. Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. Genome Res. 2012; 22(2):299–306.
4. Kulangara K, Zhang N, Corigliano E et al. Clinical utility of the combined positive score for programmed death ligand-1 expression and the approval of pembrolizumab for treatment of gastric cancer. Arch. Pathol. Lab. Med. 2019; 143(3):330–337.

Supp. Table S1. Primary antibodies and protocols used for immunohistochemistry.

Supp. Table S2. Two-way comparisons of RNA-ISH densities of cell lines infected with *F. nucleatum* with increasing multiplicity of infection (0, 1, 10, 100). Student T test p value is shown for each comparison.

Supp. Table S3. Two-way table with observed frequencies of *F. nucleatum* positive (POS) and negative (NEG) cases by RNA-ISH and qPCR.

Supp. Table S4. Clinico-pathological variables of treated cohort.

Supp. Table S5. Association between clinical-pathological variables and relapse-free survival (left) and response to neoadjuvant therapy from treated cohort.

Supp. Table S6. *F. nucleatum* (density values and frequency of positives cases) by sample type and cohort.

Supp. Table S7. Comparison of clinic-pathological variables between patients from untreated samples by *F. nucleatum* status.

Supp. Table S8. Contingency table of *F. nucleatum* status and response to nCRT by sample type from treated cohort (OR: odds ratio, Responders: pNCR and pCR).

Supp. Table S9. Multivariate Cox regression model for assessing the relationship between relapse-free survival time and covariates (*F. nucleatum* status in post-nCRT samples and type of neoadjuvant treatment) from treated cohort (HR: hazard ratio, RT: radiotherapy).

Supp. Table S10. Univariate Cox regression model for relapse-free survival time and the *F. nucleatum* status group change type (N-N: negative-negative, P-N: positive-negative and P-P: positive-positive) from paired treated cohort.

Supplementary Table S1

Antigen	Clone	Dilution	Manufacturer and reference	Primary antibody incubation time and temperature
CD3	2GV6	Ready-to-use	Ventana Medical Systems (# 790-4341)	32 min 36ºC
CD8	144B	1/100	Dako Agilent (#M7103)	32 min RT
PD-L1	SP263	Ready-to-use	Ventana Medical Systems (#741-4905)	16 min 36ºC
Pan Keratin	AE1/AE3 & PCK26	Ready-to-use	Ventana Medical Systems (#760-2595)	36 min RT

Supplementary Table S2

	0	**1**	**10**	**100**
0	-	.002	.0001	.004
1	.002	-	.001	.005
10	.0001	.001	-	.007
100	.004	.005	0.007	-

Supplementary Table S3

		qPCR	
		NEG	POS
RNA-ISH	NEG	41	6
	POS	4	20

Supplementary Table S4

Variable	*N*	Frequency (%)
Num. of pts.	87	
SEX		
MALE	53	60.9
FEMALE	34	39.1
HISTOLOGY		
CONVENTIONAL	78	91.8
MUCINOUS	7	8.2
NA's	2	
STAGE		
I	0	0.0
II	7	8.1
III	79	91.9
NA's	1	
TYPE OF SAMPLES		
ENDOSCOPY	71	81.6
SURGERY	87	100.0
***F.NUCLEATUM* STATUS IN ENDOSCOPY SAMPLES**		
POSITIVE	45	64.3
NEGATIVE	25	35.7
NA's	17	
NEOADJUVANT REGIMEN		
CHEMO + RT	69	79.3
RT	18	20.7
TUMOUR REGRESSION GRADE		
GR0	2	2.4
GR1	6	7.1
GR2	13	15.3
GR3	38	44.7
GR4	23	27.1
GR5	3	3.5
NA's	2	
RESPONSE TO NEOADJUVANT THERAPY		
pCR	8	9.4
pNCR	13	15.3
NO RESPONSE	64	75.3
NA's	2	
***F.NUCLEATUM* STATUS IN SURGERY SAMPLES**		
POSITIVE	22	25.9
NEGATIVE	63	74.1
NA's	2	
RELAPSE		
0	67	77.0
1	20	23.0
SURVIVAL		
ALIVE	67	71.3
DEAD	25	28.7

Supplementary Table S5

	RELAPSE FREE-SURVIVAL					RESPONSE TO NEOADJUVANT THERAPY			
Factor (vs. Ref.)	**N**	**Events**	**Median (95% CI)**	**HR univariate (95%CI)**	**P value**	**pCR+pNCR**	**No response**	**OR (95%CI)**	**P value**
(ref.) Female	34	7	NR (NR-NR)	1.3 (0.5-3.3)	0.6	10 (29%)	24 (71%)	1.5 (0.5-4.6)	0.45
Male	53	13	NR (NR-NR)			11 (22%)	40 (78%)		
(ref.) Stage II	7	0	NR (NR-NR)	>100 (0-NR)	0.1	4 (67%)	2 (33%)	1.6 (0.5-4.8)	0.44
Stage III	79	20	NR (NR-NR)			17 (22%)	61 (78%)		
(ref.) Conventional	78	20	NR (NR-NR)	<1 (0-NR)	0.1	19 (25%)	58 (75%)	>100 (0.3-NR)	0.33
Mucinous	7	0	NR (NR-NR)			0 (0%)	6 (100%)		
(ref.) pCR+pNCR	21	2	NR (NR-NR)	3.5 (0.8-15.0)	0.07				
No response	64	18	NR (NR-NR)						

Supplementary Table S6

Cohort	Sample type	*N*	Frequency *F.nucleatum* positive	Median	IQR	All samples	Treated vs Untreated endoscopy	Surgery vs Treated endoscopy
Treated	Endoscopy	71	64%	13.4	2.4-92.7	P=0.26	P=0.23	P=0.13
Untreated	Endoscopy	40	53%	4.4	1.3-36.6			
Untreated	Surgery	56	50%	4.0	2.1-22.6			

Supplementary Table S7

	PATIENTS BY *F. NUCLEATUM* STATUS IN UNTREATED SAMPLES		
	NEGATIVE (N=53)	**POSITIVE (N=73)**	**Fisher test**
AGE AT DIAGNOSIS	71.4 (45.2-86.9)	71.2 (45.0-90.6)	P=0.94
SEX			
MALE	37 (47%)	42 (53%)	P=0.19
FEMALE	16 (34%)	31 (66%)	
HISTOLOGY			
CONVENTIONAL	44 (41%)	63 (59%)	P=0.59
MUCINOUS	8 (50%)	8 (50%)	
STAGE			
I	8 (50%)	8 (50%)	P=0.55
II	8 (35%)	15 (65%)	
III	36 (42%)	50 (58%)	
RELAPSE			
0	45 (44%)	58 (56%)	P=0.49
1	8 (35%)	15 (65%)	
SURVIVAL			
ALIVE	39 (42%)	54 (58%)	P=1.00
DEAD	14 (42%)	19 (58%)	
nCRT			
YES	25 (36%)	45 (56%)	P=0.15
NO	28 (50%)	28 (50%)	

Supplementary Table S8

Sample type	*F. nucleatum* status	Evaluable (%)	**Responder**	**Non-responder**	NA's	OR (CI95%)	p-value
Pre-treatment N=71 endoscopies	**Positive**	44 (65%)	15 (34%)	29 (66%)	1	3.6 (1.0-17.1)	0.08
	Negative	24 (35%)	3 (13%)	21 (88%)	1		
	NA's		0	1			
Post-treatment N=87 surgeries	**Positive**	21 (25%)	2 (10%)	19 (90%)	1	0.3 (0.03-1.4)	0.13
	Negative	62 (75%)	17 (27%)	45 (73%)	1		
	NA's		2	0			

Supplementary Table S9

Factor (vs. Ref.)	**HR**	**95%CI**	**p-value**
F. nucleatum			
Positive (vs Negative)	7.3	2.9-18.8	<0.001
Treatment regimen			
RT (vs Chemo+RT)	1.0	0.4-2.8	0.97

Supplementary Table S10

Factor (vs. Ref.)	*N*	Events	Median	(95% CI)	HR univariate	95% CI	*P* value
(Ref.) N-N	23	3	NR	NR-NR			
P-N	30	2	NR	NR-NR	0.5	0.1-3.0	0.45
P-P	15	9	23.8	12.5-NR	6.5	1.7-24.0	0.005
(Ref.) N-N + P-N	53	5	NR	NR-NR			
P-P	15	9	23.8	12.5-NR	9	3.0-27.2	<0.001

Supp. Figure S1. Flowchart of samples and patients included for study. Of the 254 samples analysed, 158 belong to patients treated with nCRT (Treated cohort) and 96 are from untreated patients (Control cohort). In both cohorts, there are patients who provided paired samples. Samples were also classified as treated (87) and untreated (127). Only in the treated cohort are there samples (90) with immune cell data.

Supp. Figure S2. Kaplan-Meier curves for relapse-free survival (RFS) of: (A) Unstratified Control cohort. (B) Control cohort by *F. nucleatum* status in stage II and III surgery samples. (C) Treated cohort by *F. nucleatum* post-nCRT after excluding patients achieving pCR.

Supp. Figure S3. Change of *F. nucleatum* in paired samples. A) Parallel coordinate plot and boxplot of the change of *F. nucleatum* density values (in logarithmic scale for a better visualization) between pre-nCRT and post-nCRT samples. Each line joins the paired samples from the same patient. Colors group patients according to the *F. nucleatum* status group change type (N-N, P-N and P-P). B) Representative RNA-ISH images of F. nucleatum in pre-nCRT and post-nCRT paired samples. *F. nucleatum* is shown in red. Microbiotype status group change type: N-N, negative-negative; P-N, positive-negative; P-P positive-positive (digital magnification, 10x).

Supp. Figure S4. Boxplot of the time between nCRT end and surgery (weeks) stratified by *F. nucleatum* status group change type. Time from end of nCRT to surgery was not significantly different in the three microbiotype status group change types. N-N, negative-negative; P-N, positive-negative; P-P positive-positive.

Supp. Figure S5. Multiple boxplot of CD3 + (A), CD8 + (B), PD-L1 + (C) cell density values by pre-nCRT and post-nCRT samples (left) and stratified by *F. nucleatum* status (right), respectively. Differences were found according to the type of sample only in CD8 + ($P < 0.001$). *F. nucleatum* status was significantly different in all post treatment samples (POST) from the three types of immune cells ($P = 0.004$ (A), $P = 0.001$ (B), $P = 0.02$ (C)). However, no differences were found in pre-treatment samples.