

# SCIENTIFIC REPORTS

OPEN

## Antibody responses to flagellin C and *Streptococcus gallolyticus* pilus proteins in colorectal cancer

Julia Butt<sup>1,2</sup>, Nerea Fernández de Larrea<sup>3,4</sup>, Harold Tjalsma<sup>5</sup>, Rian Roelofs<sup>6</sup>, Ikuko Kato<sup>7</sup>, Vicente Martín<sup>4,8</sup>, Beatriz Pérez-Gómez<sup>3,4,9</sup>, Víctor Moreno<sup>4,10,11,12</sup>, Trinidad Dierssen-Sotos<sup>4,13</sup>, Jesús Castilla<sup>4,14</sup>, Guillermo Fernández-Tardón<sup>4,15</sup>, Pilar Amiano<sup>4,16</sup>, Dolores Salas<sup>4,17,18</sup>, Juan Alguacil<sup>4,19</sup>, José Juan Jiménez-Moleón<sup>4,20,21</sup>, José María Huerta<sup>4,22</sup>, Silvia de Sanjosé<sup>4,23,24</sup>, Rosa del Campo<sup>4,25,26</sup>, Manolis Kogevinas<sup>4,27,28,29</sup>, Marina Pollán<sup>3,4</sup>, Michael Pawlita<sup>1</sup>, Tim Waterboer<sup>1</sup>, Annemarie Boleij<sup>30</sup> & Nuria Aragonés<sup>4,31</sup>

Antibodies to *Streptococcus gallolyticus* subspecies *gallolyticus* (SGG) have been associated with colorectal cancer (CRC). Because SGG may correlate with impaired gut epithelia, we assessed the association of antibodies to bacterial flagellin C (FliC), a measure potentially related to this impairment, with CRC and the CRC-specific interaction with antibodies to SGG proteins. Antibodies to FliC and SGG pilus proteins Gallo2178 and Gallo2179 were measured in two independent studies, a combined study from Nijmegen and Detroit (93 CRC cases, 74 controls) and a replication data set including 576 cases and 576 controls from the Spanish multicenter multicase-control study (MCC-Spain). Logistic regression was applied to assess whether antibodies to FliC were associated with CRC and modified the association of antibodies to SGG proteins with CRC. Antibodies to FliC were associated with those to SGG Gallo2178 among CRC cases, resulting in an interaction in the association of antibodies to Gallo2178 with CRC ( $p = 0.007$ ). This association was only present among individuals with high antibody responses to FliC (OR: 2.42, 95% CI: 1.45–4.06). In conclusion, our findings suggest that colorectal tumorigenesis could be accompanied by an impaired integrity of the epithelium that could result in associated increased antibody responses to bacterial proteins.

<sup>1</sup>Infection and Cancer Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany. <sup>2</sup>Faculty of Biosciences, University of Heidelberg, Heidelberg, Germany. <sup>3</sup>Cancer and Environmental Epidemiology Unit, National Center for Epidemiology, Carlos III Institute of Health, Madrid, Spain. <sup>4</sup>Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP), Madrid, Spain. <sup>5</sup>Independent researcher. Experimental work conducted at the Department of Laboratory Medicine, Radboud University Medical Centre (RadboudUMC), Nijmegen, The Netherlands. <sup>6</sup>Department of Laboratory Medicine, Radboud Institute for Molecular Life Sciences (RIMLS), Radboud university medical centre (Radboudumc), Nijmegen, The Netherlands. <sup>7</sup>Departments of Oncology and Pathology, Wayne State University, Detroit, Michigan, USA. <sup>8</sup>The Research Group in Gene Environment and Health Interactions, University of León, León, Spain. <sup>9</sup>Cardiovascular and Metabolic Diseases Unit, National Center for Epidemiology, Carlos III Institute of Health, Madrid, Spain. <sup>10</sup>Oncology Data Analytics Program, Catalan Institute of Oncology (ICO), Hospitalet de Llobregat, Barcelona, Spain. <sup>11</sup>Colorectal Cancer Group, ONCOBELL Program, Bellvitge Biomedical Research Institute (IDIBELL), Hospitalet de Llobregat, Barcelona, Spain. <sup>12</sup>Department of Clinical Sciences, Faculty of Medicine, University of Barcelona, Barcelona, Spain. <sup>13</sup>University of Cantabria – IDIVAL, Santander, Spain. <sup>14</sup>Public Health Institute of Navarra, IdiSNA- Navarra Institute for Health Research, Pamplona, Spain. <sup>15</sup>IUOPA, University of Oviedo, Oviedo, Spain. <sup>16</sup>Public Health Division of Gipuzkoa, BioDonostia Research Institute, San Sebastian, Spain. <sup>17</sup>Cancer and Public Health Area, FISABIO – Public Health, Valencia, Spain. <sup>18</sup>General Directorate Public Health, Valencian Community, Valencia, Spain. <sup>19</sup>Natural Resources, Health and Environment Research Center (RENSMA), University of Huelva, Huelva, Spain. <sup>20</sup>Granada Health Research Institute (ibs.GRANADA), Granada, Spain. <sup>21</sup>Department of Preventive Medicine and Public Health, University of Granada, Granada, Spain. <sup>22</sup>Department of Epidemiology, Murcia Regional Health Council, IMIB-Arrixaca, Murcia, Spain. <sup>23</sup>Cancer Epidemiology Research Program, Catalan Institute of Oncology-IDIBELL, L'Hospitalet de Llobregat, Barcelona, Spain. <sup>24</sup>PATH, Reproductive Health, Seattle, USA. <sup>25</sup>Department of Microbiology, Ramón y Cajal Health Research Institute (IRYCIS), Hospital Universitario Ramón y Cajal, Madrid, Spain. <sup>26</sup>Spanish Network for Research in Infectious Diseases (REIPI), Madrid, Spain. <sup>27</sup>ISGlobal, Barcelona, Spain. <sup>28</sup>IMIM (Hospital del Mar Medical Research

Institute), Barcelona, Spain. <sup>29</sup>Pompeu Fabra University (UPF), Barcelona, Spain. <sup>30</sup>Department of Pathology, Radboud Institute for Molecular Life Sciences (RIMLS), Radboud university medical centre (Radboudumc), Nijmegen, The Netherlands. <sup>31</sup>Epidemiology Section, Public Health Division, Department of Health of Madrid, Madrid, Spain. Annemarie Boleij and Nuria Aragonés jointly supervised this work. Correspondence and requests for materials should be addressed to J.B. (email: [julia.butt@duke.edu](mailto:julia.butt@duke.edu))

Colorectal cancer (CRC) is the fourth most common cancer worldwide with in total 1.36 million newly diagnosed cases in 2012<sup>1</sup>. Risk factors include older age, male sex, certain dietary habits, tobacco and alcohol consumption, obesity, family history of CRC as well as history of inflammatory bowel diseases<sup>2</sup>. Interest in a potential association of bacteria residing in the gut with CRC development has increased during the past years<sup>3,4</sup>. In particular, *Streptococcus gallolyticus* subspecies *gallolyticus* (SGG) was shown to be associated with CRC in several epidemiological studies but also in some mechanistic studies<sup>5–14</sup>. Boleij *et al.* showed that antibody responses to SGG pilus proteins Gallo2178 and Gallo2179, both being hypothesized to be involved in virulence of the bacterium<sup>15,16</sup>, were selectively detectable in adenoma and CRC cases in two studies from Nijmegen (Netherlands), and Detroit (USA)<sup>6</sup>. We reproduced this finding in two case-control studies, including a subset of the multicentric multicase-control study MCC-Spain, as well as one prospective study and demonstrated a significant association of antibody responses to Gallo2178 and Gallo2179 with an approximately 2-fold increased risk of developing CRC<sup>8,9,12</sup>.

SGG is a rare gut microbe in humans and occasionally causes systemic infection, facilitated by increased mucosal permeability, as seen in tumor development, as well as by strain specific virulence<sup>4</sup>. An impaired gut barrier, however, should also enable other organisms to invade the otherwise well-protected gut epithelium. Bacterial products like lipopolysaccharide (LPS), also called endotoxin, or motility-proteins like flagellin C (FliC) are wide-spread among bacterial species and upon infection are targets of the host immune system<sup>17</sup>. Several studies assessed the association of these common bacterial products with diseases of the gut, e.g. Crohn's disease or CRC and its precursors<sup>7,18–21</sup>. Kang *et al.* found that individuals with high plasma endotoxin levels had a 1.4-fold increased risk of having adenomas<sup>20</sup>, and Kong *et al.* found that high levels of antibody responses to *Escherichia coli* (*E. coli*) LPS and FliC were prospectively associated with CRC in men (OR: 1.66, 95% CI: 1.10–2.51)<sup>19</sup>. In two other studies from the Netherlands and USA, *Salmonella* FliC was associated with colorectal adenomas (OR 4.71, 95% CI 1.10–20.14) and carcinomas (OR 3.09, 95% CI 1.22–7.79), while endotoxin levels were not<sup>7,21</sup>.

Using serological data obtained in the CRC and adenoma case-control studies from Nijmegen and Detroit and data obtained in MCC-Spain as a replication study, we here assessed whether antibody responses to FliC were simultaneously associated with CRC and with antibody responses to Gallo2178 and Gallo2179, as suggested by the gut permeability hypothesis.

## Results

**Association of antibody responses to FliC with CRC.** In the Nijmegen/Detroit study, the OR for adenoma/CRC significantly increased with higher FliC quartile ( $p_{\text{trend}} = 0.011$ ) with a statistically significant association of the fourth *versus* the first quartile (OR: 2.32, 95% CI: 1.01–5.33) (Table 1, see Supplementary Table S2 for individual studies). In the replication study MCC-Spain, the OR for CRC did not show a positive trend, although the highest risk was found in the highest FliC quartile (Q4 vs Q1, OR: 1.14, 95% CI: 0.81–1.59).

**Association of antibody responses to FliC with antibody positivity to SGG pilus proteins Gallo2178 and Gallo2179.** We assessed in both studies, Nijmegen/Detroit and MCC-Spain, whether positivity to SGG pilus proteins was associated with antibody responses to FliC. In the total population of the Nijmegen/Detroit study, there was an association between antibody response to FliC and positivity to Gallo2179 ( $p = 0.040$ ) but not to Gallo2178 ( $p = 0.072$ ) (Table 2). However, when analyzing cases and controls separately we observed that antibody positivity to Gallo2178 and Gallo2179 were significantly associated with FliC antibody response among cases ( $p = 0.023$  and  $p = 0.021$ , respectively) but not among controls ( $p = 0.658$  and  $p = 0.471$ , respectively) (see Supplementary Table S3 for individual studies).

In the replication study MCC-Spain, we partly reproduced the findings observed in the Nijmegen/Detroit study (Table 2): While there was no association between FliC quartile and positivity to neither Gallo2178 nor to Gallo2179 in the whole population, antibody positivity to Gallo2178 was significantly associated with FliC antibody response among cases ( $p = 0.011$ ). The majority of cases positive for Gallo2178 (88% in the Nijmegen/Detroit study and 71% in MCC-Spain) had simultaneously antibody responses to FliC in the upper quartiles 3 and 4 ( $>379.5$  MFI).

**Association of antibody responses to SGG Gallo2178 and Gallo2179 with CRC stratified by FliC antibody response in MCC-Spain.** The abovementioned analysis revealed a significant association between antibody positivity to SGG proteins with high antibody responses to FliC only among CRC cases. We performed an analysis of the association between antibody responses to SGG pilus proteins and CRC stratified by median MFI to FliC. The analysis was performed in the MCC-Spain study only since the sample size of the Nijmegen/Detroit study did not allow analysis with sufficient statistical power.

Among individuals with antibody responses to FliC below the median MFI, the association between antibody responses to SGG pilus proteins and CRC was absent (Gallo2178: OR: 0.85, 95%CI: 0.48–1.49; Gallo2179: OR: 1.26, 95% CI: 0.72–2.20) (Table 3). However, among individuals with an antibody response to FliC above the median MFI, the OR for CRC with antibody positivity to Gallo2178 and Gallo2179 was increased and reached statistical significance for Gallo2178 (Gallo2178: OR: 2.42, 95%CI: 1.45–4.06; Gallo2179: OR: 1.63, 95% CI: 0.99–2.68). This interaction was significant for Gallo2178 ( $p_{\text{interaction}} = 0.007$ ).

Study	FliC	Total n (%)	Controls n (%)	Cases <sup>a</sup> n (%)	OR (95%CI) <sup>b</sup>	P <sub>trend</sub>
Nijmegen/Detroit studies	Q1	38 (23)	19 (26)	19 (20)	ref	
	Q2	29 (17)	19 (26)	10 (11)	0.53 (0.20–1.43)	
	Q3	37 (22)	17 (23)	20 (22)	1.18 (0.48–2.91)	
	Q4	63 (38)	19 (26)	44 (47)	<b>2.32 (1.01–5.33)</b>	<b>0.011</b>
MCC-Spain	Q1	296 (26)	145 (25)	151 (26)	ref	
	Q2	254 (22)	143 (25)	111 (19)	0.75 (0.53–1.06)	
	Q3	288 (25)	144 (25)	144 (25)	0.99 (0.70–1.38)	
	Q4	314 (27)	144 (25)	170 (30)	1.14 (0.81–1.59)	0.247

**Table 1.** Association of antibody response to FliC with adenoma/CRC in the Nijmegen/Detroit and MCC-Spain studies. <sup>a</sup>Cases include n = 23 adenoma and n = 70 CRC cases for Nijmegen/Detroit studies, all cases were CRC in MCC Spain; <sup>b</sup>Logistic regression model, Nijmegen/Detroit study: no further adjustment, MCC-Spain: adjustment for age, sex, province, education, BMI, smoking and family history of CRC; significant associations (p-value < 0.05) are marked in bold font; Q = quartile.

FliC quartile	Gallo2178		p-value <sup>b</sup>	Gallo2179		p-value <sup>b</sup>
	neg n (%)	pos n (%)		neg n (%)	pos n (%)	
<b>Nijmegen/Detroit studies</b>						
<b>Total</b>	<b>150 (100)</b>	<b>17 (100)</b>		<b>151 (100)</b>	<b>16 (100)</b>	
Q1	36 (24)	2 (12)		35 (23)	3 (19)	
Q2	27 (18)	2 (12)		28 (19)	1 (6)	
Q3	29 (19)	8 (47)		29 (19)	8 (50)	
Q4	58 (39)	5 (29)	0.072	59 (39)	4 (25)	<b>0.040</b>
<b>Controls</b>	<b>65 (100)</b>	<b>9 (100)</b>		<b>66 (100)</b>	<b>8 (100)</b>	
Q1	17 (26)	2 (22)		16 (24)	3 (38)	
Q2	18 (28)	1 (11)		18 (27)	1 (13)	
Q3	14 (22)	3 (33)		14 (21)	3 (38)	
Q4	16 (25)	3 (33)	0.658	18 (27)	1 (13)	0.471
<b>Cases<sup>a</sup></b>	<b>85 (100)</b>	<b>8 (100)</b>		<b>85 (100)</b>	<b>8 (100)</b>	
Q1	19 (22)	0 (0)		19 (22)	0 (0)	
Q2	9 (11)	1 (13)		10 (12)	0 (0)	
Q3	15 (18)	5 (63)		15 (18)	5 (63)	
Q4	42 (49)	2 (25)	<b>0.023</b>	41 (48)	3 (37)	<b>0.021</b>
<b>MCC-Spain</b>						
<b>Total</b>	<b>1012 (100)</b>	<b>140 (100)</b>		<b>1015 (100)</b>	<b>137 (100)</b>	
Q1	268 (26)	28 (20)		269 (27)	27 (20)	
Q2	226 (22)	28 (20)		225 (22)	29 (21)	
Q3	247 (24)	41 (29)		250 (25)	38 (28)	
Q4	271 (27)	43 (31)	0.242	271 (27)	43 (31)	0.300
<b>Controls</b>	<b>519 (100)</b>	<b>57 (100)</b>		<b>520 (100)</b>	<b>56 (100)</b>	
Q1	130 (25)	15 (26)		135 (26)	10 (18)	
Q2	126 (24)	17 (30)		127 (24)	16 (29)	
Q3	132 (25)	12 (21)		126 (24)	18 (32)	
Q4	131 (25)	13 (23)	0.757	132 (25)	12 (21)	0.354
<b>Cases<sup>a</sup></b>	<b>493 (100)</b>	<b>83 (100)</b>		<b>495 (100)</b>	<b>81 (100)</b>	
Q1	138 (28)	13 (16)		134 (27)	17 (21)	
Q2	100 (20)	11 (13)		98 (20)	13 (16)	
Q3	115 (23)	29 (35)		124 (25)	20 (25)	
Q4	140 (28)	30 (36)	<b>0.011</b>	139 (28)	31 (38)	0.268

**Table 2.** Antibody response to FliC and association with SGG Gallo2178 and Gallo2179 antibody positivity. <sup>a</sup>Cases include n = 23 adenoma and n = 70 CRC cases for Nijmegen/Detroit studies, all cases were CRC in MCC Spain; <sup>b</sup>Chi-Square test; significant associations (p-value < 0.05) are marked in bold font.

## Discussion

In this study we assessed the association of antibody responses to FliC with CRC as well as a potential interaction with antibody responses to SGG pilus proteins Gallo2178 and Gallo2179. We found a significant trend for higher antibody responses to FliC in CRC cases compared to controls in the Nijmegen/Detroit study, which could not be

	Antigen		Controls n (%)	Cases n (%)	OR (95%CI) <sup>a</sup>	p-value <sup>a</sup>	Pinteraction
Overall	Gallo2178	Neg	519 (90)	493 (86)			
		Pos	57 (10)	83 (14)	<b>1.57 (1.09–2.27)</b>	<b>0.016</b>	
FliC Q1/Q2	Gallo2178	Neg	256 (89)	238 (91)			
		Pos	32 (11)	24 (9)	0.85 (0.48–1.49)	0.564	
FliC Q3/Q4	Gallo2178	Neg	263 (91)	255 (81)			
		Pos	25 (9)	59 (19)	<b>2.42 (1.45–4.06)</b>	<b>0.0007</b>	<b>0.007</b>
Overall	Gallo2179	Neg	520 (90)	495 (86)			
		Pos	56 (10)	81 (14)	<b>1.46 (1.01–2.11)</b>	<b>0.045</b>	
FliC Q1/Q2	Gallo2179	Neg	262 (91)	232 (89)			
		Pos	26 (9)	30 (11)	1.26 (0.72–2.20)	0.427	
FliC Q3/Q4	Gallo2179	Neg	258 (90)	263 (84)			
		Pos	30 (10)	51 (16)	1.63 (0.99–2.68)	0.056	0.499

**Table 3.** Association of SGG pilus proteins Gallo2178 and Gallo2179 with CRC, overall and stratified by FliC median antibody response in MCC-Spain. <sup>a</sup>Logistic regression model with adjustment for age, sex, province, education, BMI, smoking and family history of CRC; significant associations (p-value < 0.05) are marked in bold font.

reproduced in the replication study MCC-Spain. However, antibody responses to FliC were associated with antibody responses to SGG pilus protein Gallo2178 in CRC cases of both studies resulting in a significantly stronger association of CRC with antibody positivity to Gallo2178 among individuals with high antibody responses to FliC compared to those exhibiting low responses to FliC.

FliC is a motility protein conserved among several gram-negative bacteria including *Salmonella* and *Escherichia coli*. It is recognized by the host immune system, specifically by TLR5, when it crosses the gut epithelium as it might be given in the case of a disturbed gut epithelial integrity as a result of a developing tumor<sup>22</sup>. We hypothesize that increased gut permeability in CRC development is related to an increased immune response to certain gut bacteria and that therefore antibody responses to FliC and CRC-related SGG Gallo2178 and Gallo2179 will be increased and correlated in CRC cases in contrast to controls. Our results support this hypothesis since antibody responses to SGG Gallo2178 were specifically associated with CRC in the presence of those to FliC, suggesting increased antigen sampling by the immune system. In this respect, the association of CRC with antibodies to bacterial proteins could be attributed to a complex interplay of several bacteria rather than to a single bacterial species, either invading colorectal tissue simultaneously or consecutively dependent on tumor development as proposed in the bacteria-driver passenger model<sup>4</sup>. Certain bacteria might simply be bystanders without influence on CRC development, while others might contribute to progression, e.g. by induction of inflammation.

SGG has previously been shown to promote tumor growth, *in vitro* using cell lines but also *in vivo* in mouse models. SGG may benefit from cancer metabolites in the tumor environment, and may shape its own niche environment in the developing tumor<sup>11,13–15</sup>. The effects on cell proliferation are correlated with adherence abilities of the respective SGG strains to the epithelium and can depend on the encoded pili, i.e. those expressed from loci *pil1* (Gallo2178 and Gallo2179), *pil2* and *pil3*<sup>16</sup>. Furthermore, certain strain dependencies within SGG might relate to unidentified virulence factors besides *pil1–3* to adhere to and colonize colon epithelia of CRC patients<sup>13</sup>. Comparative genomics of eight SGG strains from human blood and feces recently revealed that complete *pil1–3* loci are only present in SGG causing bacteremia and/or endocarditis, while the fecal SGG may carry truncated versions of the *pil1–3* loci<sup>23,24</sup>. This could also explain the fact that not every patient with a high FliC-titer also expresses measurable antibody responses towards both, Gallo2178 and Gallo2179.

A major strength of this work is the analysis of two separate studies, the Nijmegen/Detroit study as well as MCC-Spain as a replication study, which allowed us to assess reproducibility of the results in studies that were independently conducted. Their differences in study design and serological methodology applied, however, could also have led to the observed association of antibody responses to FliC with adenoma/CRC being present in the Nijmegen/Detroit study only. The Nijmegen/Detroit study included adenoma patients and stage I and II CRC cases whereas MCC-Spain comprised only CRC cases, including all stages. We performed an analysis separately by stage (stage I/II versus stage III/IV) in MCC-Spain and found no difference in the association of antibody responses to FliC with CRC by stage (data not shown). However, it still remains to be elucidated whether these bacteria have a role primarily in adenoma development. Furthermore, as described above, different SGG strains exhibit a distinct virulence potential by differential expression of pilus loci. In addition, the natural history of antibody responses raised against different proteins of SGG is unknown. Longitudinal studies are needed to give further insight when in relation to CRC development sero-conversion to the respective bacterial proteins appears and whether this could even differ in time for proteins of the same pilus loci. The effect on the reported associations of using ELISA in the Nijmegen/Detroit study as opposed to multiplex serology in MCC-Spain cannot be determined from the given data since direct comparisons with both assays in the same samples were not available. We therefore cannot exclude that the applied serological method further may have affected the observed differences in results between studies.

A further limitation of this study was the design and small sample size of the Nijmegen/Detroit study. We combined two independent small studies, one from Nijmegen and one from Detroit (Supplementary Tables S2 and S3). Additionally, adjusting for potential confounders was only possible in the replication study MCC-Spain.

The low number of sero-positives (e.g.  $n = 83$  Gallo2178 and  $n = 81$  Gallo2179 sero-positive cases in MCC-Spain) further refrained us from analyzing the antibody responses to Gallo2178 and Gallo2179 in quartiles in relation to FliC quartiles. This analysis of a correlation between antibody level to FliC and Gallo2189 and Gallo2179 would further strengthen the reported results but needs larger datasets for sufficient statistical power. In addition, reverse causality cannot be excluded in a case-control study and the results presented here need to be verified in prospective studies.

In conclusion, we demonstrated that antibody responses to SGG pilus protein Gallo2178 were associated with antibody responses to FliC specifically in CRC cases. This finding suggests that CRC development may be associated with increased gut permeability and concomitant antibody responses to specific CRC-associated bacteria, like SGG.

## Methods

**Study population.** *Nijmegen and detroit studies.* The study populations have been described previously<sup>6,7,21</sup>. Serum samples from 37 stage I/II CRC and 12 adenoma patients, who had been admitted to the Radboud University Medical Center (Nijmegen, the Netherlands) were collected. Serum from 27 healthy blood donors >50 years of age were used as controls. Plasma samples from patients who participated in a population-based case-control study in Metropolitan Detroit (USA) were included in the study comprising 26 stage I/II CRC patients, 7 CRC patients with unknown stage, 11 adenoma patients and 47 healthy controls. The use of the samples was approved by the local medical ethical committees (CMO Nijmegen/Arnhem #2006/078 and Wayne State University Human Investigation Committee #0409000504). Serum and plasma samples were stored at  $-80^{\circ}\text{C}$  until use.

*MCC-Spain study.* MCC-Spain is a multicentric multicase-control study assessing the impact of environmental exposures and their interaction with genetic factors on five cancers in Spain<sup>8,25</sup>. Individuals with histologically confirmed malignant disease were recruited between 2008 and 2013 in hospitals of 12 Spanish provinces. Controls free of these cancers were randomly sampled from general practitioner's lists at primary health centers located in the catchment areas of the hospitals where cases were recruited and frequency-matched to cases by age, sex and province. Participants completed a questionnaire through personal interview and donated a blood sample. For the present study, 576 CRC cases and 576 controls were analyzed. Population characteristics are shown in Supplementary Table S1.

The protocol of MCC-Spain was approved by the Clinical Research Ethic Committee of the Clínico San Carlos Hospital, the Clinical Research Ethic Committee of the Ramón y Cajal Hospital, the Clinical Research Ethic Committee of the Municipal Healthcare Institute of Barcelona, the Clinical Research Ethic Committee of Navarra, and the Research and Animal Well-being Ethic Committee of the Carlos III Health Institute. All the procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. All participants signed an informed consent.

**Gallo2178, Gallo2179 and FliC ELISA.** The ELISA assays for Gallo2178, Gallo2179 and FliC have been described previously<sup>6,7,21</sup>. In short, ELISA-plates were coated with FliC (InvivoGen, San Diego, CA, USA), or nickel-affinity purified recombinant expressed proteins Gallo2178 or Gallo2179 (antigen, Ag)<sup>6</sup>. For each antigen-coated well, a second well on the same plate was incubated with coating buffer without antigen (*blank*). Serum or plasma samples were added to the wells and incubated. Bound serum antibodies were detected using horseradish peroxidase (HRP)-labeled goat anti-human IgG (1:25.000; Jackson Immunoresearch). Optical density of the samples was measured in duplicate and titers were calculated as the mean  $\text{OD}_{450, \text{Ag}} - \text{OD}_{450, \text{blank}}$  and expressed as arbitrary FliC, Gallo2178 or Gallo2179 units based on reference samples from *Salmonella typhimurium*- (*S. typhimurium*), or *S. gallolyticus*-infected patients, respectively. Cut-off values for positive Gallo2179 and Gallo2178 response were set as published before with 78% specificity (Gallo2179 > 12.4 (Detroit), Gallo2179 > 9.2 (Nijmegen), Gallo2178 > 2.0 (Detroit) and Gallo2178 > 3.7 (Nijmegen)). Quartiles for FliC were calculated based on the FliC titers of the control group and set at Q1: <2.14, Q2: 2.14 – < 3.18, Q3: 3.18 – < 6.63, and Q4:  $\geq 6.63$  (Detroit) and Q1: <1.16, Q2: 1.16 – < 1.70, Q3: 1.70 – < 3.00, and Q4:  $\geq 3.00$  (Nijmegen).

**Gallo2178, Gallo2179 and FliC multiplex serology.** MCC-Spain serum samples were sent on dry ice to the German Cancer Research Center (DKFZ, Heidelberg, Germany) and analyzed in a final 1:500 dilution. Multiplex serology was performed as described previously<sup>8,26</sup>. His<sub>6</sub>-tagged recombinantly expressed and affinity-purified SGG pilus proteins Gallo2178 and Gallo2179<sup>6</sup> as well as *S. typhimurium* FliC (InvivoGen, San Diego, CA, USA) were directly cross-linked onto fluorescence-labeled polystyrene beads (SeroMap, Luminex Corp., Austin, TX, USA). Mixing of the different antigen-coupled bead sets allowed analysis of antibody responses in a serum against several antigens in one reaction. A Luminex analyser (Luminex Corp., Austin, TX, USA) distinguishes the bead sets and simultaneously quantifies the amount of bound serum antibody by a secondary antibody conjugated with a fluorescent reporter (Streptavidin-R-Phycoerythrin). The level of antibody response was given as median fluorescence intensity (MFI) of at least 100 beads per set measured.

Cut-off definition for antibody responses to SGG Gallo2178 and Gallo2179 was arbitrary as previously described<sup>8</sup> and defined 10% of controls as positive (cut-off: Gallo2178: 21 MFI; Gallo2179: 506 MFI). Antibody responses to FliC were analyzed in quartiles as defined in controls (Q1:  $\leq 159$  MFI; Q2: 159.5–379.5 MFI; Q3: 380–938.5 MFI; Q4: >938.5 MFI).

**Statistical analysis.** Nijmegen and Detroit studies were analyzed as a combined study in the main analysis to increase statistical power.

Associations between antibody responses to FliC, and Gallo2178 and Gallo2179 in the Nijmegen/Detroit study and in the MCC-Spain study were analyzed using a Chi-square test.

Associations of antibody responses to FliC, Gallo2178 and Gallo2179 with CRC were assessed using logistic regression models to compute odds ratios (OR) and 95% confidence intervals (95% CI). A p-value of <0.05 was considered statistically significant. For the Nijmegen/Detroit studies, additional sociodemographic information was available only for part of the samples and thus no further adjustment was performed. For the replication in the MCC-Spain study, multivariable logistic regression models were applied with adjustment for age, sex, province, education (low (primary school or lower), intermediate (secondary school), high (University)), BMI (kg/m<sup>2</sup>, <25, 25–<30, ≥30), smoking (never, former, current) and family history of CRC in first or second-degree relatives. In the MCC-Spain study, we further performed a stratified analysis by median FliC MFI (379.5 MFI) to address whether antibody responses to FliC affected the association of antibody positivity to SGG pilus proteins with CRC. We added an interaction term between FliC (below/above median MFI) and Gallo2178 or Gallo2179 antibody positivity and performed a likelihood ratio test to assess statistical significance.

## Data Availability

The authors confirm that the study data of the Nijmegen/Detroit studies are available as supplementary material. The study data of MCC-Spain is available only upon reasonable request to Nerea Fernández de Larrea (nfernandez@isciii.es) due to personal data safety protection regulations.

## References

1. Ferlay, J. *et al.* GLOBOCAN 2012 v1.0., Lyon, France: International Agency for Research on Cancer. *Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11* [Internet], <http://globocan.iarc.fr> (2013).
2. Brenner, H., Kloor, M. & Pox, C. P. Colorectal cancer. *Lancet* **383**, 1490–1502, [https://doi.org/10.1016/S0140-6736\(13\)61649-9](https://doi.org/10.1016/S0140-6736(13)61649-9) (2014).
3. Ahn, J. *et al.* Human gut microbiome and risk for colorectal cancer. *Journal of the National Cancer Institute* **105**, 1907–1911, <https://doi.org/10.1093/jnci/djt300> (2013).
4. Tjalsma, H., Boleij, A., Marchesi, J. R. & Dutilh, B. E. A bacterial driver-passenger model for colorectal cancer: beyond the usual suspects. *Nature reviews. Microbiology* **10**, 575–582, <https://doi.org/10.1038/nrmicro2819> (2012).
5. Abdulamir, A. S., Hafidh, R. R., Mahdi, L. K., Al-jeboori, T. & Abubaker, F. Investigation into the controversial association of *Streptococcus gallolyticus* with colorectal cancer and adenoma. *BMC cancer* **9**, 403, <https://doi.org/10.1186/1471-2407-9-403> (2009).
6. Boleij, A. *et al.* Selective antibody response to *Streptococcus gallolyticus* pilus proteins in colorectal cancer patients. *Cancer Prev Res (Phila)* **5**, 260–265, <https://doi.org/10.1158/1940-6207.CAPR-11-0321> (2012).
7. Boleij, A. *et al.* Increased exposure to bacterial antigen Rpl7/L12 in early stage colorectal cancer patients. *Cancer* **116**, 4014–4022, <https://doi.org/10.1002/cncr.25212> (2010).
8. Butt, J. *et al.* Association of *Streptococcus gallolyticus* subspecies *gallolyticus* with colorectal cancer: Serological evidence. *International journal of cancer. Journal international du cancer* **138**, 1670–1679, <https://doi.org/10.1002/ijc.29914> (2016).
9. Butt, J. *et al.* Serology of *Streptococcus gallolyticus* subspecies *gallolyticus* and its association with colorectal cancer and precursors. *International journal of cancer. Journal international du cancer. https://doi.org/10.1002/ijc.30765* (2017).
10. Garza-Gonzalez, E. *et al.* Immune response against *Streptococcus gallolyticus* in patients with adenomatous polyps in colon. *International journal of cancer. Journal international du cancer* **131**, 2294–2299, <https://doi.org/10.1002/ijc.27511> (2012).
11. Kumar, R. *et al.* *Streptococcus gallolyticus* subsp. *gallolyticus* promotes colorectal tumor development. *PLoS pathogens* **13**, e1006440, <https://doi.org/10.1371/journal.ppat.1006440> (2017).
12. Butt, J. *et al.* Prospective evaluation of antibody response to *Streptococcus gallolyticus* and risk of colorectal cancer. *International journal of cancer. Journal international du cancer. https://doi.org/10.1002/ijc.31283* (2018).
13. Kumar, R., Herold, J. L., Taylor, J., Xu, J. & Xu, Y. Variations among *Streptococcus gallolyticus* subsp. *gallolyticus* strains in connection with colorectal cancer. *Scientific reports* **8**, 1514, <https://doi.org/10.1038/s41598-018-19941-7> (2018).
14. Aymeric, L. *et al.* Colorectal cancer specific conditions promote *Streptococcus gallolyticus* gut colonization. *Proceedings of the National Academy of Sciences of the United States of America* **115**, E283–E291, <https://doi.org/10.1073/pnas.1715112115> (2018).
15. Boleij, A. *et al.* Novel clues on the specific association of *Streptococcus gallolyticus* subsp. *gallolyticus* with colorectal cancer. *J Infect Dis* **203**, 1101–1109, <https://doi.org/10.1093/infdis/jiq169> (2011).
16. Danne, C. *et al.* Molecular characterization of a *Streptococcus gallolyticus* genomic island encoding a pilus involved in endocarditis. *J Infect Dis* **204**, 1960–1970, <https://doi.org/10.1093/infdis/jir666> (2011).
17. Sanders, C. J., Yu, Y., Moore, D. A. 3rd, Williams, I. R. & Gewirtz, A. T. Humoral immune response to flagellin requires T cells and activation of innate immunity. *J Immunol* **177**, 2810–2818 (2006).
18. Lodes, M. J. *et al.* Bacterial flagellin is a dominant antigen in Crohn disease. *The Journal of clinical investigation* **113**, 1296–1306, <https://doi.org/10.1172/JCI20295> (2004).
19. Kong, S. Y. *et al.* Serum Endotoxins and Flagellin and Risk of Colorectal Cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC) Cohort. *Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **25**, 291–301, <https://doi.org/10.1158/1055-9965.EPI-15-0798> (2016).
20. Kang, M. *et al.* Association of plasma endotoxin, inflammatory cytokines and risk of colorectal adenomas. *BMC cancer* **13**, 91, <https://doi.org/10.1186/1471-2407-13-91> (2013).
21. Kato, I. *et al.* Partial associations of dietary iron, smoking and intestinal bacteria with colorectal cancer risk. *Nutrition and cancer* **65**, 169–177, <https://doi.org/10.1080/01635581.2013.748922> (2013).
22. Ramos, H. C., Rumbo, M. & Sirard, J. C. Bacterial flagellins: mediators of pathogenicity and host immune responses in mucosa. *Trends in microbiology* **12**, 509–517, <https://doi.org/10.1016/j.tim.2004.09.002> (2004).
23. Pasquereau-Kotula, E., Martins, M., Aymeric, L. & Dramsi, S. Significance of *Streptococcus gallolyticus* subsp. *gallolyticus* Association With Colorectal Cancer. *Frontiers in microbiology* **9**, 614, <https://doi.org/10.3389/fmicb.2018.00614> (2018).
24. Jans, C. & Boleij, A. The Road to Infection: Host-Microbe Interactions Defining the Pathogenicity of *Streptococcus bovis*/*Streptococcus equinus* Complex Members. *Front. Microbiol.* **9**, 603, <https://doi.org/10.3389/fmicb.2018.00603> (2018).
25. Castano-Vinyals, G. *et al.* Population-based multicase-control study in common tumors in Spain (MCC-Spain): rationale and study design. *Gaceta sanitaria / S.E.S.P.A.S* **29**, 308–315, <https://doi.org/10.1016/j.gaceta.2014.12.003> (2015).
26. Waterboer, T. *et al.* Multiplex human papillomavirus serology based on *in situ*-purified glutathione s-transferase fusion proteins. *Clinical chemistry* **51**, 1845–1853, <https://doi.org/10.1373/clinchem.2005.052381> (2005).

## Acknowledgements

We thank Shaynoor Dramsi and Camille Danne (Department of Microbiology, Institut Pasteur, Paris, France) for providing the purified proteins Gallo2178 and Gallo2179. This work was supported by the National Institutes of Health (grant number R01-CA93817, Ikuko Kato); The Netherlands Organization for Scientific Research (NWO 016.166.089 to A.B.); the “Acción Transversal del Cáncer”, approved on the Spanish Ministry Council on the 11th October 2007; the Instituto de Salud Carlos III-FEDER funds –a way to build Europe- (grants numbers PI08/1770, PI08/0533, PI08/1359, PS09/00773, PS09/01662, PS09/01286, PS09/01903, PS09/02078, PI11/00226, PI11/01403, PI11/01810, PI11/01889, PI11/02213, PI12/00150, PI12/00265, PI12/00488, PI12/00715, PI12/01270, PI14/00613, PI17/00092); the Fundación Marques de Valdecilla (grant number API 10/09); Obra Social CAJASTUR (grant number SV-CAJASTUR-1); Recercaixa (grant number 2010ACUP 00310); Spanish Association Against Cancer (AECC) Scientific Foundation; Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR) –Generalitat de Catalunya (Catalonian Government) (grant numbers 2009SGR1026, 2009SGR1465 and 2017SGR723); and Junta de Castilla y León (grant number LE022A10-2). Sample collection and storage was partially supported by the Instituto de Salud Carlos III-FEDER (grant number RD09/0076/00036), Xarxa de Bancs de Tumors de Catalunya sponsored by Pla Director d'Oncologia de Catalunya (XBTC).

## Author Contributions

J.B., A.B., N.A. and N.F. are responsible for the conception of the study, statistical analysis, interpretation of results and drafting the manuscript. V.M., B.P., V.M.o., T.D., J.C., G.F., P.A., D.S., J.A., J.J., J.H., S.S., R.C., M.K., M.P., N.A. are the principal investigators of the MCC-Spain study; I.K. is the principal investigator of the Detroit cohort; H.T. and A.B. are the principal investigators of the Nijmegen cohort; and responsible for the conception of the respective study and data collection. A.B., H.T. and R.R. were responsible for acquisition and analysis of ELISA data. J.B., M.P. and T.W. were responsible for acquisition and analysis of multiplex serology data. All authors have seen and approved the final version of the manuscript. J.B. is the guarantor.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-019-47347-6>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019