A pathogenicity locus of *Streptococcus gallolyticus* subspecies *gallolyticus*

John Culver Taylor[1], Ritesh Kumar[1#], Juan Xu[1], and Yi Xu[1,2,3*]

[1]Center for Infectious and Inflammatory Diseases, Texas A&M Health Science Center Institute of Biosciences of Technology, Houston, Texas, United States of America

[2]Department of Microbiology and Molecular Genetics, McGovern Medical School, UT Health, Houston, Texas, United States of America

[3]Department of Microbial Pathogenesis and Immunology, College of Medicine, Texas A&M Health Science Center, Texas, United States of America

[#]Current affiliation: IFF Health and Biosciences

*Corresponding author: Yi Xu

E-mail: yi-xu1@tamu.edu

Key words: Colorectal cancer, colon cancer, microbiome, *Streptococcus gallolyticus*, *Streptococcus bovis*, cell proliferation, colonization.

Abstract

Streptococcus gallolyticus subspecies *gallolyticus* (*Sgg*) is known to be strongly associated with colorectal cancer (CRC). Recent functional studies further demonstrated that *Sgg* actively stimulates CRC cell proliferation and promotes the development of colon tumors. However, the *Sgg* factors important for the pro-proliferative and pro-tumor activities of *Sgg* remain unclear. Here, we identified a chromosomal locus in *Sgg* strain TX20005. Deletion of this locus significantly reduced *Sgg* adherence to CRC cells and abrogated the ability of *Sgg* to stimulate CRC cell proliferation. Thus, we designate this locus as the *Sgg* pathogenicity-associated region (*SPAR*). More importantly, we found that *SPAR* is important for *Sgg* pathogenicity *in vivo*. In a gut colonization model, mice exposed to the *SPAR* deletion mutant showed significantly reduced *Sgg* load in the colonic tissues and fecal materials, suggesting that *SPAR* contributes to the colonization capacity of *Sgg*. In a mouse model of CRC, deletion of *SPAR* abolished the ability of *Sgg* to promote the development of colon tumors growth. Taken together, these results highlight *SPAR* as a critical pathogenicity determinant of *Sgg*.

30 **Introduction**

31 *Streptococcus gallolyticus* subspecies *gallolyticus* (*Sgg*) is a group D streptococcal
32 bacterium, and an opportunistic pathogen belonging to the *Streptococcus bovis*/*Streptococcus*
33 *equinis* complex (SBSEC) (1). *Sgg* causes bacteremia and infective endocarditis (IE), and has a
34 long-standing clinical association with CRC, supported by extensive studies and case reports
35 dating back to the 1950s (2-15). Patients with *Sgg* IE and/or bacteremia have concomitant colon
36 adenomas or adenocarcinomas at a much higher rate (~60%) than that of the general population
37 (8). Moreover, a prospective study found that 45% of patients with *Sgg* IE developed colonic
38 neoplastic lesions within 5 years of IE diagnosis compared to 21% of patients with IE due to
39 closely-related enterococci, further supporting the association of *Sgg* with CRC (6). In CRC
40 patients with no symptoms of IE or bacteremia, *Sgg* was found to preferentially associate with
41 tumor tissues compared to normal adjacent tissues or normal tissues in multiple studies around
42 the world (4, 16-19).

43 Functionally, *Sgg* stimulates the proliferation of human CRC cells in a β-catenin
44 dependent manner *in vitro* and in a xenograft model (4). Exposure to *Sgg* also led to larger and
45 more advanced tumors in an azoxymethane (AOM)-induced CRC model and in a colitis-
46 associated CRC model (4, 15). These results indicate that *Sgg* possesses pro-proliferation and
47 pro-tumor activities. Previous work also indicated that there are variations among *Sgg* strains in
48 their ability to stimulate human CRC cell proliferation. Some *Sgg* strains such as ATCC43143
49 showed significantly reduced capacity to adhere to HT29 cells, to stimulate the proliferation of
50 these cells and to promote tumor development *in vivo* (2), suggesting that there are specific *Sgg*
51 factors contributing to the pro-proliferative and pro-tumor activities of *Sgg*. The identity of these
52 specific *Sgg* factors was unknown.

53 Here, we describe work characterizing a chromosomal locus of *Sgg* strain TX20005.
54 Deletion of this locus significantly impaired the ability of *Sgg* to adhere to cultured CRC cells
55 and abrogated the ability of *Sgg* to stimulate CRC cell proliferation. *In vivo*, the deletion mutant
56 displayed significantly reduced capacity to colonize normal mouse colons. More importantly, the
57 mutant lost the ability to promote the development of colon tumors. Given the role of this locus
58 in *Sgg* pathogenesis, we have coined this locus as the *Sgg* pathogenicity-associated region
59 (*SPAR*).
60

61 **Materials and Methods**

62 **Bacterial strains, cell lines and growth conditions.** *Sgg* strains were routinely grown in
63 brain heart infusion (BHI) broth or tryptic soy broth at 37°C with shaking, or on BHI or TSB
64 agar plates at 37°C overnight (Teknova). For co-culture experiments and animal studies,
65 stationary phase bacterial cultures were pelleted, washed with phosphate-buffered saline (PBS),
66 pH 7.4, resuspended in PBS containing 15% glycerol, aliquoted and stored at -80°C. Aliquoted
67 stocks were thawed, washed with PBS, diluted in appropriate media to obtain the desired
68 concentration and directly added to cells or administered to mice. Human colon cancer cell lines
69 HT29, HCT116, and SW480, as well as the HEK293 cell line were grown in Dulbecco's
70 Modified Eagle's Medium F-12 50/50 (DMEM/F-12, GIBCO) supplemented with 10% FBS at
71 37°C with 5% CO_2 in a humidified chamber. Cells from less than 30 passages were used in the
72 experiments.

73 **Preparation of culture supernatants (CSs).** Supernatants from stationary phase *Sgg*
74 cultured in BHI were filtered-sterilized through a 0.2μm filter and concentrated approximately
75 10 – 20-fold using centrifugal concentrators (3kD molecular weight cut off). The concentrated

76 CSs were aliquoted and stored at -80°C. A vial is thawed and diluted in the appropriate media for
77 adherence or cell proliferation assays and used immediately.

78 **Adherence assay.** This was performed as described previously (3). Briefly, HT29 cells
79 were seeded at a density of $1x10^6$ cells/well in a 24 well plate and allowed to attach overnight.
80 Cells were incubated with bacteria at a multiplicity of infection (MOI) of 10 in the absence or
81 presence of CSs (diluted to 1x) for 1 hour at 37°C with 5% CO_2. Cells were washed thrice in
82 PBS to remove unattached bacteria, lysed with 1mL of 0.01% Triton X-100 (Sigma) and dilution
83 plated onto BHI or TSB agar plates. Adherence was calculated as the percentage of adhered
84 bacteria vs. total bacteria added.

85 **Cell proliferation assay.** This was performed as described previously (3) with slight
86 modifications. Briefly, cells were seeded in 96 well plates at a concentration of ~$1x10^4$ cells/well
87 and incubated overnight. Cells were then incubated in fresh DMEM/F-12 containing 10% FBS in
88 the absence or presence of *Sgg* bacteria (MOI = 1) or *Sgg* CSs (1x) for a total of 24 hours.
89 Trimethoprim was added after 6 hours of incubation (1 μg/mL final concentration) to inhibit
90 bacterial growth, as previously described. The number of viable cells was determined using the
91 cell counting kit (CCK)-8 kit following the instructions of the supplier (Apex Bio).

92 **Western Blot**. Detection of β-catenin and PCNA was carried out as described previously
93 (3). Briefly, HT29 cells were seeded at a density of $1x10^6$ cells per well in 6-well plates and
94 incubated overnight. The cells were then incubated with media only or media containing bacteria
95 (MOI=1) for 9 hours. Total cell lysates were subjected to SDS-PAGE, transferred, and probed
96 with antibodies against β-catenin (1:1000), PCNA (1:1000), and β-actin (1:1000) (Cell Signaling
97 Technology), followed by incubation with HRP-conjugated secondary antibodies (1:3000). Band
98 intensities were measured using FIJI ImageJ and normalized to those of β-actin.

99 **Deletion of *SPAR***. This was performed following a procedure described previously (3).
100 Briefly, the ~ 1kb region upstream of *sparA* and the ~1kb region downstream of *sparL* were
101 synthesized by Genscript and cloned into pUC57 (Genscript). The insert was then subcloned into
102 a temperature sensitive conjugative plasmid pG1-$oriT_{TnGBS1}$. The insert sequence was verified by
103 Sanger sequencing. The construct was introduced into *S. agalactiae* NEM316 by electroporation
104 and then into *Sgg* strain TX20005 by conjugation under the permissive temperature. PCR was
105 used to screen for double cross-over deletion mutants. Deletion was further confirmed by PCR
106 amplification of the regions spanning the deleted fragment and DNA sequencing of the PCR
107 product.
108 **RNA extraction, cDNA synthesis and PCR.** This was performed following the method
109 described previously (3). Briefly, RNA was extracted from TX20005 cultured in the presence of
110 HT29, as well as from the colonic tissues of mice that had been orally gavaged with TX20005.
111 RNA was treated with DNase and synthesized using a ProtoScript II First Strand cDNA
112 Synthesis Kit (NEB). Primers used in PCR amplification are: *sparA*, forward
113 5'GCAAGCTGGTCGAACAGAAC and reverse 5'GCTTCTATGGTTGGGGCTAGA; *sparD*,
114 forward 5' GGAGGTGGATCCAACAAGGG and reverse 5'CAGGTTCCTCGATAGCCAGC;
115 and *sparG*, forward 5'TCAGTTGTTAGCGGATGCGT and reverse 5'
116 CCCTTTATTGCTTGTGCTCCC.
117 **Growth curves**. Overnight cultures of *Sgg* were inoculated into fresh BHI broth at 1:100
118 dilution and grown at 37°C with shaking. Samples were taken at 0, 3, 6, 9, 12, and 24 hours,
119 dilution plated onto agar plates, incubated for 24 hours and colonies enumerated.
120 **Animal experiments**. Animal studies were performed in accordance with protocols
121 approved by the Institutional Animal Care and Use Committee at the Texas A&M Health

122 Science Center, Institute of Biosciences and Technology. Mice were fed with standard ProLab
123 IsoPro RMH3000 (LabDiet). **1) Colonization.** This was performed as previously described (3)
124 with slight modifications. Briefly, 6-week-old A/J mice, sex matched (Jackson Laboratory), were
125 treated with ampicillin at a concentration of 1 g/L in drinking water for 3 days and switched to
126 antibiotic-free water 24 hours prior to the administration of bacteria. *Sgg* was orally gavaged at a
127 dose of ~1 x 10^9 CFU/mouse. Colons and fecal materials were collected at day 1, 3, and 7 post-
128 gavage. Samples were weighed, homogenized in sterile PBS in a TissueLyser (Qiagen), dilution
129 plated onto Enterococcus Selective Agar (ESA) plates and incubated at 37°C for 24–48 hours to
130 enumerate *Sgg* colonies. **2) AOM-induced model of CRC.** This was performed as previously
131 described (3) with slight modifications. Briefly, 6-week-old A/J mice, sex matched were treated
132 with 4 weekly i.p. injections of AOM (10 mg/kg body weight), followed by ampicillin in
133 drinking water (1 g/L) for 7 days and switched to antibiotic-free water 24 hours prior to the first
134 oral gavage with *Sgg*. Mice were orally gavaged with bacteria at ~ 1x10^9 CFU/mouse or saline (n
135 = 11-12 per group) once per week for 12 weeks. Mice were euthanized one week after the last
136 oral gavage by CO_2 inhalation followed by cervical dislocation. Colon and fecal pellets were
137 collected. The number and size of macroscopic were recorded. Visual evaluation of colons was
138 carried out by a blinded observer. A random subset of fecal pellets were weighed, homogenized
139 in sterile PBS and dilution plated onto ESA plates.

140 **Statistical analysis**. GraphPad Prism 9 was used for statistical analyses. Two-tailed
141 unpaired *t*-test was used for pairwise comparisons to assess the significance of differences
142 between two groups in cell proliferation assays, western blot analysis, and adherence assays. The
143 non-parametric Mann-Whitney test was used to assess the significance of differences of results

144 between groups in animal studies. Two-way ANOVA was used to analyze the bacterial growth
145 curves. Ns, not significant; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$.
146 **Ethics statement.** Animal studies were performed in accordance with protocols
147 (IACUC#2017-0420-IBT) approved by the Institutional Animal Care and Use Committee at the
148 Texas A&M Health Science Center, Institute of Biosciences and Technology. The Texas A&M
149 University Health Science Center—Institute of Biosciences and Technology is registered with the
150 Office of Laboratory Animal Welfare per Assurance A4012-01. It is guided by the PHS Policy on
151 Human Care and Use of Laboratory Animals (Policy), as well as all applicable provisions of the
152 Animal Welfare Act.
153

154 **Results**

155 **The *Sgg SPAR* locus**. We compared the genome sequence of TX20005
156 (NZ_CP077423.1) with that of ATCC 43143 (NC_017576.1), an *Sgg* strain that is defective in
157 stimulating CRC cell proliferation or promoting the development of colon tumors, using the
158 multiple genome alignment tool MAUVE (20). The SPAR locus is one of the regions in the
159 chromosome of TX20005 that display differences from ATCC 43143. The locus is comprised of
160 12 genes (*sparA* to *sparL*) (Fig. 1A). The protein sequences encoded by these genes were
161 analyzed for secretion signals using SignalP - 6.0 (21). None of them contains typical secretion
162 signals recognized by the Sec or Tat translocon. SparA, SparK and SparL are predicted to
163 contain transmembrane helices (TMHMM2.0) (22) and are thus putative transmembrane
164 proteins. Homology search using Protein BLAST showed that SparA to C are hypothetical
165 proteins of unknown function (Table 1). SparD to SparF exhibit homology to predicted ATP-
166 dependent endonuclease of the OLD family, GIY-YIG endonuclease/PcrA/UvrD helicase, and

167 GntR family transcriptional regulator, respectively. SparG to SparJ display features of effectors
168 secreted by the type VII secretion system (T7SS). SparG, H and J are homologous to T7SS
169 effectors, LXG domain-containing protein and T7SS effector LXG polymorphic toxin. SparG, H
170 and I are of 91, 102 and 137 amino acids in length, in keeping with T7SS effectors belonging to
171 the WXG100 family. Analysis using HHpred (23) showed that SparG and SparI fold into a four-
172 helical bundle structure typical of ESAT-6-like WXG100 proteins, subfamily of T7SS effectors,
173 with a probability of 96.48% and 97.97%, respectively. Spar H, I and J also contain a conserved
174 sequence motif HxxxD/ExxhxxxH (H denotes highly conserved and h less conserved
175 hydrophobic residues) that is considered to be important for the secretion of WXG100 proteins
176 by T7SS (24) (Supplemental Fig. S1). Immediately downstream of the *sparJ* gene are two genes
177 that encode putative TipC family immunity proteins. Genes encoding T7SS effector LXG
178 polymorphic toxins are typically clustered together with genes encoding their cognate immunity
179 proteins (25-27). The arrangement of *sparJ*, *K* and *L* further suggests that SparJ is a putative
180 T7SS effector LXG polymorphic toxin and SparK and L are likely the cognate immunity
181 proteins.

182 We next tested if the *SPAR* genes were expressed *in vitro* and *in vivo* by RT-PCR (Fig.
183 1B). We focused on the expression of *sparA*, *D* and *G*, under conditions that are relevant to the
184 pro-proliferation and pro-tumor activity of *Sgg*. All three genes were expressed when the
185 bacteria were co-cultured with HT29 cells and in the colonic tissues collected from mice orally
186 gavaged with TX20005 (Fig. 1B).

187 ***SPAR* is important for *Sgg* adherence.** To investigate the role of the *SPAR* locus in the
188 pathogenic potential of *Sgg*, we generated a mutant in the TX20005 background, in which all 12
189 genes were deleted (TX20005Δ*SPAR*). We verified that the deletion does not affect bacterial

190 growth (Fig. 2A). We sought to determine if deletion of the *SPAR* locus had any effect on the
191 adherence capacity of *Sgg*. The results showed that the mutant strain exhibited a significantly
192 reduced adherence to HT29 cells compared to the wild type (WT) parent strain TX20005 (~2%
193 vs ~9%) (Fig. 2B), suggesting that the *SPAR* locus is important for *Sgg* adherence to host cells.
194 Previous work showed that T7SS-secreted factors of *Sgg* significantly enhanced the adherence
195 capacity of the WT bacteria (3). We tested whether filter sterilized culture supernatants (CS)
196 from TX20005Δ*SPAR* can enhance the adherence capacity of TX20005. Addition of CS from the
197 WT TX20005 significantly enhanced the adherence of WT bacteria, whereas addition of the
198 mutant CS did not cause any increase in the adherence of WT bacteria (Fig. 2B, indicated by the
199 arrow), suggesting that *SPAR*-dependent secreted factors contribute to *Sgg* adherence .

200 ***SPAR* is required for *Sgg* to stimulate CRC cell proliferation.** *Sgg* has been previously
201 shown to stimulate cell proliferation in human CRC cell lines HT29 and HCT116, but not in
202 SW480 or HEK293 cell lines (2-4). We examined the ability of TX20005Δ*SPAR* to stimulate
203 cell proliferation in these cell lines. HT29, HCT116, SW480, and HEK293 cells were co-
204 cultured with TX20005 or TX20005Δ*SPAR*. As anticipated, the WT strain significantly
205 stimulated HT29 and HCT116 cell proliferation compared to untreated cells (Fig. 3A and 3B). In
206 contrast, HT29 and HCT116 cells co-cultured with the mutant did not exhibit any significant
207 increase in cell proliferation, indicating that *SPAR* is required for *Sgg*-stimulated CRC cell
208 proliferation. Neither the WT nor the mutant had any effect on the proliferation of SW480 or
209 HEK293 cells (Fig. 3C and 3D), as expected.

210 To further validate the results, we investigated the effect of the mutant on cell
211 proliferation markers β-catenin and proliferating cell nuclear antigen (PCNA) in HT29 cells (Fig.
212 3E-G). Cells co-cultured with TX20005 exhibited a significant increase in the level of both β-

213 catenin and PCNA compared to cells cultured in media only, as expected. In contrast, cells co-
214 cultured with the *SPAR* mutant showed no significant increase in the level of either marker
215 compared to cells cultured in media only, further confirming that the *SPAR* locus is essential for
216 *Sgg* to stimulate cell proliferation.

217 Previous work demonstrated that filter sterilized CSs from WT *Sgg* strains induce CRC
218 cell proliferation (3, 28). Therefore, we examined the effect of CS from TX20005Δ*SPAR* on cell
219 proliferation (Fig. 2H). HT29 cells treated with the WT CS exhibited a significant increase in
220 cell proliferation, as expected, whereas cells treated with the mutant CS did not exhibit any
221 significant increase in cell proliferation. These results suggest that the defect in the *SPAR* mutant
222 to stimulate host cell proliferation is due, at least in part, to the absence of certain secreted
223 factors in the CS. Taken together, these results indicate that *SPAR*, and particularly secreted
224 factors dependent on *SPAR*, is required for *Sgg* to stimulate CRC cell proliferation *in vitro*.

225 ***Sgg* gut colonization is impaired by *SPAR* deletion.** The ability of *Sgg* to colonize the
226 host gut is an important aspect of its pathogenic potential (2). We sought to determine if the
227 *SPAR* locus is involved in *Sgg* gut colonization. Mice were orally gavaged with WT or the *SPAR*
228 deletion mutant. Colon tissues as well as fecal pellets were collected at day 1, 3, and 7 post-
229 bacterial gavage to determine the bacterial load in the samples. At day 1, the *Sgg* bacterial load
230 in the colonic tissues from mice gavage with the WT and the mutant did not significantly differ.
231 At day 3 and 7, the bacterial load of the mutant was significantly decreased compared to that of
232 the WT (Fig. 4A). In the fecal pellets, a similar patten was observed, such that at day 1, the
233 bacterial load of mutant and WT strains did not significantly differ, while At days 3 and 7, the
234 bacterial load of the mutant was decreased significantly compared to that of the WT (Fig. 4B).

235 Taken together, these results indicate that the *SPAR* locus contributes to the colonization capacity
236 of *Sgg* in the normal colon.

237 ***SPAR* is essential for the ability of *Sgg* to promote the development of colon tumors**.
238 Next, we investigated if *SPAR* is important for *Sgg* to promote the development of colon tumors
239 using an AOM-induced CRC model (Fig. 5A). Mice gavaged with the WT strain exhibited a
240 significant increase in the colon tumor burden (Fig. 5B) compared to saline-gavaged mice,
241 consistent with previous results (3, 4). In contrast, mice exposed to the *SPAR* mutant showed no
242 increase in the tumor burden compared to the saline control and a significant reduction compared
243 to the WT-treated mice, suggesting that *SPAR* is important for *Sgg* to promote colon tumors. We
244 also determined the bacterial load in the fecal pellets isolated at the end of the experiment and
245 found no difference in the *Sgg* burden between WT or mutant-treated groups (Fig. 5C). We note
246 that this result could be due to the repeated gavages with bacteria during the experiment and does
247 not necessarily reflect the ability of the strains to colonize tumor-bearing colons in this model.
248 These results together suggest that *SPAR* plays a critical functional role in promoting the
249 development of colon tumors.

250 **Discussion**

251 In this study, we describe an *Sgg* chromosomal locus, *SPAR*, that plays a key role in *Sgg*
252 pathogenesis. Deletion of *SPAR* led to striking phenotypes *in vitro* and *in vivo*, including a
253 reduced capacity to adhere to CRC cells, decreased colonization of the normal colon, and a
254 complete loss of the ability to stimulate CRC cell proliferation *in vitro* and to promote the
255 development of colon tumors *in vivo*. These results highlight *SPAR* as a critical pathogenicity
256 determinant of *Sgg*.

257 Our data showed that deletion of the *SPAR* locus resulted in significantly reduced
258 adherence of *Sgg* to HT29 cells. Previous studies showed that T7SS-secreted factors enhance *Sgg*
259 adherence (3). Interestingly, our results revealed that deletion of *SPAR* eliminated the adherence-
260 enhancing activity in the CS, suggesting that the deletion might have affected the T7-secreted
261 factors involved in adherence. In terms of cell proliferation, we showed that the *SPAR* locus is
262 essential for *Sgg* to stimulate certain CRC cell proliferation. Furthermore, the results indicate that
263 *SPAR* is also important for the production of secreted factors responsible for stimulating host cell
264 proliferation. In this aspect, the *SPAR* mutant exhibits a very similar phenotype as that described
265 for a T7SS defective mutant of *Sgg* (TX20005*Δesx*) with respect to adherence and stimulation of
266 cell proliferation (3). One possible reason for the phenotypic similarity is that the *SPAR* locus
267 encodes several putative T7SS effectors (SparG to SparJ). Examination of the TX20005 genome
268 reveals that the previously reported T7SS locus (*Sgg*T7SS^{T05}) (3) is the only locus that encodes a
269 set of proteins comprising the T7SS secretion apparatus. Thus, SparG to SparJ are likely secreted
270 by *Sgg*T7SS^{T05}. This would imply that one or more of the *SPAR*-encoded putative T7SS effectors
271 are involved in the adherence and the pro-proliferative activity of *Sgg*. On the other hand, the
272 *sparF* gene is predicted to encode a GntR family transcriptional regulator (29). A recent

273 publication indicated that a GntR family transcriptional regulator (OG1RF_11099) of
274 *Enterococcus faecalis* controls the expression of T7SS genes (30). Using Global Align (NCBI),
275 we found that SparF is highly homologous to OG1RF_11099, showing an overall 74% similarity
276 at the amino acid sequence level. Hence, a second possible reason for the phenotypic similarity
277 between TX20005*Δspar* and TX20005*Δesx* is that SparF regulates the expression of genes in the
278 *Sgg*T7SS^{T05} locus.

279 In a gut colonization model, we observed that deletion of *SPAR* resulted in reduced *Sgg*
280 bacterial load in the normal colon. This could be due to the involvement of *SPAR* in *Sgg*
281 adherence to the colonic epithelium. Additionally, SparJ is a putative T7SS-secreted
282 polymorphic toxin. This family of toxins are often involved in interbacterial competition (26).
283 Thus, SPAR may contribute to *in vivo* colonization by antagonizing other gut commensal
284 bacteria through SparJ.

285 In the AOM model of CRC, we observed a striking difference between the *SPAR* mutant
286 and the WT bacteria in promoting the development of colon tumors, such that the mutant has
287 completely lost the ability to promote the development of colon tumors. This phenotype is much
288 stronger than that previously reported for the T7SS mutant (3), raising the possibility that in
289 addition to T7SS genes, SparF also regulates the expression of other genes that contribute tumor
290 development *in vivo*. Studies to further investigate the biological activities of SparF, and SparG
291 to SparL are needed to delineate the specific contribution of the *SPAR* locus to the observed
292 phenotypes.

293 It is also possible that other genes within the *SPAR* locus, SparA-SparE, also play a role,
294 however, their predicted function based on homology search does not provide clues regarding
295 their respective contribution. SparA-C are putative hypothetical proteins. SparD is homologous

296 to OLD family endonucleases. OLD family endonucleases are widely present among bacteria
297 and archaea (31). While their biological function is not completely understood, they contain a C-
298 terminal ATPase domain, as well as an N-terminal Toprim domain, which is believed to be
299 important in DNA replication, recombination, and repair (32, 33). SparE is homologous to GIY-
300 YIG endonuclease (34-38) and PcrA/UvrD helicase. The PcrA/UvrD helicase is an essential
301 helicase in gram-positive bacteria and has been shown to play a role in DNA repair, as well as
302 the replication of small drug-resistance plasmids in *Staphylococcus aureus* and *Bacillus subtilis*
303 *(39-45)*.

304 In summary, we have identified a chromosomal locus in *Sgg*, *SPAR*, that is critical to the
305 pathogenicity of *Sgg*. We report that deletion of *SPAR* significantly reduces the capacity of *Sgg*
306 to adhere to CRC cells and to colonize the gut. Furthermore, deletion of *SPAR* abrogates the
307 ability for *Sgg* to stimulate CRC cell proliferation and to accelerate colon tumor development.
308 Examination of genes with the locus highlighted several potential candidates responsible for the
309 observed phenotypes. The results suggest a connection between the *SPAR* locus and the T7SS of
310 *Sgg* in terms of additional T7SS effectors encoded by *SPAR* genes and a potential regulator of
311 T7SS expression. Further investigations into the activities of specific *SPAR* proteins will be
312 important for dissecting the intricate *Sgg* pathogenic mechanisms and will open up the path
313 ahead to identify biomarker candidates and targets for clinical prevention and intervention
314 strategies.

315

316 **Acknowledgements**

317 This study was supported by funding from the Hamill Foundation, the Cancer Prevention

318 Research Institute of Texas (RP170653) and NIAID (1R21AI151914) to Y. Xu and

319 (T32AI055449-16) to J. Taylor. The content of this manuscript is also available at bioRxiv.org

320 (https://doi.org/10.1101/2022.03.15.484266).

321

322 **References**

323 1. Schlegel L, Grimont F, Ageron E, Grimont PAD, Bouvet A. Reappraisal of the taxonomy
324 of the Streptococcus bovis/Streptococcus equinus complex and related species: description of
325 Streptococcus gallolyticus subsp. gallolyticus subsp. nov., S. gallolyticus subsp. macedonicus
326 subsp. nov. and S. gallolyticus subsp. pasteurianus subsp. nov. Int J Syst Evol Microbiol.
327 2003;53(Pt 3):631-45.
328 2. Kumar R, Herold JL, Taylor J, Xu J, Xu Y. Variations among Streptococcus gallolyticus
329 subsp. gallolyticus strains in connection with colorectal cancer. Scientific reports.
330 2018;8(1):1514.
331 3. Taylor JC, Gao X, Xu J, Holder M, Petrosino J, Kumar R, et al. A type VII secretion
332 system of Streptococcus gallolyticus subsp. gallolyticus contributes to gut colonization and the
333 development of colon tumors. PLoS Pathog. 2021;17(1):e1009182.
334 4. Kumar R, Herold JL, Schady D, Davis J, Kopetz S, Martinez-Moczygemba M, et al.
335 Streptococcus gallolyticus subsp. gallolyticus promotes colorectal tumor development. PLoS
336 Pathog. 2017;13(7):e1006440.
337 5. Marmolin ES, Hartmeyer GN, Christensen JJ, Nielsen XC, Dargis R, Skov MN, et al.
338 Bacteremia with the bovis group streptococci: species identification and association with
339 infective endocarditis and with gastrointestinal disease. Diagn Microbiol Infect Dis.
340 2016;85(2):239-42.
341 6. Corredoira J, Garcia-Pais MJ, Coira A, Rabunal R, Garcia-Garrote F, Pita J, et al.
342 Differences between endocarditis caused by Streptococcus bovis and Enterococcus spp. and their
343 association with colorectal cancer. Eur J Clin Microbiol Infect Dis. 2015;34(8):1657-65.

344 7. Boleij A, Tjalsma H. Gut bacteria in health and disease: a survey on the interface
345 between intestinal microbiology and colorectal cancer. Biol Rev Camb Philos Soc.
346 2012;87(3):701-30.
347 8. Boleij A, van Gelder MM, Swinkels DW, Tjalsma H. Clinical Importance of
348 Streptococcus gallolyticus infection among colorectal cancer patients: systematic review and
349 meta-analysis. Clin Infect Dis. 2011;53(9):870-8.
350 9. Abdulamir AS, Hafidh RR, Abu Bakar F. The association of Streptococcus
351 bovis/gallolyticus with colorectal tumors: the nature and the underlying mechanisms of its
352 etiological role. J Exp Clin Cancer Res. 2011;30:11.
353 10. Gupta A, Madani R, Mukhtar H. Streptococcus bovis endocarditis, a silent sign for
354 colonic tumour. Colorectal Dis. 2010;12(3):164-71.
355 11. Waisberg J, Matheus CeO, Pimenta J. Infectious endocarditis from Streptococcus bovis
356 associated with colonic carcinoma: case report and literature review. Arq Gastroenterol.
357 2002;39(3):177-80.
358 12. Alazmi W, Bustamante M, O'Loughlin C, Gonzalez J, Raskin JB. The association of
359 Streptococcus bovis bacteremia and gastrointestinal diseases: a retrospective analysis. Dig Dis
360 Sci. 2006;51(4):732-6.
361 13. Gold JS, Bayar S, Salem RR. Association of Streptococcus bovis bacteremia with colonic
362 neoplasia and extracolonic malignancy. Arch Surg. 2004;139(7):760-5.
363 14. Corredoira J, Grau I, Garcia-Rodriguez JF, Alonso-Garcia P, Garcia-Pais MJ, Rabunal R,
364 et al. The clinical epidemiology and malignancies associated with Streptococcus bovis biotypes
365 in 506 cases of bloodstream infections. J Infect. 2015;71(3):317-25.

366 15. Zhang Y, Weng Y, Gan H, Zhao X, Zhi F. Streptococcus gallolyticus conspires myeloid
367 cells to promote tumorigenesis of inflammatory bowel disease. Biochem Biophys Res Commun.
368 2018;506(4):907-11.
369 16. Thomas AM, Manghi P, Asnicar F, Pasolli E, Armanini F, Zolfo M, et al. Metagenomic
370 analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and
371 a link with choline degradation. Nat Med. 2019;25(4):667-78.
372 17. Perichon B, Lichtl-Hafele J, Bergsten E, Delage V, Trieu-Cuot P, Sansonetti P, et al.
373 Detection of Streptococcus gallolyticus and Four Other CRC-Associated Bacteria in Patient
374 Stools Reveals a Potential "Driver" Role for Enterotoxigenic Bacteroides fragilis. Frontiers in
375 cellular and infection microbiology. 2022;12:794391.
376 18. Kwong TNY, Wang X, Nakatsu G, Chow TC, Tipoe T, Dai RZW, et al. Association
377 Between Bacteremia From Specific Microbes and Subsequent Diagnosis of Colorectal Cancer.
378 Gastroenterology. 2018;155(2):383-90 e8.
379 19. Abdulamir AS, Hafidh RR, Bakar FA. The association of Streptococcus
380 bovis/gallolyticus with colorectal tumors: The nature and the underlying mechanisms of its
381 etiological role. J Exp Clin Cancer Res. 2011;30:11.
382 20. Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved
383 genomic sequence with rearrangements. Genome Res. 2004;14(7):1394-403.
384 21. Teufel F, Almagro Armenteros JJ, Johansen AR, Gislason MH, Pihl SI, Tsirigos KD, et
385 al. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat
386 Biotechnol. 2022.

387 22. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein
388 topology with a hidden Markov model: application to complete genomes. J Mol Biol.
389 2001;305(3):567-80.
390 23. Hildebrand A, Remmert M, Biegert A, Soding J. Fast and accurate automatic structure
391 prediction with HHpred. Proteins. 2009;77 Suppl 9:128-32.
392 24. Poulsen C, Panjikar S, Holton SJ, Wilmanns M, Song YH. WXG100 protein superfamily
393 consists of three subfamilies and exhibits an α-helical C-terminal conserved residue pattern.
394 PLoS One. 2014;9(2):e89313.
395 25. Zhang D, de Souza RF, Anantharaman V, Iyer LM, Aravind L. Polymorphic toxin
396 systems: Comprehensive characterization of trafficking modes, processing, mechanisms of
397 action, immunity and ecology using comparative genomics. Biol Direct. 2012;7:18.
398 26. Whitney JC, Peterson SB, Kim J, Pazos M, Verster AJ, Radey MC, et al. A broadly
399 distributed toxin family mediates contact-dependent antagonism between gram-positive bacteria.
400 Elife. 2017;6.
401 27. Klein TA, Pazos M, Surette MG, Vollmer W, Whitney JC. Molecular Basis for Immunity
402 Protein Recognition of a Type VII Secretion System Exported Antibacterial Toxin. J Mol Biol.
403 2018;430(21):4344-58.
404 28. Taddese R, Garza DR, Ruiter LN, de Jonge MI, Belzer C, Aalvink S, et al. Growth rate
405 alterations of human colorectal cancer cells by 157 gut bacteria. Gut Microbes. 2020;12(1):1-20.
406 29. Liu GF, Wang XX, Su HZ, Lu GT. Progress on the GntR family transcription regulators
407 in bacteria. Yi Chuan. 2021;43(1):66-73.

408 30. Chatterjee A, Willett JLE, Dunny GM, Duerkop BA. Phage infection and sub-lethal
409 antibiotic exposure mediate Enterococcus faecalis type VII secretion system dependent inhibition
410 of bystander bacteria. PLoS Genet. 2021;17(1):e1009204.
411 31. Schiltz CJ, Lee A, Partlow EA, Hosford CJ, Chappie JS. Structural characterization of
412 Class 2 OLD family nucleases supports a two-metal catalysis mechanism for cleavage. Nucleic
413 Acids Res. 2019;47(17):9448-63.
414 32. Aravind L, Leipe DD, Koonin EV. Toprim--a conserved catalytic domain in type IA and
415 II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. Nucleic
416 Acids Res. 1998;26(18):4205-13.
417 33. Koonin EV, Gorbalenya AE. The superfamily of UvrA-related ATPases includes three
418 more subunits of putative ATP-dependent nucleases. Protein Seq Data Anal. 1992;5(1):43-5.
419 34. Kowalski JC, Belfort M, Stapleton MA, Holpert M, Dansereau JT, Pietrokovski S, et al.
420 Configuration of the catalytic GIY-YIG domain of intron endonuclease I-TevI: coincidence of
421 computational and molecular findings. Nucleic Acids Res. 1999;27(10):2115-25.
422 35. Van Roey P, Meehan L, Kowalski JC, Belfort M, Derbyshire V. Catalytic domain
423 structure and hypothesis for function of GIY-YIG intron endonuclease I-TevI. Nat Struct Biol.
424 2002;9(11):806-11.
425 36. Truglio JJ, Rhau B, Croteau DL, Wang L, Skorvaga M, Karakas E, et al. Structural
426 insights into the first incision reaction during nucleotide excision repair. EMBO J.
427 2005;24(5):885-94.
428 37. Dunin-Horkawicz S, Feder M, Bujnicki JM. Phylogenomic analysis of the GIY-YIG
429 nuclease superfamily. BMC Genomics. 2006;7:98.

430 38. Ibryashkina EM, Sasnauskas G, Solonin AS, Zakharova MV, Siksnys V. Oligomeric
431 structure diversity within the GIY-YIG nuclease family. J Mol Biol. 2009;387(1):10-6.
432 39. Bruand C, Ehrlich SD. UvrD-dependent replication of rolling-circle plasmids in
433 Escherichia coli. Mol Microbiol. 2000;35(1):204-10.
434 40. Chang TL, Naqvi A, Anand SP, Kramer MG, Munshi R, Khan SA. Biochemical
435 characterization of the Staphylococcus aureus PcrA helicase and its role in plasmid rolling circle
436 replication. J Biol Chem. 2002;277(48):45880-6.
437 41. del Solar G, Giraldo R, Ruiz-Echevarría MJ, Espinosa M, Díaz-Orejas R. Replication and
438 control of circular bacterial plasmids. Microbiol Mol Biol Rev. 1998;62(2):434-64.
439 42. Naqvi A, Tinsley E, Khan SA. Purification and characterization of the PcrA helicase of
440 Bacillus anthracis. J Bacteriol. 2003;185(22):6633-9.
441 43. Khan SA. Rolling-circle replication of bacterial plasmids. Microbiol Mol Biol Rev.
442 1997;61(4):442-55.
443 44. Khan SA. Plasmid rolling-circle replication: recent developments. Mol Microbiol.
444 2000;37(3):477-84.
445 45. Soultanas P, Dillingham MS, Papadopoulos F, Phillips SE, Thomas CD, Wigley DB.
446 Plasmid replication initiator protein RepD increases the processivity of PcrA DNA helicase.
447 Nucleic Acids Res. 1999;27(6):1421-8.
448

449 **Table 1. Proteins encoded by the *SPAR* genes.**

Gene	Length (aa)[a]	Predicted Gene Function[b]
sparA	424	Hypothetical Protein
sparB	53	Hypothetical Protein
sparC	582	Hypothetical Protein
sparD	656	OLD-Family Endonuclease
sparE	609	GIY-YIG Endonuclease/PcrA/UvrD Helicase
sparF	232	GntR-Family Transcriptional Regulator
sparG	91	TIGR04197-Family T7SS Effector
sparH	102	Hypothetical Protein
sparI	137	LXG domain-containing protein
sparJ	417	T7SS effector LXG polymorphic toxin
sparK	199	TipC Family Immunity Protein
sparL	197	TipC Family Immunity Protein

450

451 [a]Based on the annotated genome of TX20005 (NZ_CP077423.1)

452 [b]The amino acid sequence of each protein was used to search the non-redundant protein

453 sequences database (updated on 04/05/2022) at NCBI using Protein BLAST.

454 **Figure Legends**

455 **Fig 1**. ***SPAR* genetic organization and expression. A.** The *SPAR* locus. Genes encoding
456 proteins with homology to proteins of known functions or containing characteristics of certain
457 protein families are colored in blue, while genes encoding hypothetical proteins of unknown
458 function are colored in gray. **B. RT-PCR.** cDNA synthesized from RNA extracted from HT29
459 cells co-cultured with TX20005 and the colonic tissues of three mice that had been orally
460 gavaged with TX20005 were used as a template for the RT-PCR. cDNA from HT29 cells only
461 and from mice orally gavaged with saline were used as controls to show the specificity of the
462 PCR primers. RNA samples that had not been treated with reverse transcriptase (RT) were used
463 as controls to show the lack of DNA contamination.

464

465 **Fig 2**. **SPAR is important for *Sgg* adherence but does not impact growth. A. Growth Curve**.
466 A single colony of WT or mutant bacteria was inoculated into 1mL of fresh BHI and incubated
467 at 37ºC overnight with shaking. The overnight culture was then diluted 1:100 with fresh BHI and
468 incubated at 37ºC with shaking. 100μL of the culture were taken at the 0, 3, 6, 9, 12, and 24
469 hours for serial dilution and plating. Results were combined from three biological replicates. **B.**
470 **Adherence**. HT29 cells were seeded at a density of $1x10^6$ cells per well and incubated with
471 TX20005 or TX20005*ΔSPAR* (MOI=10) in media only or in the presence of CS from TX20005
472 or TX20005*ΔSPAR* for 1 hour. Percentage adherence was calculated as the percentage of
473 adhered bacteria vs total bacteria added. Results were combined from three biological replicates.
474 Unpaired two-tailed *t* tests were used. **, $p<0.01$; NS, not significant.

475

476 **Fig 3**. ***SPAR* is essential for *Sgg* to stimulate CRC cell proliferation. A-D**. **Cell proliferation**
477 **assay.** HT29 (**A**), HCT116 (**B**), SW480 (**C**), and HEK293 (**D**) cells were incubated in the
478 presence or absence of bacteria (MOI=1) or CSs for 24 hours, as described in the Materials and
479 Methods section. Cell proliferation was measured using a CCK-8 kit. Cell-free wells containing
480 only the culture media were served as a blank control to which the absorbance values were
481 normalized. **E-G. Western blot**. Cell lysates from HT29 cells co-cultured in the presence or
482 absence of bacteria (MOI=1) were analyzed by western blot, probed with antibodies against β-
483 catenin, PCNA, and β-actin (**E**). Band intensity was quantified and normalized to β-actin. Fold
484 change in β-catenin (**F**) and PCNA (**G**) were calculated against cells incubated in media only.
485 Data was combined from three biological replicates. **H. Cell proliferation in the presence of**
486 ***Sgg* CSs.** HT29 cells were cultured in the presence or absence of WT or mutant *Sgg* (MOI=1) or
487 CSs from WT or mutant *Sgg* for 24 hours. Cell proliferation was measured by using a CCK-8 kit.
488 Results were combined from at least three biological replicates. Unpaired two-tailed *t* tests were
489 used for the comparisons. *, $p<0.05$; **, $p<0.01$; ***, $p<0.001$; ****, $p<0.0001$; ns, not
490 significant. Significance in panel H indicates comparison to untreated cells.
491
492 **Fig 4. *SPAR* deletion reduced the colonization capacity of *Sgg***. This was performed as
493 described in the Materials and Methods section. Colon tissues (**A**) and fecal pellets (**B**) were
494 weighed, homogenized in sterile PBS and dilution plated onto ESA plates. Mann-Whitney test
495 was used for the comparisons. *, $p<0.05$; **, $p<0.01$; ****, $p<0.0001$; NS, not significant.
496
497 **Fig 5. *SPAR* is critical for *Sgg* to promote the development of colon tumors**. This was
498 performed as described in the Materials and Methods section. The procedure is illustrated in **A**.

499 **B.** The number and size of macroscopic tumors were recorded by blinded observers. Tumor
500 burden is calculated as the sum of tumor volumes per mouse. **C.** *Sgg* load in the fecal pellets
501 collected at the experimental endpoint was determined by dilution plating of homogenized
502 pellets onto ESA plates. Mann-Whitney test was used for all comparisons. *, $p<0.05$; **, $p<0.01$;
503 ***, $p<0.001$; ****, p, <0.0001; NS, not significant.

Fig 1

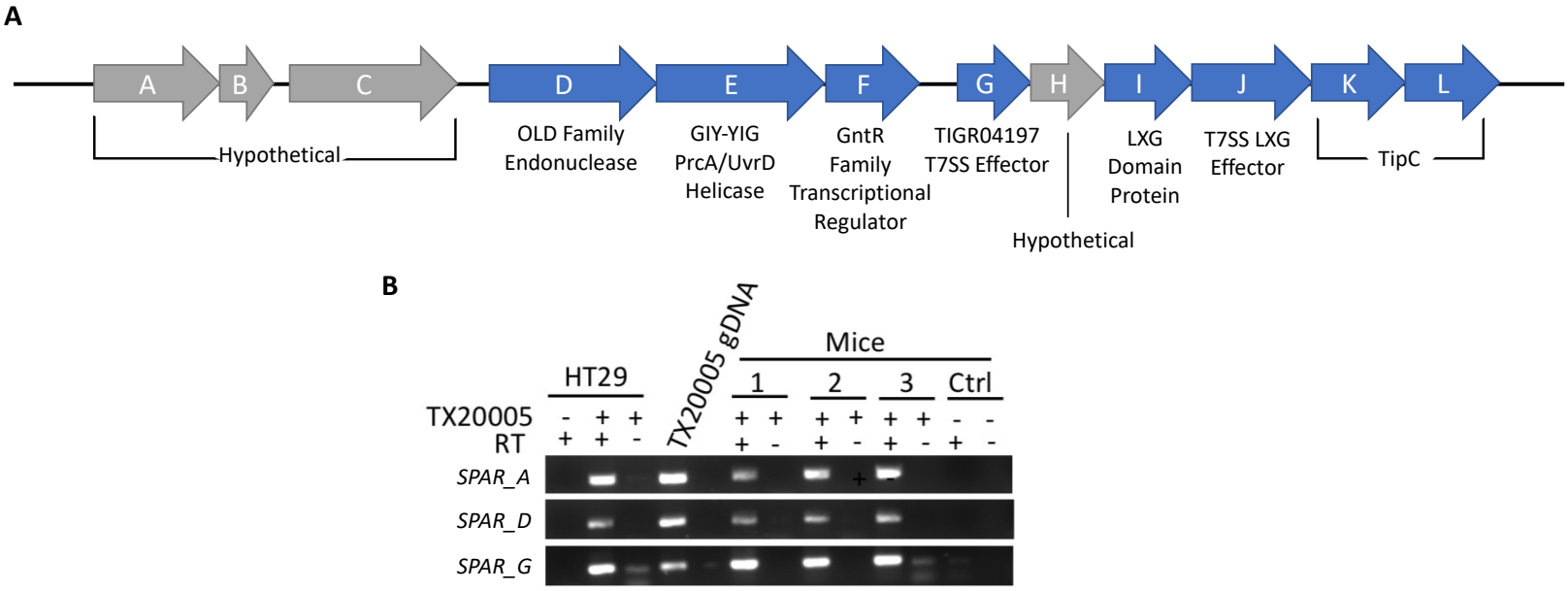
A
A
B
C
D
E
F
G
H
I
J
K
L
Hypothetical
OLD Family Endonuclease
GIY-YIG PrcA/UvrD Helicase
GntR Family Transcriptional Regulator
TIGR04197 T7SS Effector
Hypothetical
LXG Domain Protein
T7SS LXG Effector
TipC
B
HT29
TX20005 gDNA
Mice
1
2
3
Ctrl
TX20005
RT
SPAR_A
SPAR_D
SPAR_G

Fig 2

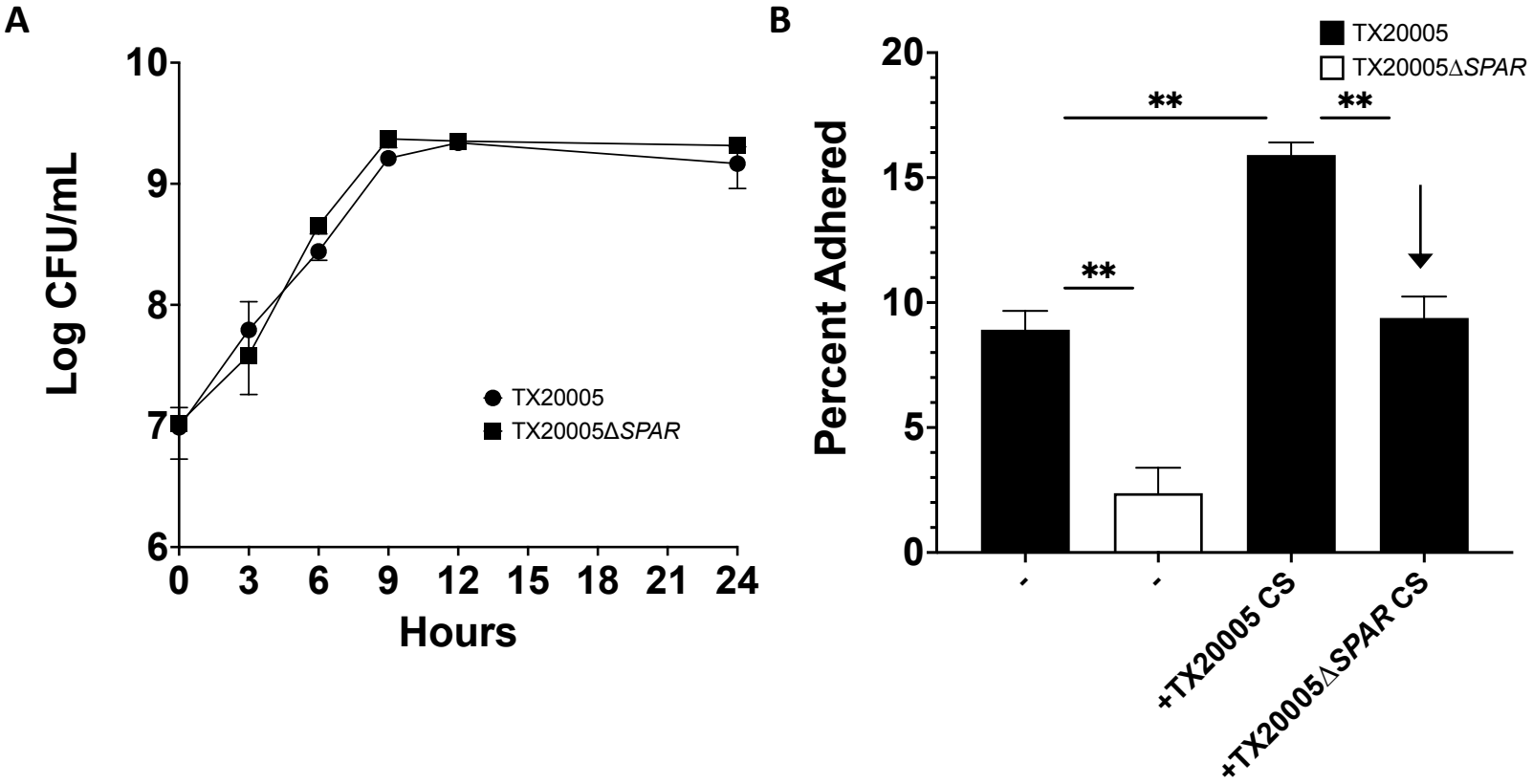
A
Log CFU/mL
10
9
8
7
6
0
3
6
9
12
15
18
21
24
Hours
TX20005
TX20005ΔSPAR
B
Percent Adhered
20
15
10
5
0
TX20005
TX20005ΔSPAR
**
**
**
-
-
+TX20005 CS
+TX20005ΔSPAR CS

Fig 3

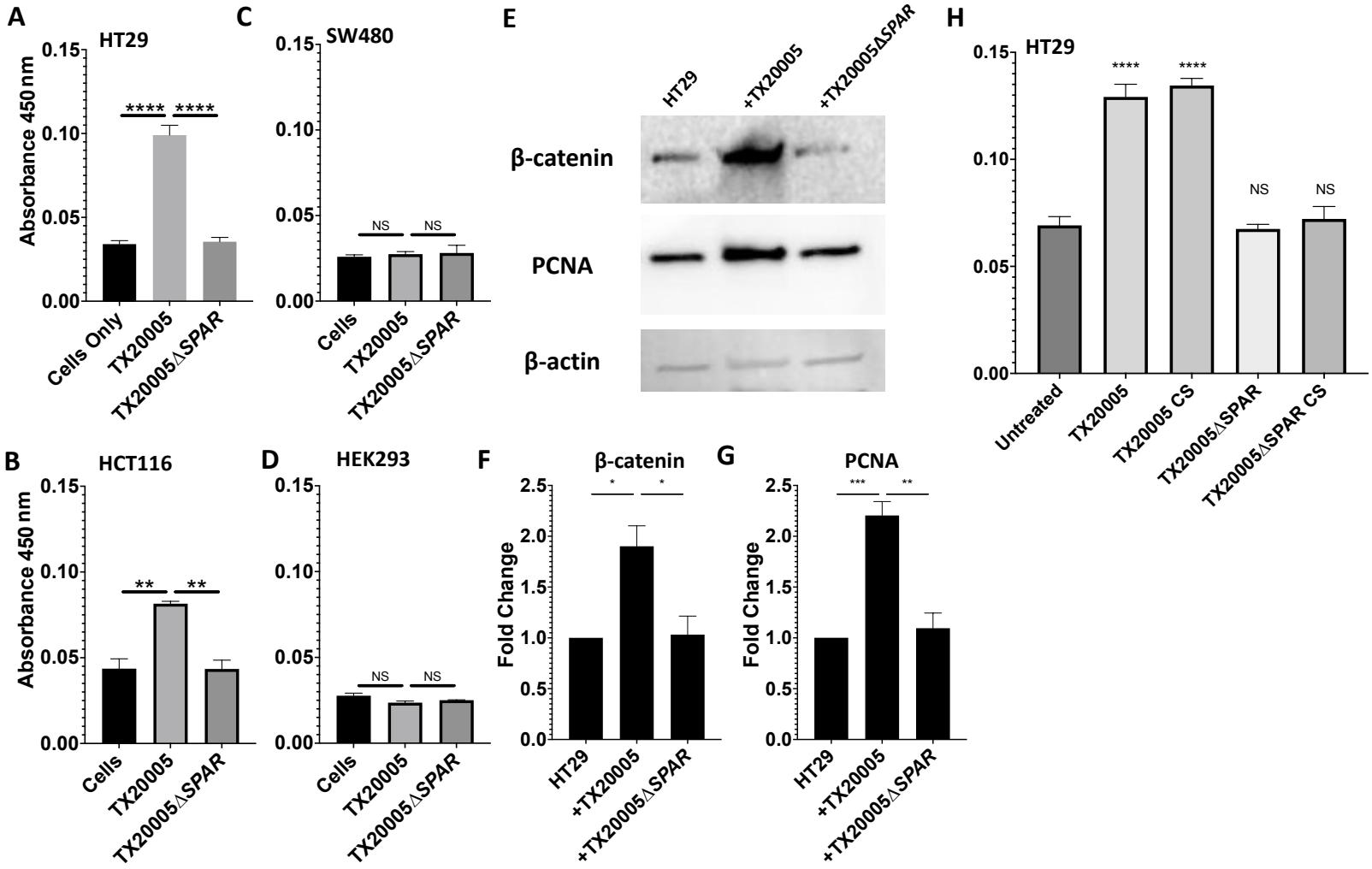
A
HT29
Absorbance 450 nm
0.15
0.10
0.05
0.00
****
****
Cells Only
TX20005
TX20005ΔSPAR
C
SW480
NS
NS
Cells
TX20005
TX20005ΔSPAR
E
HT29
+TX20005
+TX20005ΔSPAR
β-catenin
PCNA
β-actin
H
HT29
****
****
NS
NS
Untreated
TX20005
TX20005 CS
TX20005ΔSPAR
TX20005ΔSPAR CS
B
HCT116
**
**
D
HEK293
NS
NS
F
β-catenin
Fold Change
2.5
2.0
1.5
1.0
0.5
0.0
*
*
G
PCNA
***
**

Fig 4

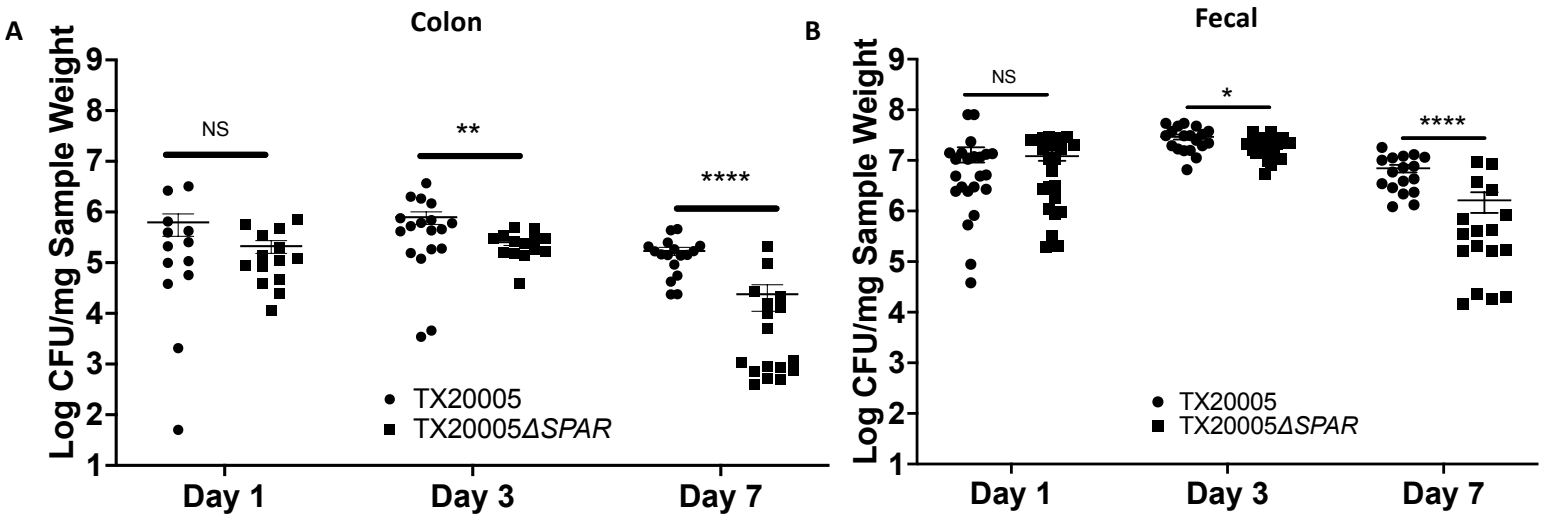
A
Colon
Log CFU/mg Sample Weight
9
8
7
6
5
4
3
2
1
NS
**
****
TX20005
TX20005ΔSPAR
Day 1
Day 3
Day 7
B
Fecal
Log CFU/mg Sample Weight
NS
*
****
TX20005
TX20005ΔSPAR
Day 1
Day 3
Day 7

Fig 5

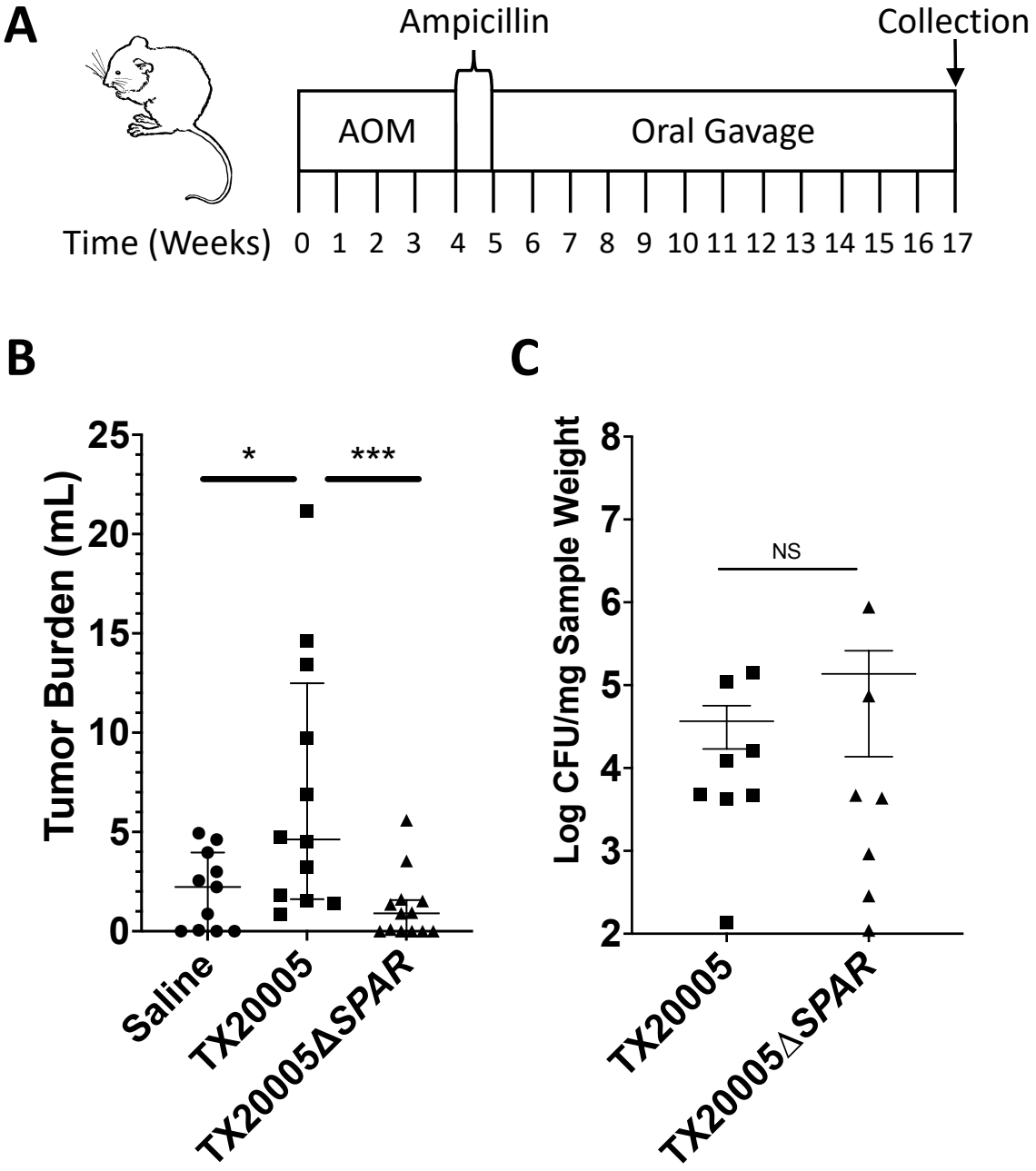
A
Ampicillin
Collection
AOM
Oral Gavage
Time (Weeks) 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
B
Tumor Burden (mL)
25
20
15
10
5
0
*
***
Saline
TX20005
TX20005ΔSPAR
C
Log CFU/mg Sample Weight
8
7
6
5
4
3
2
NS
TX20005
TX20005ΔSPAR

Fig S1

```
SPAR_H      GETDEALAELKKQEDDLEDALVDLLRKEAQDEH
SPAR_I      DQQQQCLNVLEEAELQLSLISVMSLERKTSLHQ
SPAR_J      EVKTTIFSSLVPKEVVAINPVTGAPIYASVSEDEAMAYLQENYEGSYLFIQSLEEEENLFGEDN
Consensus             HxxxDxxhxxxH
                          E
```

Supplemental Figure S1: The conserved T7SS sequence in SparH-J. H denotes highly conserved and h less conserved hydrophobic residues.