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

[#2]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 a pathobiont of the gut with clinical significance. *Sgg* adheres to the colonic epithelium and stimulates colorectal cancer (CRC) cell proliferation. In pre-clinical models of CRC, *Sgg* promotes the development of colon tumors. However, the molecular determinants of *Sgg* virulence remain unclear. Here, we identified a 12-gene locus in *Sgg* strain TX20005. Deletion of this locus resulted significantly reduced capacity to adhere to CRC cells and abrogated the ability of *Sgg* stimulate CRC cell proliferation. Thus, we designate this locus as the *Sgg* pathogenicity-associated region (SPAR). In a gut colonization model, mice exposed to the SPAR mutant showed significantly reduced *Sgg* load in the colon tissues and fecal materials. In a mouse mode of CRC, deletion of *SPAR* abolished the ability of *Sgg* to promote the development of colon tumors growth, while the fecal *Sgg* load in the CRC model does not show a significant difference between the wild type and the *SPAR* mutant. The results also show a connection between the *SPAR* locus and a previously described type VII secretion system of *Sgg*. Taken together, these results indicate that SPAR genes are critical pathogenicity determinants of *Sgg*. Studies to elucidate the specific activities of individual SPAR genes will be important.

Introduction

Streptococcus gallolyticus subspecies *gallolyticus* (*Sgg*) is a group D streptococcal bacterium, and an opportunistic pathogen belonging to the *Streptococcus bovis*/*Streptococcus equinis* complex[1]. *Sgg* primarily causes bacteremia and infective endocarditis (IE), and has had a long-standing clinical association with CRC, supported by extensive studies and case reports dating back to the 1950s[2-16]. Patients with *Sgg* IE and/or bacteremia have concomitant colon adenomas or adenocarcinomas at a much higher rate (~60%) than that of the general population[8]. Moreover, a prospective study found that 45% of patients with *Sgg* IE (45%) developd colonic neoplastic lesions within 5 years compared to 21% of patients with IE due to closely-related enterococci, further supporting the association of *Sgg* with CRC[6]. A recent metagenomic study of datasets from multiple countries across the continents confirmed that *Sgg* is enriched in colon tumors and can be a biomarker for CRC[17].

Functionally, *Sgg* stimulates the proliferation of human CRC cells in a β-catenin dependent manner *in vitro* and in a xenograft model[4]. Exposure to *Sgg* led to larger and more advanced tumors in an azoxymethane (AOM)-induced CRC model and in a colitis-associated CRC model[4,15]. These results highlight a pro-proliferation and pro-tumor activity of *Sgg*. Recent studies showed that there are variations among *Sgg* strains in their ability to stimulate human CRC cell proliferation. Some *Sgg* strains such as ATCC43143 showed significantly reduced capacity to adhere to HT29 cells, to stimulate the proliferation of these cells and to promote tumor development in the AOM model[2], suggesting that there are specific *Sgg* factors contributing to the pro-proliferative and pro-tumor activity of *Sgg*. The identity of these specific *Sgg* factors was unknown.

54 Here, we describe work characterizing a 12-gene locus in the chromosome of *Sgg* strain
55 TX20005. Deletion of this locus significantly impaired the ability of *Sgg* to adhere to cultured
56 CRC cells and to stimulate CRC cell proliferation. *In vivo*, the deletion mutant displayed
57 significantly reduced capacity to colonize normal mouse colon. More importantly, the mutant
58 lost the ability to promote the development of colon tumors. Given the role of this locus in *Sgg*
59 pathogenesis, we have coined this locus as the *Sgg* pathogenicity-associated region (*SPAR*).
60

61 **Results**

62 **The *Sgg SPAR* locus**. The SPAR locus was identified by comparing the genome
63 sequence of TX20005 (NZ_CP077423.1) with that of ATCC 43143 (NC_017576.1) using the
64 multiple genome alignment tool MAUVE[18]. The locus is situated in a region displaying
65 substantial differences between the strains. The locus is comprised of 12 genes (*sparA* to *sparL*),
66 organized into 3 putative operons, *SPAR 1, 2,* and *3*, and one standalone gene, *sparC* (Fig. 1A).
67 The majority of the *spar* genes except for *sparD*, *E*. and *F* encode hypothetical proteins of
68 unknown function (Table 1). SparD to SparF exhibits homology to Predicted ATP-dependent
69 endonuclease of the OLD family, GIY-YIG endonuclease/PcrA/UvrD helicase, and GntR family
70 transcriptional regulator, respectively.

71 We next sought to determine if the *SPAR* genes were expressed *in vitro* and *in vivo* by
72 RT-PCR (Fig 1B). We focused on the expression of the first gene of the three putative operons,
73 *sparA*, *D* and *G*, under conditions that are relevant to the pro-proliferation and pro-tumor activity
74 of *Sgg*. All three genes were expressed when the bacteria were co-cultured with HT29 cells (Fig.
75 1B). We further analyzed the expression of these genes in the colonic tissues collected from mice

76 orally gavaged with TX20005. The result showed that these genes were also expressed in the
77 colon of mice (Fig. 1B).

78 ***SPAR* is important for *Sgg* adherence.** To investigate the role of the *SPAR* locus in the
79 pathogenic potential of *Sgg*, we generated a mutant in the TX20005 background, in which all 12
80 genes were deleted (TX20005Δ*SPAR*). We verified that the deletion does not affect bacterial
81 growth (Fig. 2A).

82 We sought to determine if deletion of the *spar* locus had any effect on the adherence
83 capability of *Sgg*. The results showed that the mutant strain exhibited a significantly decreased
84 capacity for adherence to HT29 cells compared to the WT parent strain (Fig 2B). Previous work
85 indicated that type VII-secreted factors of *Sgg* are involved in adherence by significantly
86 enhancing the adherence capacity of the WT bacteria[3]. We tested whether filter-sterilized culture
87 supernatants (CS) from TX20005Δ*SPAR* can enhance the adherence capacity of TX20005.
88 Addition of the WT CS significantly enhanced the adherence of WT bacteria, but not that of the
89 mutant. Further, addition of the mutant supernatant did not render any increase in WT adherence,
90 nor did it enhance mutant adherence. These results suggest that the *SPAR* locus is important for
91 *Sgg* to adhere to CRC cells. Furthermore, the results also suggest that the contribution of the
92 *SPAR* locus to adherence involves secreted factors and surface adhesins.

93 ***SPAR* is required for *Sgg* to stimulate CRC cell proliferation.** *Sgg* has been previously
94 shown to stimulate cell proliferation in human CRC cell lines HT29 and HCT116, but not in
95 HEK293 or SW480 cell lines[2-4]. We examined the ability of TX20005Δ*SPAR* to stimulate cell
96 proliferation in these cell lines. HT29, HCT116, SW480, and HEK293 cells were co-cultured
97 with the wild type (WT) strain TX20005 or TX20005Δ*SPAR* (Fig 3A-D). As anticipated, the WT
98 strain significantly stimulated HT29 and HCT116 cell proliferation compared to untreated cells.

99 In contrast, HT29 and HCT116 cells exposed to the mutant did not exhibit any significant
100 increase in cell proliferation, indicating that *SPAR* is required for *Sgg*-stimulated CRC cell
101 proliferation. Neither the WT or the mutant had any effect on the proliferation of SW480 or
102 HEK293 cells, as expected.

103 To further validate the results, we investigated the effect of the mutant on cell
104 proliferation markers β-catenin and proliferating cell nuclear antigen (PCNA) in HT29 cells (Fig.
105 3E-G). Cells co-cultured with WT TX20005 exhibited a significant increase in the levels of both
106 β-catenin and PCNA compared to cells cultured in media only, as expected. Notably, cells co-
107 cultured with the *SPAR* mutant showed no difference in the level of β-catenin or PCNA
108 compared to cells cultured in media only, further confirming that the *spar* locus is essential for
109 *Sgg* to stimulate cell proliferation.

110 Previous work demonstrated that filter-sterilized CS from *Sgg* is able to induce cell
111 proliferation[3,19]. Therefore, we examined the effect of CS from TX20005Δ*SPAR* (Fig. 2H). Cells
112 treated with the WT CS exhibited a significant increase in cell proliferation, as expected, while
113 cells treated with the mutant CS did not exhibit any significant increase in cell proliferation.
114 These results suggest that the defect in the *SPAR* mutant to stimulate host cell proliferation is due
115 to, at least in part, the absence of secreted factors in the CS. Taken together, these results indicate
116 that *SPAR*, and particularly secreted factors dependent on *SPAR* is required for *Sgg* to stimulate
117 CRC cell proliferation *in vitro*.

118 ***Sgg* gut colonization is impaired by *SPAR* deletion.** The ability of *Sgg* to colonize the
119 host gut is an important aspect of its pathogenic potential[2]. We sought to determine if the *SPAR*
120 locus is involved in *Sgg* gut colonization (Fig 4, A and B). Mice were orally gavaged with WT
121 and the mutant bacteria. Colon tissues as well as fecal materials were collected at days 1, 3, and

122 7 post gavage to determine the bacterial load in the samples. In the colon tissues, the bacterial
123 load of the WT and mutant strains did not significantly differ at day 1. However, the bacterial
124 load of the mutant was significantly decreased compared to that of the WT at days 3 and 7 (Fig.
125 4A). In the fecal pellets, a similar patten was observed, such that at the 1-day mark, the bacterial
126 loads of mutant and WT strains did not significantly differ. At days 3 and 7, the bacterial load of
127 the mutant was decreased significantly compared to that of the WT (Fig. 4B). Taken together,
128 these results indicate that the *SPAR* locus is an important contributor to the colonization capacity
129 of *Sgg* in the normal colon.

130 ***SPAR* is essential for the ability of *Sgg* to promote the development of colon tumors**.
131 Next, we investigated if *SPAR* is important for the ability of *Sgg* to promote the development of
132 colon tumors. We examined the effect of TX20005Δ*SPAR* on macroscopic tumors in the AOM-
133 induced CRC model (Fig 5A). Mice gavaged with the WT strain exhibited a significant increase
134 in tumor incidence (Fig. 5B), tumor multiplicity (Fig. 5C), tumor size (Fig. 5D) and total tumor
135 burden (Fig 5E) compared to saline-gavaged mice. In contrast, mice exposed to the mutant
136 showed no significant changes in incidence, multiplicity or size compared to saline treated mice,
137 and significant reduction compared to WT-treated mice (Fig. 5B-E). We also determined the
138 bacterial load in the fecal pellets isolated at collection and found no difference in the *Sgg* burden
139 between WT or mutant-treated groups (Fig. 5F). We note that this result could be due to the
140 repeated gavage with bacteria during the experiment and is not a definitive measure of the ability
141 of the strains to colonize tumor-bearing colons. Taken together, these results suggest that the
142 *SPAR* locus plays a critical role in *Sgg*-induced of colon tumor development.

143 **Discussion**

144 In this study, we describe a 12-gene *Sgg* chromosomal locus, *SPAR*, that plays a key role
145 to *Sgg* pathogenicity in the gut. Deletion of *SPAR* renders *Sgg* deficient in adherence to CRC
146 cells and colonization of the colon, abrogates the ability for *Sgg* to stimulate CRC cell
147 proliferation *in vitro* and to promote the development of colon tumors *in vivo*. Furthermore,
148 results from this study also suggested that multiple factors within the *SPAR* locus may contribute
149 to its phenotype in a direct or indirect manner.

150 Our data showed that deletion of the *SPAR* locus resulted in significantly reduced
151 adherence of *Sgg* to HT29 cells and colonization of the colon. Previous studies showed that
152 T7SS-secreted factors enhance *Sgg* adherence[3]. Interestingly, our results revealed that *SPAR*
153 deletion eliminated the adherence-enhancing activity in the CS, suggesting that the deletion
154 might have affected the T7-secreted factors involved in adherence. Furthermore, previous work
155 showed that T7SS-dependent surface anchored factor(s) is required for adherence[3]. The *SPAR*
156 mutant also exhibited a similar deficiency in the surface anchored factor(s), in that the CS from
157 WT bacteria could not enhance the adherence of the *SPAR* mutant. In terms of cell proliferation,
158 we showed that the *SPAR* locus is essential for *Sgg* to stimulate certain CRC cell proliferation.
159 Furthermore, the results suggest that *SPAR* is also important for the production of secreted
160 factors that stimulate host cell proliferation. Thus, the *SPAR* mutant exhibits a very similar
161 phenotype as that described for a T7SS defective mutant of *Sgg* (TX20005*Δesx*) with respect to
162 adherence and induction of cell proliferation[3]. In the AOM model of CRC, we observed a
163 striking difference between the *SPAR* mutant and the WT bacteria in promoting the development
164 of colon tumors. This phenotype is much stronger than previously reported for the T7SS mutant[3],

165 suggesting that the *SPAR* locus has a more extensive tumor promoting activity than the T7SS of
166 *in vivo*.

167 Further examination of genes within the *SPAR* locus showed that *sparG – sparK* in the
168 third putative operon display some characteristics of proteins secreted by T7SS. SparG, H and I
169 are of 91, 102 and 137 amino acids in length. Analysis with HHpred showed that SparG and
170 SparI fold into a four-helical bundle structure typical of ESAT-6-like WXG100 proteins with a
171 probability of 96.48% and 97.97%, respectively. SparH to SparK contain a conserved sequence
172 motif HxxxD/ExxhxxxH (H denotes highly conserved and h less conserved hydrophobic
173 residues) that is considered to be important for the secretion of WXG100 proteins by T7SS[20]
174 (Supplemental Fig. S1). SparH – SparK also contain an LXG motif present in WXG100 proteins
175 or T7SS-secreted LXG-toxins. Thus, SparG to SparK are potential T7SS secreted proteins.
176 Genes encoding the T7 secretion machinery are not identified within the *SPAR* locus. In fact, the
177 previously identified T7SS (*Sgg*T7SS^{T05}) is the only T7SS locus in TX20005 that contains a
178 complete set of secretion machinery genes. It is likely that SparG to SparK are secreted through
179 *Sgg*T7SS^{T05}.

180 The *sparF* gene is predicted to encode a GntR family transcriptional regulator. This
181 family of regulators contain a conserved N-terminal DNA binding domain and a C-terminal
182 effector-binding/oligomerization domain which varies between different GntR family members[21].
183 A recent publication indicated that a GntR family transcriptional regulator (OG1RF_11099) of
184 *Enterococcus faecalis* played an important role in the activation of T7SS gene transcription[22].
185 Using Global Align (NCBI), we found that SparF is highly homologous to the OG1RF_11099,
186 showing a 74% positivity at the amino acid sequence level. This indicates that in addition to the
187 putative T7SS secreted proteins in the *SPAR* locus, SparF can potentially regulate the expression

188 of T7SS genes. Further studies are needed to reveal the specific activities of SparF to SparK in
189 the context of adherence and cell proliferation. However, the presence of these genes within the
190 *SPAR* locus is likely a reason for the phenotypic similarity between the *SPAR* and T7SS mutants.
191 Our data also showed that the *SPAR* mutant has a more pronounced phenotype than the
192 T7SS mutant *in vivo*, suggesting that in addition to *Sgg*T7SS^{T05}-mediated activities, the *SPAR*
193 locus may encode or regulate functions that faciliate tumor development. One possibility is that
194 SparF can regulate the expression of genes in addition to T7SS genes. Those genes may
195 contribute to the more severe defects observed for the *SPAR* mutant. It is also possible that other
196 genes within the *SPAR* locus, SparA-SparE, play a role, however, their predicted function based
197 on homology search does not provide clues regarding their possible contribution. SparA-C are
198 putative hypothetical proteins. SparD is homologous to OLD family endonucleases. OLD family
199 endonucleases are widely present among bacteria and archaea[23]. While their biological function
200 is not completely understood, they contain a C-terminal ATPase domain, as well as an N-
201 terminal Toprim domain, which is believed to be important in DNA replication, recombination,
202 and repair[24,25]. SparE is homologous to GIY-YIG endonucleas[26-30] and PcrA/UvrD helicase. The
203 PcrA/UvrD helicase is an essential helicase in gram-positive bacteria and has been shown to play
204 a role in DNA repair, as well as the replication of small drug-resistance plasmids in
205 *Staphylococcus aureus* and *Bacillus subtilis*[31-37]. This raises the possibility that the *SPAR* locus
206 may be a remnant of a plasmid or mobile genetic element that had been integrated into the
207 chromosome.
208 In summary, we have identified a chromosomal locus in *Sgg*, *SPAR*, that is vital to the
209 pathogenicity of *Sgg*. We report that deletion of *SPAR* significantly reduces the capacity of *Sgg*
210 to adhere to CRC cells and to colonize the gut. Further, *SPAR* deletion abrogates the ability for

211 *Sgg* to stimulate CRC cell proliferation and to accelerate colon tumor development. Examination
212 of genes with the locus provided potential candidates responsible for the observed phenotypes
213 and suggested a connection between the *SPAR* locus and T7SS in terms of additional T7SS
214 effectors and a possible regulator of T7SS expression. Further investigation into the activities of
215 specific *SPAR* proteins will be important for unveiling the intricate *Sgg* virulence mechanisms
216 and will open up the path ahead to identify biomarker candidates and targets for clinical
217 prevention and intervention strategies.

Materials and Methods

Bacterial strains, cell lines and growth conditions *Sgg* strains were routinely grown in brain heart infusion (BHI) broth or tryptic soy broth with shaking, or on BHI or TSB agar plates at 37 ˚C overnight (Teknova). For co-culture experiments and animal studies, stationary phase bacterial cultures were pelleted, washed with phosphate-buffered saline (PBS), pH 7.4, resuspended in PBS containing 15% glycerol, aliquoted and stored at -80 ˚C. Aliquoted stocks were thawed and washed with PBS, and diluted in appropriate media to obtain the desired concentration. Human colon cancer cell lines HT29, HCT116, and SW480, as well as the HEK293 cell line (ATCC), were grown in Dulbecco's Modified Eagle's Medium F-12 50/50 (DMEM/F-12, GIBCO) supplemented with 10% FBS at 37 ˚C with 5% CO2 in a humidified chamber. Cells from less than 30 passages were used in the experiments.

Adherence assay. This was performed as described previously[3]. Briefly, HT29 cells were seeded at a density of $1x10^6$ cells/well in a 24 well plate and allowed to attach overnight. Cells were incubated with bacteria at a multiplicity of infection (MOI) of 10 for 1 hour at 37 ˚C with 5% CO2. Cells were then washed with PBS, lysed with PBS containing 0.025% Triton-X 100 and dilution plated. Adherence was calculated as the percentage of adherent bacteria vs. total bacteria added.

Cell proliferation assay. This was performed as described previously[3] with slight modifications. Briefly, cells were seeded in 96 well plates at a concentration of ~$1x10^4$ cells/well and incubated overnight. Cells were then incubated in fresh DMEM/F-12 with 10% FBS and *Sgg* bacteria (MOI = 1) for a total of 24 hours. Trimethoprim was added after 6 hours of incubation (1 μg/mL final concentration) to inhibit bacterial growth, as previously described. The number of

240 viable cells was determined using the cell counting kit (CCK)-8 kit following the instructions of
241 the supplier (Apex Bio).

242 **Western Blot**. Detection of β-catenin and PCNA was carried out as described
243 previously[3]. Briefly, total cell lysates from HT29 cells cultured in the presence or absence of
244 bacteria were subjected to SDS-PAGE, transferred, and probed with antibodies against β-catenin
245 (1:1000), PCNA (1:1000), and β-actin (1:1000) (Cell Signaling Technology), followed by
246 incubation with HRP-conjugated secondary antibodies (1:3000). Band intensities were measured
247 using FIJI (ImageJ) and normalized to those of β-actin.

248 **Deletion of *SPAR***. This was performed as described previously[3]. Briefly, the ~ 1kb
249 region upstream of *SPAR_A* and the ~1kb region downstream of *SPAR_L* were synthesized by
250 Genscript and cloned into pUC57 (Genscript). The insert was then subcloned into a temperature
251 sensitive conjugative plasmid pG1-*oriT*TnGBS1. The insert sequence was verified by Sanger
252 sequencing. The construct was introduced into *S. agalactiae* NEM316 by electroporation and
253 then into *Sgg* strain TX20005 by conjugation under the permissive temperature. Erythromycin-
254 resistant transconjugants were then cultured under a non-permissive temperature to select for
255 single cross-over recombinants, followed by serial passages in antibiotic-free BHI and screening
256 for double cross-over deletion mutants by PCR. Deletion was confirmed by PCR amplification of
257 the regions spanning the deleted fragment and DNA sequencing of the PCR product.

258 **RNA extraction, cDNA synthesis and PCR.** This was performed following the method
259 described previously [Taylor]. Primers used in PCR amplification are: *sparA*, forward
260 5'GCAAGCTGGTCGAACAGAAC and reverse 5'GCTTCTATGGTTGGGGCTAGA; *sparD*,
261 forward 5' GGAGGTGGATCCAACAAGGG and reverse 5'CAGGTTCCTCGATAGCCAGC;

262 and *sparG*, forward 5'TCAGTTGTTAGCGGATGCGT and reverse 5'
263 CCCTTTATTGCTTGTGCTCCC.
264 **Growth curve**. Overnight cultures of *Sgg* were inoculated into fresh BHI broth at 1:100
265 dilution and grown at 37 ˚C with shaking. Samples were taken at 0, 3, 6, 9, 12, and 24 hours,
266 dilution plated, incubated for 24 hours and colonies enumerated.
267 **Preparation of culture supernatants.** Supernatants from stationary phase Sgg cultured
268 in BIH were filtered-sterilized through a 0.2μm filter and concentrated approximately 10 – 20
269 fold using centrifugal concentrators (3kD molecular weight cut off). The concentrated CSs were
270 aliquoted and stored at -80°C. A vial is thawed and diluted in the appropriate media for
271 adherence or cell proliferation assays.
272 **Animal experiments**. Animal studies were performed in accordance with protocols
273 approved by the Institutional Animal Care and Use Committee at the Texas A&M Health
274 Science Center, Institute of Biosciences and Technology. Mice were fed with standard ProLab
275 IsoPro RMH3000 (LabDiet). **Colonization experiments.** This was performed as previously
276 described[3] with slight modifications. Briefly, 6-week old A/J mice, sex matched (Jackson
277 Laboratory), were treated with ampicillin at a concentration of 1 g/L in drinking water for 3 days
278 and switched to antibiotic-free water 24 hours prior to administration of bacteria. *Sgg* was orally
279 gavaged at a dose of ~1 x 10^9 CFU/mouse. Colons and fecal materials were collected at day 1, 3,
280 and 7 post-gavage. Samples were weighed, homogenized in PBS in a TissueLyser (Qiagen),
281 dilution plated onto Enterococcus Selective Agar (ESA) plates and incubated at 37 ˚C for 24–48
282 hours to enumerate *Sgg* colonies, as previously described. **AOM-induced model of CRC.** This
283 was performed as previously described[3] with slight modifications. Briefly, A/J mice were treated
284 with 4 weekly i.p. injections of AOM (10 mg/kg body weight), followed by ampicillin in

drinking water (1 g/L) for 7 days and switched to antibiotic-free water 24 hours prior to the first oral gavage with *Sgg*. Mice were orally gavaged with bacteria at ~ $1x10^9$ CFU/mouse or saline once per week for 12 weeks. Colons of mice were collected and cut opened longitudinally. The number and size of macroscopic were recorded. Tumor burden was calculated as the sum of all tumor volumes per mouse. Visual evaluation of colons was carried out by a blinded observer.

Statistical analysis. GraphPad Prism 8 was used for statistical analyses. Two-tailed unpaired *t*-test was used for pairwise comparisons to assess the significance of differences between two groups in cell proliferation assays, western blot analysis, and adherence assays. The non-parametric Mann-Whitney test was used to assess the significance of differences in the bacterial load and tumor burden in animal studies. Two-way ANOVA was used to analyze the bacterial growth curves. Ns, not significant; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$.

Acknowledgements

This study was supported by funding from the Hamill Foundation, the Cancer Prevention Research Institute of Texas and NIAID.

300 **References**

301 1 Schlegel, L., Grimont, F., Ageron, E., Grimont, P. A. D. & Bouvet, A. Reappraisal of the
302 taxonomy of the Streptococcus bovis/Streptococcus equinus complex and related species:
303 description of Streptococcus gallolyticus subsp. gallolyticus subsp. nov., S. gallolyticus
304 subsp. macedonicus subsp. nov. and S. gallolyticus subsp. pasteurianus subsp. nov. *Int J*
305 *Syst Evol Microbiol* **53**, 631-645, doi:10.1099/ijs.0.02361-0 (2003).

306 2 Kumar, R., Herold, J. L., Taylor, J., Xu, J. & Xu, Y. Variations among Streptococcus
307 gallolyticus subsp. gallolyticus strains in connection with colorectal cancer. *Scientific*
308 *reports* **8**, 1514, doi:10.1038/s41598-018-19941-7 (2018).

309 3 Taylor, J. C. *et al*. A type VII secretion system of Streptococcus gallolyticus subsp.
310 gallolyticus contributes to gut colonization and the development of colon tumors. *PLoS*
311 *Pathog* **17**, e1009182, doi:10.1371/journal.ppat.1009182 (2021).

312 4 Kumar, R. *et al*. Streptococcus gallolyticus subsp. gallolyticus promotes colorectal tumor
313 development. *PLoS Pathog* **13**, e1006440, doi:10.1371/journal.ppat.1006440 (2017).

314 5 Marmolin, E. S. *et al*. Bacteremia with the bovis group streptococci: species
315 identification and association with infective endocarditis and with gastrointestinal
316 disease. *Diagn Microbiol Infect Dis* **85**, 239-242,
317 doi:10.1016/j.diagmicrobio.2016.02.019 (2016).

318 6 Corredoira, J. *et al*. Differences between endocarditis caused by Streptococcus bovis and
319 Enterococcus spp. and their association with colorectal cancer. *Eur J Clin Microbiol*
320 *Infect Dis* **34**, 1657-1665, doi:10.1007/s10096-015-2402-1 (2015).

321 7 Boleij, A. & Tjalsma, H. Gut bacteria in health and disease: a survey on the interface
322 between intestinal microbiology and colorectal cancer. *Biol Rev Camb Philos Soc* **87**,
323 701-730, doi:10.1111/j.1469-185X.2012.00218.x (2012).
324 8 Boleij, A., van Gelder, M. M., Swinkels, D. W. & Tjalsma, H. Clinical Importance of
325 Streptococcus gallolyticus infection among colorectal cancer patients: systematic review
326 and meta-analysis. *Clin Infect Dis* **53**, 870-878, doi:10.1093/cid/cir609
327 cir609 [pii] (2011).
328 9 Abdulamir, A. S., Hafidh, R. R. & Abu Bakar, F. The association of Streptococcus
329 bovis/gallolyticus with colorectal tumors: the nature and the underlying mechanisms of
330 its etiological role. *J Exp Clin Cancer Res* **30**, 11, doi:10.1186/1756-9966-30-11 (2011).
331 10 Gupta, A., Madani, R. & Mukhtar, H. Streptococcus bovis endocarditis, a silent sign for
332 colonic tumour. *Colorectal Dis* **12**, 164-171, doi:CDI1814 [pii]
333 10.1111/j.1463-1318.2009.01814.x (2010).
334 11 Waisberg, J., Matheus, C. e. O. & Pimenta, J. Infectious endocarditis from Streptococcus
335 bovis associated with colonic carcinoma: case report and literature review. *Arq*
336 *Gastroenterol* **39**, 177-180, doi:10.1590/s0004-28032002000300008 (2002).
337 12 Alazmi, W., Bustamante, M., O'Loughlin, C., Gonzalez, J. & Raskin, J. B. The
338 association of Streptococcus bovis bacteremia and gastrointestinal diseases: a
339 retrospective analysis. *Dig Dis Sci* **51**, 732-736, doi:10.1007/s10620-006-3199-7 (2006).
340 13 Gold, J. S., Bayar, S. & Salem, R. R. Association of Streptococcus bovis bacteremia with
341 colonic neoplasia and extracolonic malignancy. *Arch Surg* **139**, 760-765,
342 doi:10.1001/archsurg.139.7.760 (2004).

343 14 Corredoira, J. *et al*. The clinical epidemiology and malignancies associated with
344 Streptococcus bovis biotypes in 506 cases of bloodstream infections. *J Infect* **71**, 317-
345 325, doi:10.1016/j.jinf.2015.05.005 (2015).
346 15 Zhang, Y., Weng, Y., Gan, H., Zhao, X. & Zhi, F. Streptococcus gallolyticus conspires
347 myeloid cells to promote tumorigenesis of inflammatory bowel disease. *Biochem Biophys*
348 *Res Commun* **506**, 907-911, doi:10.1016/j.bbrc.2018.10.136 (2018).
349 16 Kwong, T. N. Y. *et al*. Association Between Bacteremia From Specific Microbes and
350 Subsequent Diagnosis of Colorectal Cancer. *Gastroenterology* **155**, 383-390.e388,
351 doi:10.1053/j.gastro.2018.04.028 (2018).
352 17 Thomas, A. M. *et al*. Metagenomic analysis of colorectal cancer datasets identifies cross-
353 cohort microbial diagnostic signatures and a link with choline degradation. *Nat Med* **25**,
354 667-678, doi:10.1038/s41591-019-0405-7 (2019).
355 18 Darling, A. C., Mau, B., Blattner, F. R. & Perna, N. T. Mauve: multiple alignment of
356 conserved genomic sequence with rearrangements. *Genome Res* **14**, 1394-1403,
357 doi:10.1101/gr.2289704 (2004).
358 19 Corredoira, J. *et al*. Colorectal neoplasm in cases of Clostridium septicum and
359 Streptococcus gallolyticus subsp. gallolyticus bacteraemia. *Eur J Intern Med* **41**, 68-73,
360 doi:10.1016/j.ejim.2017.02.009 (2017).
361 20 Poulsen, C., Panjikar, S., Holton, S. J., Wilmanns, M. & Song, Y. H. WXG100 protein
362 superfamily consists of three subfamilies and exhibits an α-helical C-terminal conserved
363 residue pattern. *PLoS One* **9**, e89313, doi:10.1371/journal.pone.0089313 (2014).
364 21 Liu, G. F., Wang, X. X., Su, H. Z. & Lu, G. T. Progress on the GntR family transcription
365 regulators in bacteria. *Yi Chuan* **43**, 66-73, doi:10.16288/j.yczz.20-245 (2021).

366 22 Chatterjee, A., Willett, J. L. E., Dunny, G. M. & Duerkop, B. A. Phage infection and sub-
367 lethal antibiotic exposure mediate Enterococcus faecalis type VII secretion system
368 dependent inhibition of bystander bacteria. *PLoS Genet* **17**, e1009204,
369 doi:10.1371/journal.pgen.1009204 (2021).
370 23 Schiltz, C. J., Lee, A., Partlow, E. A., Hosford, C. J. & Chappie, J. S. Structural
371 characterization of Class 2 OLD family nucleases supports a two-metal catalysis
372 mechanism for cleavage. *Nucleic Acids Res* **47**, 9448-9463, doi:10.1093/nar/gkz703
373 (2019).
374 24 Aravind, L., Leipe, D. D. & Koonin, E. V. Toprim--a conserved catalytic domain in type
375 IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR
376 proteins. *Nucleic Acids Res* **26**, 4205-4213, doi:10.1093/nar/26.18.4205 (1998).
377 25 Koonin, E. V. & Gorbalenya, A. E. The superfamily of UvrA-related ATPases includes
378 three more subunits of putative ATP-dependent nucleases. *Protein Seq Data Anal* **5**, 43-
379 45 (1992).
380 26 Kowalski, J. C. *et al*. Configuration of the catalytic GIY-YIG domain of intron
381 endonuclease I-TevI: coincidence of computational and molecular findings. *Nucleic*
382 *Acids Res* **27**, 2115-2125, doi:10.1093/nar/27.10.2115 (1999).
383 27 Van Roey, P., Meehan, L., Kowalski, J. C., Belfort, M. & Derbyshire, V. Catalytic
384 domain structure and hypothesis for function of GIY-YIG intron endonuclease I-TevI.
385 *Nat Struct Biol* **9**, 806-811, doi:10.1038/nsb853 (2002).
386 28 Truglio, J. J. *et al*. Structural insights into the first incision reaction during nucleotide
387 excision repair. *EMBO J* **24**, 885-894, doi:10.1038/sj.emboj.7600568 (2005).

388 29 Dunin-Horkawicz, S., Feder, M. & Bujnicki, J. M. Phylogenomic analysis of the GIY-
389 YIG nuclease superfamily. *BMC Genomics* **7**, 98, doi:10.1186/1471-2164-7-98 (2006).
390 30 Ibryashkina, E. M., Sasnauskas, G., Solonin, A. S., Zakharova, M. V. & Siksnys, V.
391 Oligomeric structure diversity within the GIY-YIG nuclease family. *J Mol Biol* **387**, 10-
392 16, doi:10.1016/j.jmb.2009.01.048 (2009).
393 31 Bruand, C. & Ehrlich, S. D. UvrD-dependent replication of rolling-circle plasmids in
394 Escherichia coli. *Mol Microbiol* **35**, 204-210, doi:10.1046/j.1365-2958.2000.01700.x
395 (2000).
396 32 Chang, T. L. *et al*. Biochemical characterization of the Staphylococcus aureus PcrA
397 helicase and its role in plasmid rolling circle replication. *J Biol Chem* **277**, 45880-45886,
398 doi:10.1074/jbc.M207383200 (2002).
399 33 del Solar, G., Giraldo, R., Ruiz-Echevarría, M. J., Espinosa, M. & Díaz-Orejas, R.
400 Replication and control of circular bacterial plasmids. *Microbiol Mol Biol Rev* **62**, 434-
401 464, doi:10.1128/MMBR.62.2.434-464.1998 (1998).
402 34 Naqvi, A., Tinsley, E. & Khan, S. A. Purification and characterization of the PcrA
403 helicase of Bacillus anthracis. *J Bacteriol* **185**, 6633-6639, doi:10.1128/JB.185.22.6633-
404 6639.2003 (2003).
405 35 Khan, S. A. Rolling-circle replication of bacterial plasmids. *Microbiol Mol Biol Rev* **61**,
406 442-455, doi:10.1128/mmbr.61.4.442-455.1997 (1997).
407 36 Khan, S. A. Plasmid rolling-circle replication: recent developments. *Mol Microbiol* **37**,
408 477-484, doi:10.1046/j.1365-2958.2000.02001.x (2000).

409 37 Soultanas, P. *et al*. Plasmid replication initiator protein RepD increases the processivity
410 of PcrA DNA helicase. *Nucleic Acids Res* **27**, 1421-1428, doi:10.1093/nar/27.6.1421
411 (1999).
412

413 **Tables**

414 **Table 1. Features of predicted SPAR Proteins.**

Gene	Length (aa)	Predicted Gene Function
sparA	424	Hypothetical
sparB	53	Hypothetical
sparC	582	Hypothetical
sparD	656	OLD-Family Endonuclease
sparE	609	GIY-YIG Endonuclease/PcrA/UvrD Helicase
sparF	232	GntR-Family Transcriptional Regulator
sparG	91	Hypothetical, potential T7SS effector
sparH	102	Hypothetical, potential T7SS effector
sparI	137	Hypothetical, potential T7SS effector
sparJ	417	Hypothetical, potential T7SS effector
sparK	199	Hypothetical, potential T7SS effector
sparL	197	Hypothetical

415

416

417 **Figure Legends**

418 **Fig 1**. ***SPAR* Genetic Organization and Expression. A.** The *SPAR* locus. Genes encoding
419 proteins with homology to proteins of known functions or containing characteristics of certain
420 protein families are colored in blue, while genes encoding hypothetical proteins of unknown
421 function are colored in gray. **B. RT-PCR.** RNA was extracted from TX20005 cultured in the
422 presence of HT29, as well as from the colon tissues of mice that had been orally gavaged with
423 TX20005. RNA was treated with DNase and synthesized using a ProtoScript II First Strand
424 cDNA Synthesis Kit (NEB). Newly synthesized cDNA was then used as PCR template to detect
425 *sparA, D,* and *G*.

426

427 **Fig 2**. **SPAR is important for *Sgg* adherence but does not impact growth. A. Adherence**.
428 HT29 cells were seeded at a density of $1x10^6$ cells per well and incubated with TX20005 or
429 TX20005*ΔSPAR* (MOI=10) in media only or in the presence of CS from TX20005 or
430 TX20005*ΔSPAR* for 1 hour. Cells were washed thrice in PBS to remove unattached bacteria,
431 then lysed with 1mL of 0.01% Triton X-100 (Sigma). The lysate was collected, serially diluted,
432 and plated on tryptic soy agar (TSA) plates (Sigma) to be incubated for 24 hours. Colonies were
433 enumerated and percentage adherence was calculated as the percentage of adhered bacteria vs
434 total bacteria added. Results were combined from three biological replicates. Unpaired two-tailed
435 *t* tests were used. **, $p<0.01$; NS, not significant. **B. Growth Curve**. Single colonies of either
436 bacteria were inoculated into 1mL of BHI (Teknova) and incubated at 37ºC overnight with
437 shaking. The overnight culture was then diluted 1:100 with fresh BHI and incubated at 37ºC with
438 shaking. 100μL of the culture were taken at the 0, 3, 6, 9, 12, and 24 hours for serial dilution and
439 platin. Results were combined from three biological replicates.

440

441 **Fig 3**. ***SPAR* is essential for *Sgg* to stimulate CRC cell proliferation. A-D**. **Cell Proliferation**
442 **assay.** HT29 (**A**), HCT116 (**B**), SW480 (**C**), and HEK293 (**D**) cells were seeded at a density of
443 $1x10^4$ cells per well and co-cultured in a 37ºC humidified incubator with 5% CO_2 for 24 hours in
444 the presence or absence of bacteria (MOI=1), as described in the Materials and Methods section.
445 Cell proliferation was measured using a CCK-8 kit. Cell-free wells containing only the culture
446 media were served as a baseline control to which the raw values were normalized. **E-G. Western**
447 **blot**. HT29 cells were seeded at a density of $1x10^6$ cells per well, and co-cultured in the presence
448 or absence of bacteria (MOI=1) for 9 hours. Cell lysates were analyzed by western blot, probed
449 with antibodies against β-catenin, PCNA, and β-actin (**E**). Band intensity was quantified and
450 normalized to β-actin. Fold change in β-catenin (**F**) and PCNA (**G**) were calculated against
451 untreated cells. Data was combined from three biological replicates. **H. Cell proliferation in the**
452 **presence of *Sgg* CS.** HT29 cells were cultured in the presence or absence of bacteria (MOI=1)
453 or with CSs from WT or mutant *Sgg* for 24 hours. Cell proliferation was measured by using a
454 CCK-8 kit. Unpaired two-tailed *t* tests were used for all comparisons. *, $p<0.05$; **, $p<0.01$; ***,
455 $p<0.001$; ****, $p<0.0001$; ns, not significant.

456

457 **Fig 4. *SPAR* deletion reduced the colonization capacity of *Sgg***. This was performed as
458 described in the Materials and Methods section. Colon tissues (**A**) and fecal pellets (**B**) were
459 weighed, homogenized in sterile PBS and dilution plated onto selective agar plates. The non-
460 parametric Mann-Whitney test was used for the comparisons. *, $p<0.05$; **, $p<0.01$; ****,
461 $p<0.0001$; NS, not significant.

462

463 **Fig 5. *SPAR* is important for *Sgg* to promote the development of colon tumors**. This was
464 performed as described in the Materials and Methods section. The procedure is illustrated in **A**.
465 **B-D.** Macroscopic tumors were observed by blinded observers. Tumor incidence (**B**) is shown as
466 the ratio of tumor-bearing mice vs. total mice, presented as a fraction of the total. Tumor number
467 per mouse and average tumor size are shown in **C** and **D**. Tumor burden (**E**) is calculated as sum
468 of tumor volumes per mouse. **F.** *Sgg* load in the fecal pellets collected at the experimental
469 endpoint was determined by dilution plating onto selective agar plates. Mann-Whitney test was
470 used for all comparisons. *, $p<0.05$; **, $p<0.01$; ***, $p<0.001$; ****, p, <0.0001; NS, not
471 significant.

Fig 1

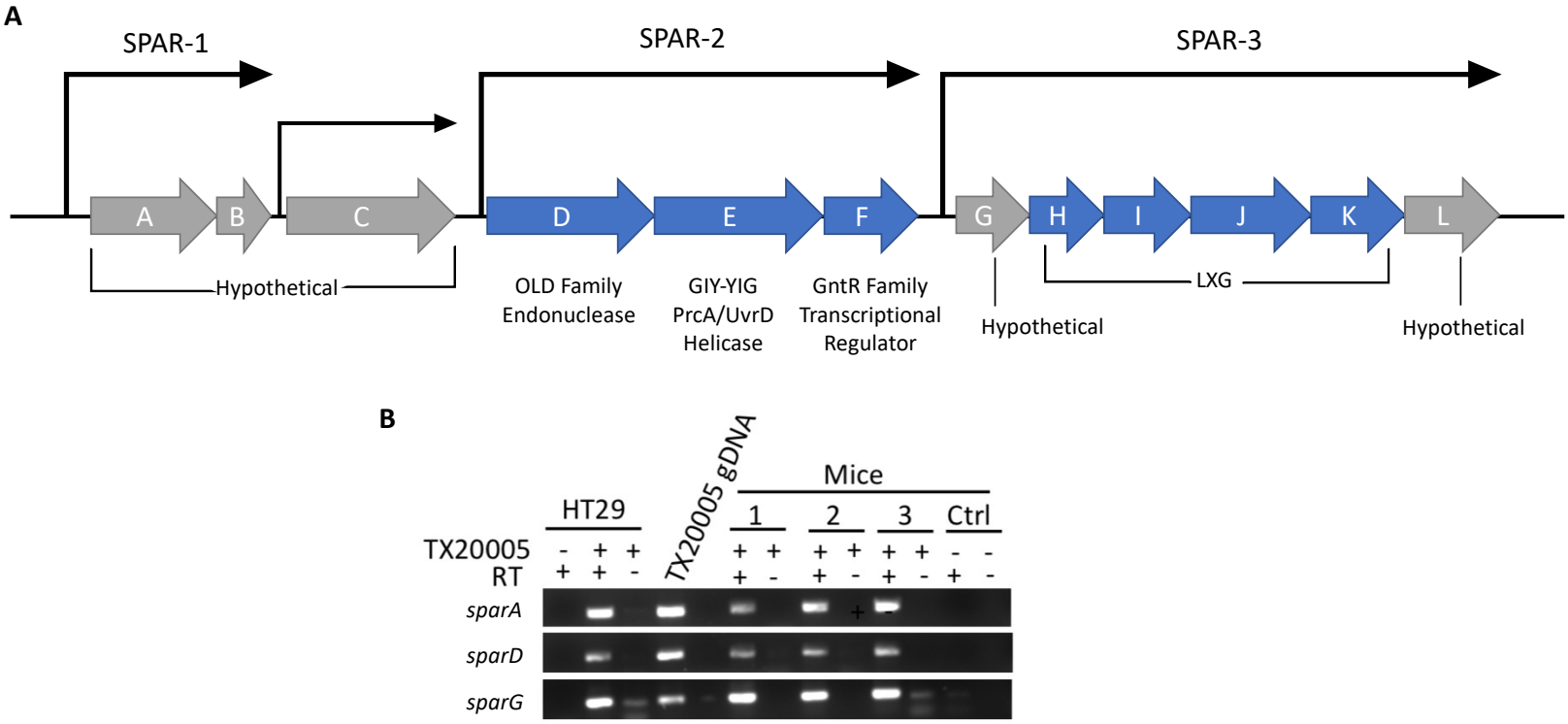
A
SPAR-1
SPAR-2
SPAR-3
A
B
C
D
E
F
G
H
I
J
K
L
Hypothetical
OLD Family
Endonuclease
GIY-YIG
PrcA/UvrD
Helicase
GntR Family
Transcriptional
Regulator
Hypothetical
LXG
Hypothetical
B
HT29
TX20005 gDNA
Mice
1
2
3
Ctrl
TX20005
RT
sparA
sparD
sparG

Fig 2

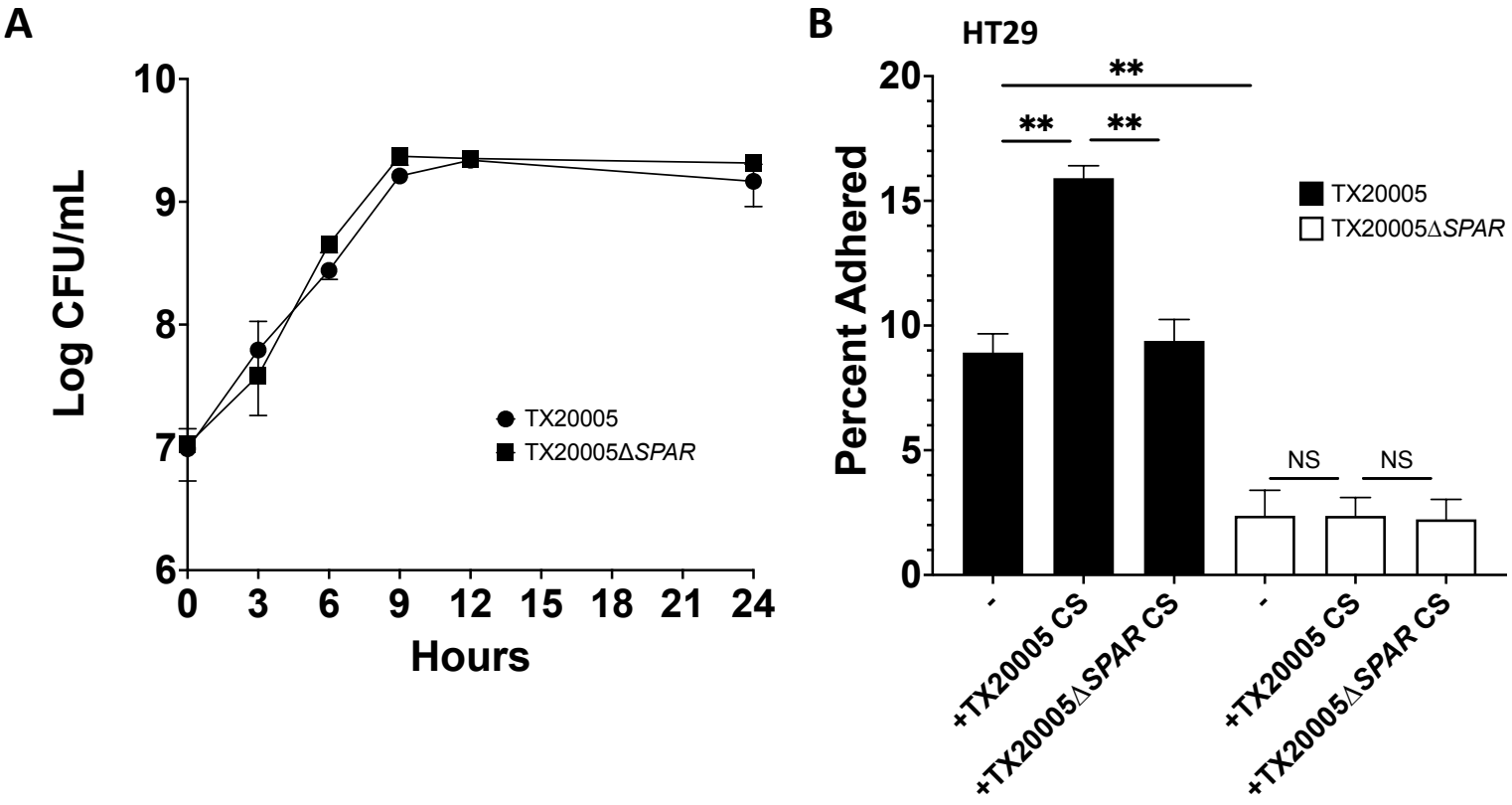
A
Log CFU/mL
10
9
8
7
6
0
3
6
9
12
15
18
21
24
Hours
TX20005
TX20005ΔSPAR
B
HT29
Percent Adhered
20
15
10
5
0
**
**
**
NS
NS
TX20005
TX20005ΔSPAR
-
+TX20005 CS
+TX20005ΔSPAR CS
-
+TX20005 CS
+TX20005ΔSPAR CS

Fig 3

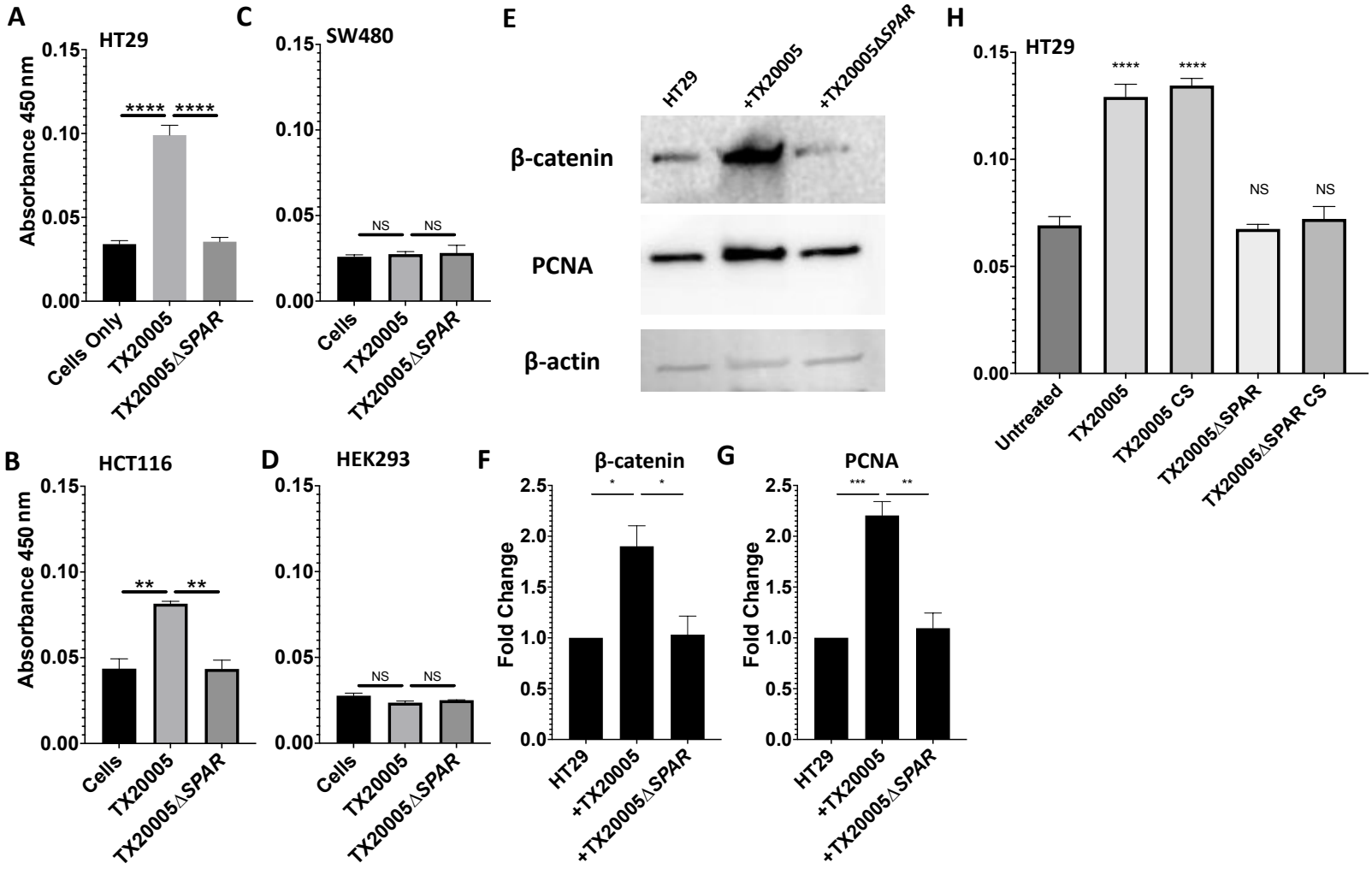
A
HT29
Absorbance 450 nm
0.15
0.10
0.05
0.00
****
****
Cells Only
TX20005
TX20005ΔSPAR
C
SW480
NS
NS
Cells
TX20005
TX20005ΔSPAR
E
HT29
+TX20005
+TX20005ΔSPAR
β-catenin
PCNA
β-actin
H
HT29
****
****
NS
NS
Untreated
TX20005
TX20005 CS
TX20005ΔSPAR
TX20005ΔSPAR CS
B
HCT116
**
**
D
HEK293
NS
NS
F
β-catenin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

Fig 4

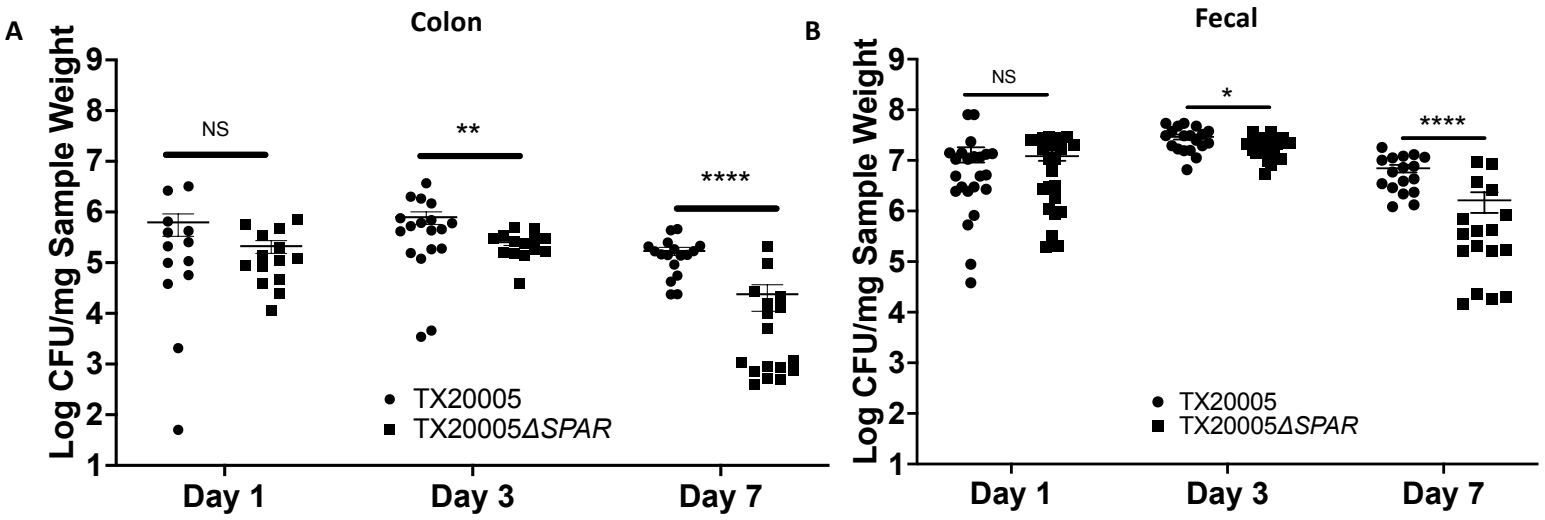
A
Colon
Log CFU/mg Sample Weight
9
8
7
6
5
4
3
2
1
NS
**
****
TX20005
TX20005ΔSPAR
Day 1
Day 3
Day 7
B
Fecal
Log CFU/mg Sample Weight
NS
*
****
TX20005
TX20005ΔSPAR
Day 1
Day 3
Day 7

Fig 5

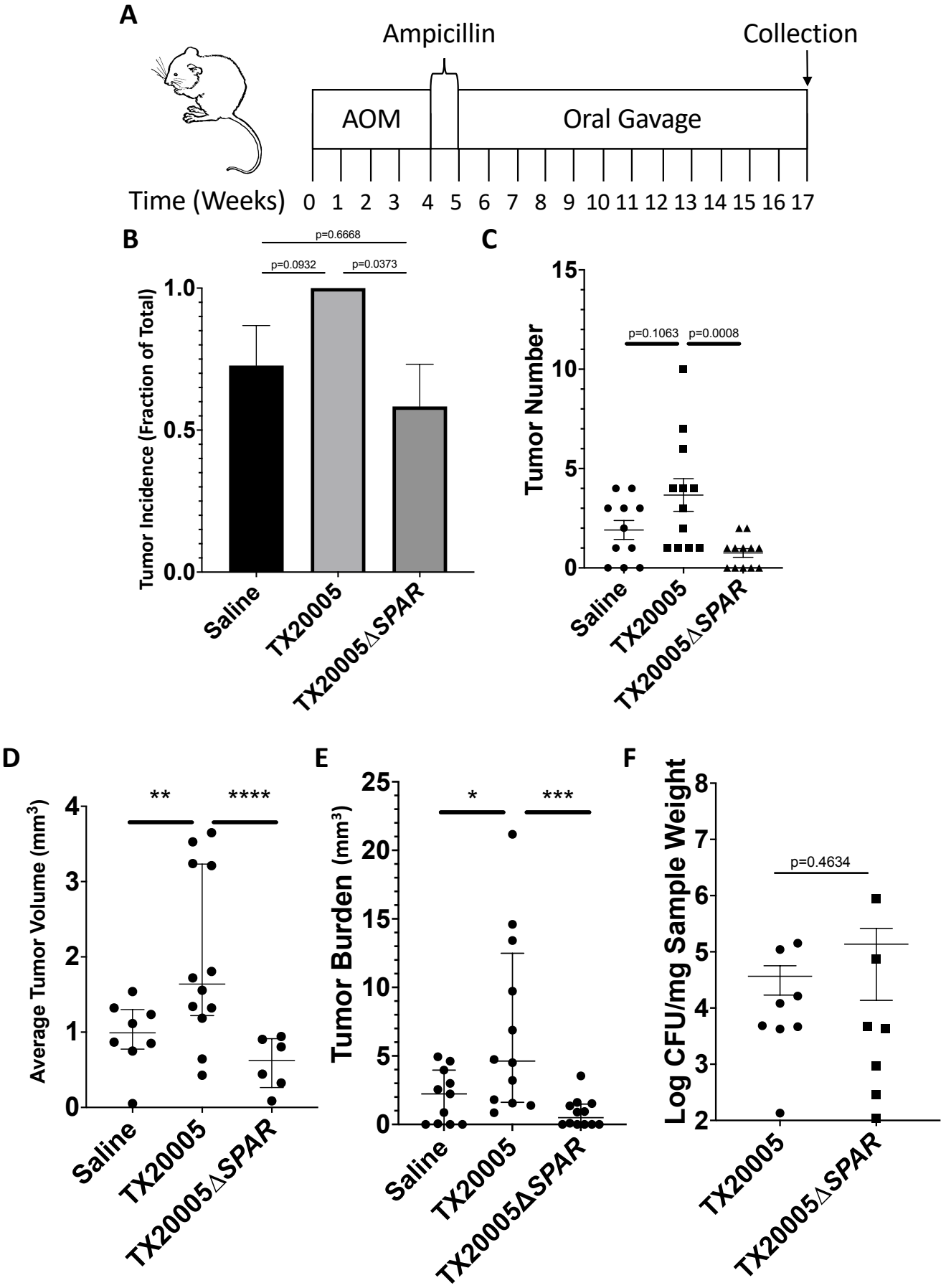
A
Ampicillin
Collection
AOM
Oral Gavage
Time (Weeks) 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
B
p=0.6668
p=0.0932
p=0.0373
Tumor Incidence (Fraction of Total)
1.0
0.5
0.0
Saline
TX20005
TX20005ΔSPAR
C
Tumor Number
15
10
5
0
p=0.1063
p=0.0008
Saline
TX20005
TX20005ΔSPAR
D
Average Tumor Volume (mm³)
4
3
2
1
0
**
****
Saline
TX20005
TX20005ΔSPAR
E
Tumor Burden (mm³)
25
20
15
10
5
0
*
***
Saline
TX20005
TX20005ΔSPAR
F
Log CFU/mg Sample Weight
8
7
6
5
4
3
2
p=0.4634
TX20005
TX20005ΔSPAR

Fig S1

```
sparH      GETDEALAELKKQEDDLEDALVDLLRKEAQDEH
sparI      DQQQQCLNVLEEAELQLSLISVMSLERKTSLHQ
sparJ      EVKTTIFSSLVPKEVVAINPVTGAPIYASVSEDEAMAYLQENYEGSYLFIQSLEEEENLFGEDN
sparK      GDDLLTQLSAYGKDLDWLQTTSQTVLEDDILGLWFKKGSHRYSLNNLGDLTITYDDVLNDN
Consensus            HxxxDxxhxxxH
                          E
```

Fig S1: The conserved T7SS sequence in SparH-K. H denotes highly conserved and h less conserved hydrophobic residues