GATA4 regionalizes intestinal metabolism and barrier function to prevent immunopathology

Zachary M. Earley[1,2], Wioletta Lisicka[1,2], Joseph J. Sifakis[3], Raúl Aguirre-Gamboa[2], Anita Kowalczyk[2], Jacob T. Barlow[4], Dustin G. Shaw[1,2], Valentina Discepolo[5], Ineke L. Tan[6], Saideep Gona[7], Jordan D. Ernest[2], Polly Matzinger[8], Luis B. Barreiro[1,2,7], Andriy Morgun[9], Albert Bendelac[1,10], Rustem F. Ismagilov[4,11], Natalia Shulzhenko[12], Samantha J. Riesenfeld[13], Bana Jabri[1,2,10,*]

Affiliations:

[1] Committee on Immunology, University of Chicago, Chicago, IL, USA

[2] Department of Medicine, University of Chicago, Chicago, IL, USA

[3] Department of Chemistry, University of Chicago, Chicago, IL, USA

[4] Division of Biology and Biological Engineering, California Institute of Technology, CA, USA

[5] Department of Medical Translational Sciences and European Laboratory for the Investigation of Food Induced Diseases, University of Federico II, Naples, IT

[6] Department of Gastroenterology and Hepatology, University of Groningen and University of Medical Center Groningen, Groningen, NL

[7] Genetics, Genomics, and Systems Biology, University of Chicago, Chicago, IL, USA

[8] Ghost Lab. National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA

[9] College of Pharmacy, Oregon State University, Corvallis, OR, USA

[10] Department of Pathology, University of Chicago, Chicago, IL, USA

[11] Division of Chemistry and Chemical Engineering, California Institute of Technology, CA, USA

[12] Department of Biomedical Sciences, Oregon State University, Corvallis, OR, USA

[13] Pritzker School of Molecular Engineering, University of Chicago, Chicago, IL, USA

* Correspondence: bjabri@bsd.uchicago.edu (B.J.)

Summary

Different regions of the gastrointestinal tract have distinct digestive and absorptive functions, which may be locally disrupted by infection or autoimmune disease. Yet, the mechanisms underlying intestinal regionalization and its dysregulation in disease are not well understood. Here, we used mouse models, transcriptomics, and immune profiling to show that regional epithelial expression of the transcription factor GATA4 prevented adherent bacterial colonization and inflammation in the proximal small intestine by regulating retinol metabolism and luminal IgA. Loss of epithelial GATA4 expression increased mortality in mice infected with *Citrobacter rodentium*. In active celiac patients with villous atrophy, low GATA4 expression was associated with metabolic alterations, mucosal *Actinobacillus,* and increased IL-17 immunity. This study reveals broad impacts of GATA4-regulated intestinal regionalization and highlights an elaborate interdependence of intestinal metabolism, immunity, and microbiota in homeostasis and disease.

One-sentence summary: Epithelial GATA4 regulates intestinal regionalization of bacterial colonization, metabolic pathways, and tissue immunity.

51 **Introduction**

52 Each region of the gastrointestinal tract plays a distinct physiological role, featuring
53 specific metabolic functions, host-microbe interactions, and immune responses. The duodenum
54 and jejunum of the proximal intestine are optimized to digest and absorb critical nutrients.
55 Despite having a large surface area in close contact with luminal contents, these regions also
56 exhibit relatively low microbial growth and adhesion (*1*). In contrast, the ileum of the distal
57 small intestine has shorter villi, less optimized for nutrient uptake, and instead specializes in
58 reabsorption of bile acids and vitamin B12 (*1*). The ileum harbors more adherent commensal
59 microbes, such as *segmented filamentous bacteria* (SFB), which promote local T_H17
60 differentiation and inflammatory IL-17 responses and protect the host from enteropathogenic
61 infections (*2-4*). Finally, the colon is characterized by a thick mucus layer that permits
62 absorption of water, electrolytes, and numerous microbial metabolites, while preventing the
63 microbiota from invading the intestinal epithelium and driving inflammatory immunity (*1, 5, 6*).
64 Together, these patterns suggest that each region uniquely balances its critical digestive and
65 barrier immune functions, yet the regulatory mechanisms controlling regionalization remain
66 elusive.

67 One known regulator of intestinal regionalization is the transcription factor GATA4,
68 which is selectively expressed in the intestinal epithelial cells (IECs) of the duodenum and
69 jejunum (*7, 8*). Deleting GATA4 in IECs was shown to alter lipid absorption in the jejunum and
70 drive major transcriptional changes, including down-regulation of jejunal-specific genes and up-
71 regulation of ileal-specific genes (*8*). These results suggest that epithelial GATA4 has a profound

72 role controlling numerous cell types in the tissue, a regulatory circuit that has not yet been
73 dissected.

74 Using GATA4-deficient mice, we investigated how GATA4 controls barrier function and
75 bacterial colonization in the jejunum through its key metabolic functions. We also assessed the
76 impact of disrupted intestinal regionalization on host susceptibility to pathology, both in the
77 context of enteric bacterial infection and human celiac disease, a dietary-induced autoimmune-
78 like disorder of the proximal small intestine.

79

80 **Results**

81 **GATA4 regulates regionalization of intestinal immunity in a microbiota-dependent manner**

82 Analysis of cytokine production in T cells from the intestinal track of specific-pathogen-
83 free (SPF) and germ-free (GF) mice, revealed regional-specific immune regulation (Fig. 1A). For
84 instance, despite the microbial load being considerably higher in the colon than in the ileum,
85 inflammatory microbiota-dependent IFN-γ^+ and IL-17^+ T cell responses were significantly higher
86 in the ileum. In contrast, regulatory IL-10^+ T cell responses were found at similar levels in the
87 ileum and colon, and were dependent on the presence of the microbiota in both locations.
88 Furthermore, the microbiota was required for the presence of intraepithelial IFN-γ^+ CD4+ T cells
89 in the ileum but not the proximal small intestine.

90 To determine how IECs shape regional tissue immunity, we analyzed mice in which
91 GATA4 was selectively depleted in IECs (GATA4fl/fl villin-cre+ or GATA4$^{\Delta IEC}$) (*8*). First, we
92 confirmed that GATA4 expression was restricted to duodenal and jejunal IECs, and that, in its
93 absence, jejunal IECs acquired an ileal-like transcriptional program (S1A, B). GATA4 strongly

94 repressed ileal-specific genes (*Fabp6*, *Slc10a2*) involved in the enterohepatic circulation of bile
95 acids (Fig. S1C). In contrast, GATA4 induced expression of lipid metabolic genes in the jejunum
96 involved in retinol metabolism (*Adh1*), and fat digestion and absorption (*Cd36*, *Fabp1*, *Dgat2*,
97 *Apoa4*). GATA4 also played an important role in controlling jejunal epithelial uptake of vitamins
98 and folate (Slc46a1, Pdxk) (Fig. S1C). To systematically assess the impact of GATA4 on regional
99 tissue functions, we performed total tissue RNA sequencing in GATA4$^{\Delta IEC}$ and littermate control
100 wild-type (WT) mice (GATA4fl/fl villin-cre-). Our analysis identified what we term "region-
101 specific, GATA4-regulated" genes, which are defined by expression that distinguishes "ileum-
102 like" tissue, *i.e.,* WT ileum and GATA4$^{\Delta IEC}$ jejunum, from WT jejunum (Fig. 1B, C). Intersecting
103 region-specific, GATA4-regulated genes with public databases of immune genes (*9*) revealed
104 distinct immune signatures of WT jejunum and ileum, a regionalization of tissue immunity that
105 was lost in GATA4$^{\Delta IEC}$ mice (Fig. 1B). Indeed, among the 625 regional-specific, GATA4-regulated
106 immune genes, more than a third (238) were published potential targets of IFNγ or IL-17
107 regulation (Fig. 1C).

108 Consistent with GATA4-regulated regionalization of IFNγ and IL-17 immune pathways, in
109 the absence of epithelial GATA4, the frequency of intraepithelial IFNγ^{+} cells among CD8αβ T
110 cells in the jejunum increased to the levels observed in the ileum (Fig. 1D). Furthermore, GATA4
111 deficiency led to a heightened T_H17 response in the jejunum, with frequencies of IL-17^{+} cells
112 among CD4 T cells surpassing those in the ileum (Fig. 1E). The heightened IFNγ and IL-17
113 immune responses observed in the jejunum of GATA4$^{\Delta IEC}$ mice were generally microbiota
114 dependent (Fig. 1F, S1D-F). Intriguingly, though, some particular immune genes, such as *Csf2,*
115 which encodes for GM-CSF, the anti-viral response genes *Ifnar2* and *Mavs*, and the tissue

116 alarmins IL-33 and IL-15, were GATA4 regulated in a microbiota-independent manner (Figure
117 1F). Among epithelial, region-specific GATA4-regulated genes, the microbiota-independent
118 subset was enriched, compared to the microbiota-dependent subset, in direct targets of GATA4
119 (Fig. S1G), as indicated by GATA4 binding of promoter regions in published ChIP-seq data
120 (hypergeometric distribution test; p-value = 3.21E-80) (Fig. S1G) (*10, 11*). Taken together, these
121 results demonstrate that GATA4 in IECs regulated regional tissue immunity and blocked
122 development of microbiota-dependent inflammatory immune responses in the proximal
123 intestine.
124
125 **GATA4 shapes regional immunity by controlling colonization of *segmented filamentous***
126 ***bacteria***
127 To investigate which specific microbiota may trigger inflammatory immune responses in
128 the absence of GATA4 in the proximal small intestine, we performed 16S ribosomal RNA
129 sequencing of luminal- and mucosal-associated bacterial communities in the jejunum and ileum
130 of littermate control WT and GATA4$^{\Delta IEC}$ mice. This analysis revealed that in the jejunum, SFB
131 (*Candidatus arthromitus*) loads were much higher in GATA4$^{\Delta IEC}$ relative to WT mice (Fig. 2A,
132 S2A). Furthermore, GATA4 was required to prevent SFB from colonizing (Fig. 2A, S2A) and
133 adhering (Fig. 2B) to IECs in the jejunum. In contrast, and in accordance with the lack of GATA4
134 expression in the ileum, no changes in bacterial composition were observed in the ileum of
135 GATA4$^{\Delta IEC}$ versus WT mice (Fig. S1A).
136 Previous studies showed that SFB triggers the upregulation of various adaptive immune
137 pathways (*2, 12-14*). To determine whether jejunal colonization by SFB was responsible for the

138 significant increase in inflammatory T cell immunity observed in the jejunum of SPF GATA4$^{\Delta IEC}$
139 mice (Fig.1 C-E), we performed a series of microbial transfers. First, transplantation of
140 GATA4$^{\Delta IEC}$ and littermate control WT microbiota into GF WT and GATA4$^{\Delta IEC}$ hosts, respectively,
141 showed that the host genotype determined the immune outcome irrespective of the input
142 microbial community (Figure S2B, C). Furthermore, transfer of either altered Schaedler flora
143 (ASF), a defined bacterial community that lacks SFB, or jejunal microbiota from a WT donor
144 within our colony with low-to-undetectable levels of SFB, failed to induce inflammatory T cells
145 in GATA4$^{\Delta IEC}$ mice. These results suggest that SFB is the causative microbe for dysregulated
146 immunity in GATA4$^{\Delta IEC}$ mice (Figure S2B, C). Finally, transfer of fecal homogenates from Jackson
147 C57BL/6J mice with or without SFB supplemented demonstrated that the increases in both
148 IFNγ^{+} CD8αβ T cells and T_H17 cells observed in the jejunum of GATA4$^{\Delta IEC}$ mice (Fig. 1D-E), were
149 dependent on the presence of SFB (Fig. 2C, D).

150 Since SFB colonization is known to induce an antigen-specific T_H17-cell response against
151 the 3340 epitope of SFB (*15*), we asked if the increased T_H17 cell response observed in the
152 jejunum was a consequence of altered T cell priming. To address that question, we transferred
153 congenically marked CD45.1^{+} 7B8^{+} CD4 T cells, specific to the 3340 epitope of SFB, into CD45.2
154 WT and GATA4$^{\Delta IEC}$ hosts. In the absence of GATA4, 7B8^{+} CD4 T cells differentiated into RORγt^{+}
155 Foxp3^{-} CD4 T cells in both the jejunal and ileal draining mesenteric lymph nodes (MLN) (Fig. 2E),
156 indicating a change in regional T cell priming against SFB in the MLN. Furthermore, nine days
157 after transfer, 7B8^{+} T cells expanded (Fig. 2F) and downregulated expression of their TCR in the
158 jejunum of GATA4$^{\Delta IEC}$, versus WT, mice (Fig. 2G). Together, these results indicate that, through

159 GATA4, IECs regulate regional colonization of the commensal microbe SFB and SFB-specific T-
160 cell priming.

161 To establish whether colonization of the jejunum by SFB depended on its ability to
162 adhere to IECs, we monocolonized GF WT and GATA4$^{\Delta IEC}$ mice with rat SFB, which is 86%
163 genetically identical to mouse SFB but unable to adhere to mouse IECs and thereby induce a
164 T_H17 response (*12, 16*). Rat SFB was unable to expand and induce inflammatory T cell responses
165 in the jejunum of GATA4$^{\Delta IEC}$ mice (Figure 2H-J). To further test the hypothesis that GATA4
166 prevents colonization of the jejunum by adherent commensals, we transplanted GATA4$^{\Delta IEC}$ mice
167 with ASF containing the mucosal-associated commensal microbe *Mucispirillum ASF 457*.
168 Consistent with the hypothesis, ASF colonization of GF GATA4$^{\Delta IEC}$ and WT mice led to an
169 expansion *Mucispirillum ASF 457* in the jejunum of GATA4$^{\Delta IEC}$ but not WT mice (Fig. S2D).

170

171 **GATA4 regulates luminal IgA levels to control colonization by SFB**

172 We next sought to understand how GATA4 in IECs restricts adherent bacterial
173 colonization in the proximal intestine. Expansion of SFB was reported in the ileum and feces of
174 IgA deficient mice (*17, 18*). Furthermore, a previous report revealed that B-cell deficient mice
175 display lipid metabolic defects in the jejunum, as well as a gene expression signature associated
176 with GATA4$^{\Delta IEC}$ mice (*19*). Together, these studies suggest the possibility that an IgA defect in
177 GATA4$^{\Delta IEC}$ mice controls SFB colonization. In accordance with this hypothesis, SFB
178 monocolonization of GF B-cell deficient Jh$^{-/-}$ mice and GF IgA$^{-/-}$ mice substantially increased SFB
179 loads in the jejunum compared to those observed in the ileum of littermate control mice
180 indicating that IgA regulates regionalization of SFB colonization (Fig. 3A, B, S3A, B).

181 Given that B cells and IgA prevented SFB from colonizing the jejunum, we asked
182 whether GATA4 may control SFB colonization by regulating the regional distribution of IgA^+
183 plasma cells in the small intestine. We observed that the jejunum contained approximately
184 three times as many IgA^+ $B220^-$ plasma cells as the ileum (Fig. 3C), and that the higher numbers
185 of IgA-producing plasma cells were associated with a greater capacity to produce IgA in tissue
186 explants (Fig. S3C). This regionalization of IgA response was GATA4-dependent, as evidenced by
187 significantly reduced numbers of IgA^+ $B220^-$ plasma cells (Fig. 3C) and IgA production in the
188 jejunum of $GATA4^{\Delta IEC}$ versus WT mice (Fig. S3C). The overall result was a substantial decrease in
189 luminal secretory IgA in both the jejunum and the ileum of $GATA4^{\Delta IEC}$ mice (Fig. 3D). Reduced
190 luminal IgA was also observed in the jejunum and ileum of GF $GATA4^{\Delta IEC}$ mice, indicating that
191 the reduction is microbiota independent (Fig. S3D). Moreover, free IgA in the jejunal luminal
192 content from $GATA4^{\Delta IEC}$ mice had less capacity than that of littermate-control WT mice to coat
193 the microbiota present in the feces of immune-deficient ($RAG^{-/-}$) mice (Fig. 3E and S3E). Taken
194 together, these results indicate that GATA4 blocks the regionalization of adherent microbes,
195 such as SFB, by increasing luminal IgA.

196 We hypothesized that epithelial GATA4 mediated IgA levels by controlling region-
197 specific metabolic processes. To identify potential candidates, we performed gene set
198 enrichment analysis in our epithelial RNA-seq data (Fig. S1B) using the KEGG database, which
199 revealed retinol metabolism as a top GATA4-dependent, region-specific pathway (Fig. S3F).
200 Genes in the retinol metabolic pathway were expressed higher in the jejunum of WT, but not
201 $GATA4^{\Delta IEC}$, mice, compared to WT ileum (Fig. 3F), supporting previous reports that the proximal
202 intestine facilitates greater vitamin A uptake and metabolism (*20, 21*). Published ChIP-seq data

203 (*10, 11*) shows direct binding of GATA4 to the promoters of 23/35 (66%) of the differentially
204 expressed epithelial genes in the retinol metabolism pathway (Fig. 3F), suggesting direct
205 transcriptional regulation of retinol metabolism in IECs. Concordantly, in the absence of GATA4,
206 jejunal epithelial cells exhibited impaired aldehyde dehydrogenase (ALDH) activity, indicating a
207 decreased capacity to produce retinoic acid (RA) (Fig. 3G).

208 RA was previously shown to regulate intestinal B cell responses, including upregulating
209 intestinal homing receptors on B cells and IgA class switching (*22*). In addition, epithelial
210 RARα/β reportedly regulates the number of IgA-producing B cells (*23, 24*). Therefore, we asked
211 whether retinol deficiency was sufficient to enable the SFB expansion in the proximal small
212 intestine that we observed in GATA4$^{\Delta IEC}$ mice. We fed GF WT mice a control or vitamin-A
213 deficient diet for 4 weeks post-weaning and subsequently colonized them with SFB. SFB
214 colonized the jejunum of mice fed a vitamin-A deficient diet but failed to colonize the mice fed
215 a control diet (Fig 3H). To determine whether the defect in luminal IgA in the small intestine of
216 GATA4$^{\Delta IEC}$ mice could be rescued with exogenous RA, we injected GATA4$^{\Delta IEC}$ mice
217 intraperitoneally with RA for two weeks. RA significantly restituted luminal IgA in the jejunum
218 of GATA4$^{\Delta IEC}$ mice (Fig. 3I). We also tested whether exogenous luminal IgA was sufficient to
219 prevent colonization of the jejunum by SFB in GATA4$^{\Delta IEC}$ mice. Polyclonal luminal sIgA, capable
220 of strongly coating microbes from RAG-/- feces, was isolated from the intestinal contents of WT
221 mice with protein L magnetic beads (Figure S3G). This luminal polyclonal IgA or phosphate-
222 buffered saline (PBS) was gavaged to GF WT or GATA4$^{\Delta IEC}$ mice prior to and after colonization
223 with SFB (Fig. S3H). This IgA was indeed sufficient to prevent SFB from colonizing the jejunum
224 and ileum of GATA4$^{\Delta IEC}$ mice (Fig. 3J, S3I). These data indicate that, by regulating regional

225 retinol metabolism in IECs, GATA4 regulates luminal IgA, which in turn restricts bacterial
226 adhesion to the proximal small intestine.
227
228 **Loss of GATA4 control of colonization by an adherent pathogen leads to severe**
229 **immunopathology**
230 Since GATA4 controlled intestinal colonization by adherent commensal microbes, we
231 tested whether it also regulated colonization by adherent pathogens. *Citrobacter rodentium* is
232 an adherent pathogen that preferentially colonizes the colon but not the small intestine of WT
233 mice (Fig. 4A). In the absence of GATA4, the niche for *C. rodentium* was altered such that the
234 pathogen colonized the small intestine at levels approaching those in the colon of WT mice (Fig.
235 4A). Furthermore, the colonization in the small intestine of GATA4$^{\Delta IEC}$ mice depended on *C.*
236 *rodentium's* ability to adhere to IECs. Mutant ΔEAE *C. rodentium,* which lacks the gene intimin
237 required for adherence to IECs, did not colonize the jejunum and ileum of GF GATA4$^{\Delta IEC}$ mice
238 beyond levels observed in GF WT mice infected with WT *C. rodentium* (Fig. 4B). These results
239 suggest that GATA4 controls regional colonization of the small intestine not only by commensal
240 bacteria, but also by adherent pathogenic microbes.
241 We tested whether the expansion of *C. rodentiu*m colonization to the small intestine of
242 GATA4$^{\Delta IEC}$ mice impacted host immune response or survival. In accordance with the change in
243 colonization pattern, *C. rodentiu*m infection induced a striking expansion of IL-17^{+} IFNγ^{-}, IL-17^{-}
244 IFNγ^{+}, and IL-17^{+} IFNγ^{+} double-positive CD4 T cells in the colon and ileum of GATA4$^{\Delta IEC}$ (Fig. 4C).
245 This contrasted with WT littermate control mice, which only showed a moderate T_H17 and T_H1
246 expansion in the colon (Figure 4C). The immune response in GATA4$^{\Delta IEC}$ mice was accompanied

247 by the development ten days after infection of severe colitis and villous atrophy in the ileum
248 (Fig. 4D), symptoms not normally associated with the infection in WT. Once the intestinal
249 barrier was compromised in GATA4$^{\Delta IEC}$ mice, *C. rodentium* translocated to systemic sites,
250 including the mesenteric lymph nodes, liver, and spleen (Fig. 4E), highlighting the substantial
251 impacts on distal sites that may be triggered by region-specific intestinal defects. By day 12
252 post-infection, 87.5% of GATA4$^{\Delta IEC}$ mice had died (Fig. 4F), punctuating the critical role of
253 GATA4-dependent intestinal regionalization in controlling host disease susceptibility to an
254 enteric pathogen.
255
256 **Loss of GATA4 expression in celiac disease is associated with increased IL-17 immunity and**
257 **changes in bacterial colonization**
258 Celiac disease (CeD) is an auto-immune–like T_H1-mediated enteropathy of the
259 duodenum caused by dietary gluten in genetically susceptible individuals (*25*). A long-standing
260 conundrum has been the increase of gluten-dependent, but not gluten-specific, duodenal T_H17
261 responses in CeD patients (*26-28*). Previous observations noted alterations in IECs (*25, 29*), as
262 well as lipid (*30*) and vitamin deficiencies (*31*) during active CeD (ACeD). We therefore
263 hypothesized that ACeD patients have decreased expression of GATA4 in IECs, and that this
264 decrease may be associated with reported changes in microbiota (*32, 33*) and increased IL-17
265 immunity. To investigate this possibility, we compared the transcriptional profiles of duodenal
266 biopsies from ACeD patients, CeD patients on a gluten free diet (GFD), and control patients with
267 non-inflammatory intestinal symptoms that required upper endoscopies. Intriguingly, our
268 analysis revealed significantly lower *Gata4* expression in ACeD patients versus controls, which

269 was restored by GFD (Fig. 5A). Immunohistochemistry staining demonstrated an ACeD-
270 associated loss of GATA4 protein production specifically in apical epithelial cells, whereas
271 intestinal crypts cells retained GATA4 production (Fig. 5B). To gain insights into the impact of
272 GATA4 in ACeD, we identified genes that were differentially expressed between "GATA4-hi"
273 individuals, defined by *Gata4* expression above the 70th percentile of the entire cohort, and
274 "GATA4-lo" ACeD patients, defined by *Gata4* expression below the 30th percentile of ACeD
275 patients (Fig. S4A). GATA4-lo specific genes were enriched in mouse genes specific to ileum-like
276 but not jejunum tissue. Furthermore, GATA4-hi specific genes were significantly enriched in
277 mouse genes specific to jejunum but not ileum-like tissue (Fig. 5C, S4B). Thus, during ACeD,
278 intestinal regionalization may be lost as the duodenum decreases GATA4-dependent jejunum-
279 specific gene expression and increases ileum-specific gene expression.

280 Many of the genes driving these enrichments are direct targets of GATA4 and involved
281 in lipid metabolic processes, such as cholesterol absorption and retinol metabolism (Fig. 5D).
282 Compared to GATA4-hi patients, GATA4-lo ACeD patients demonstrated decreased enrichment
283 in retinol metabolism genes and increased enrichment in IL-17 signaling genes (Fig. 5E). In fact,
284 expression of these pathways was correlated (or anti-correlated, in the case of IL-17 signaling
285 genes) with *GATA4* expression across all ACeD patients (Fig S4C-F). Together, these data reveal
286 that GATA4 expression can be lost in ACeD, and that this loss may play a role in regional
287 defects, such as low retinol metabolism, low plasma cholesterol (*34*), and increased IL-17
288 immunity.

289 Studies in mice revealed that the presence of T_H17 cells in the ileum is dependent on the
290 microbiota (*2*) and that bacteria can drive distinct immune responses depending on the host

291 context (*35*). Therefore, we asked whether a particular microbe was associated with the IL-17
292 response in ACeD patients. To answer this, we leveraged a quantitative framework to detect
293 absolute abundances of individual bacterial taxa in duodenal biopsies in which host
294 transcriptional analysis is performed concomitantly (*36*). Digital PCR anchoring of 16S rRNA
295 amplicon sequencing revealed no change in overall bacterial load in ACeD patients versus
296 controls (Fig. S4G) and only a trend of stratification between ACeD and control patients in the
297 second and third principal components (Fig. S4I). We did observe an expansion of *Neisseria* in
298 ACeD patients, the only microbial change significantly associated with disease state (Fig. S4H, J),
299 confirming a previous report (*37*). The increased abundance of *Neisseria* was not, however,
300 associated with decreased *Gata4* expression, metabolic defects, or IL-17 signaling in ACeD
301 patients (Fig. 5F).

302 In contrast, we found that *Actinobacillus* was associated with lower *Gata4* expression,
303 lower retinol metabolism, and higher IL-17 signaling in ACeD, but not control, patients (Fig. 5F,
304 G, S4K). Of note, we did not observe a significant association with the prototypical gluten-
305 specific T_H1 IFNγ pathway in ACeD patients, suggesting that *Actinobacillus* may play a specific
306 role in IL-17 immunity (Fig. 5F). Intriguingly, *Actinobacillus* was also associated with other
307 metabolic defects in ACeD patients, such as in xenobiotic, bile acid, and heme metabolism,
308 whereas other microbes, such as *Streptococcus,* had inverse patterns and positive associations
309 with bile and heme metabolism. The discovery of mucosal-associated bacteria, such as
310 *Actinobacillus*, associated with the loss of GATA4, metabolic dysfunction, and IL-17 immunity in
311 ACeD patients, highlights the potential importance in the pathogenesis of celiac disease of
312 GATA4 regulation of host microbial interactions.

313

314

315 **Discussion**

316 In this report, we identify mechanisms controlling regionalization of the proximal and
317 distal small intestine and reveal the importance of this segregation in homeostasis and disease.
318 We found that in part by regulating retinol metabolism and local IgA response, the epithelial
319 transcription factor GATA4 plays a critical role in limiting bacterial adhesion in the jejunum,
320 which may allow the host to focus on digestion and absorption of nutrients. On the other hand,
321 the ileum, lacking GATA4 expression and producing less IgA permits adherent microbes, such
322 as SFB, to induce inflammatory IL-17^{+} T cell responses required for controlling pathogenic
323 infections (*2*). The intriguing association between loss of GATA4 expression in IEC and microbial-
324 associated dysregulated IL-17 immune responses in active CeD, poses the intriguing question of
325 whether a decrease in GATA4 in IECs of the proximal intestine may play a role in other complex
326 immune disorders by altering host-microbial interactions and triggering dysregulated immune
327 responses to bacteria.

328

329 **Figure legends**

330 **Figure 1. Intestinal epithelial cells control microbiota-dependent and independent**
331 **regionalization of tissue immunity through GATA4.**

332 (**A**) Percentage (y axis) of cells stained for cytokines IFNγ, IL-17, or IL-10 (panel titles), as
333 measured by intracellular flow cytometry, among all CD8αβ (leftmost panel) or CD4 (other
334 panels) T cells isolated from intraepithelial lymphocytes (IELs, left set) or the lamina propria (LP,

335 right set) of each intestinal segment (x axis) of SPF (orange circle) or GF (purple square) mice
336 and then stimulated by phorbol myristate acetate (PMA) and ionomycin. ****, *P*<0.0001,
337 effect due to region; ^{oooo} *P*<0.0001, ^{ooo} *P*<0.001, ^{oo} *P*< 0.01, effect due to microbiota; two-way
338 ANOVA of microbiota and region impact on cytokine levels *N*= 5-7 mice/group.
339 (**B**) Top two principal components (PCs) (x, y axes) of the expression of region-specific, GATA4-
340 regulated genes, based on RNA-seq of tissue samples (dots) from the jejunum (dark colors) and
341 ileum (light colors) of WT (blues) and GATA4$^{\Delta IEC}$ (reds) mice. *N*= 8 mice/group.
342 (**C**) Heatmap of the z-scored expression (color) in tissue samples of region-specific, GATA4-
343 regulated immune genes (rows) from the jejunum and ileum (columns) of WT (left set) and
344 GATA4$^{\Delta IEC}$ (right set) mice. Of 625 total genes, 145 are in only the IFNγ module, 54 in only the IL-
345 17 module, 39 in both modules, and 387 in neither (annotation column) (Methods). *N*= 8
346 mice/group.
347 (**D**) Representative (left) and summary (right) plots of the frequencies of IFN-γ^{+} cells among
348 CD8αβ T cells in the IELs of the jejunum and ileum of WT and GATA4$^{\Delta IEC}$ mice. **** *P*<0.0001,
349 paired t-test. *N*= 6 mice/group.
350 (**E**) Representative (left) and summary (right) plots of the frequencies of IL-17a^{+} cells among
351 CD4 T cells in the LP of the jejunum and ileum of WT and GATA4$^{\Delta IEC}$ mice. *** *P*<0.001, paired t-
352 test. *N*= 6 mice/group
353 (**F**) Heatmap of the z-scored expression (color) in tissue samples of 50 selected microbiota
354 dependent and independent (right annotation column), region-specific, GATA4-regulated
355 immune genes (rows) in the jejunum (columns) of SPF (left set) and GF (right set) WT and
356 GATA4$^{\Delta IEC}$ mice. Gene modules (left annotation column) as in C.

357

358 **Figure 1 supplemental. GATA4 controls regionalization of tissue metabolism and immunity**

359 (**A**) Expression of GATA4 (y axis), as measured by qPCR relative to GAPDH, in the tissue of all

360 intestinal regions (x axis).

361 (**B**) Top two principal components (x, y axes) of the expression of region-specific, GATA4-

362 regulated genes in samples (dots) of sorted Epcam^{+} CD45^{-} IECs isolated from the jejunum (dark

363 colors) and ileum (light colors) of WT (blues) and GATA4$^{\Delta IEC}$ (reds) mice.

364 (**C**) Heatmap of z-scored expression (color) in IECs of the jejunum and ileum of WT and

365 GATA4$^{\Delta IEC}$ mice (columns) of selected DE genes (rows) from the following metabolic pathways:

366 metabolism of steroids (Reactome), retinol metabolism (KEGG), fat digestion and absorption

367 (KEGG), and metabolism of vitamins and cofactors (Reactome).

368 (**D**) Heatmap of the z-scored expression (color) of all region-specific, GATA4-regulated genes

369 (rows) in jejunum tissue (columns) of SPF (left set) and GF (right set) WT and GATA4$^{\Delta IEC}$ mice.

370 (**E** and **F**) Frequency (y axis) of IFNγ^{+} cells among CD8αβ T cells (**E**) or of IL-17a^{+} cells among

371 CD4 T cells (**F**) in the jejunum (left panel) and ileum (right panel) of SPF and GF WT and

372 GATA4$^{\Delta IEC}$ mice (bar color). ****P<0.0001 , *** P<0.001, ** P<0.01, * P<0.05, ANOVA with

373 Tukey multiple comparison test N= 6 mice/group.

374 (**G**) Frequency of genes bound (red) or not bound (gray) by GATA4, as determined by GATA4

375 ChIP-seq, among microbiota dependent (top) or microbiota independent (bottom) genes, as in

376 D.

377

378

379 **Figure 2. GATA4 controls the colonization of *segmented filamentous bacteria* to shape tissue**

380 **immunity.**

381 (**A**) SFB load (y axis), as measured by qPCR, relative to the amount of host DNA in mucosal

382 scrapings (dots) of jejunum (dark colors) and ileum (light colors) from WT (blues) and GATA4$^{\Delta IEC}$

383 (reds) mice. **** *P*<0.0001 , *** *P*p<0.001, ** *P*<0.01,Kruskal-Wallis with Dunn multiple

384 comparison test. *N*= 18-19 mice/group.

385 (**B**) Representative FISH staining of jejunal (left) and ileal (right) tissue of WT (top) and

386 GATA4$^{\Delta IEC}$ (bottom) mice using universal 16s rRNA probes (Alexa 546) and a probe specific for

387 SFB 16s (Alexa 488) and counterstained with DAPI. The overlay of the 16s probes represents

388 SFB.

389 (**C** and **D**) Frequency (y axis) of IFNγ^{+} cells among CD8αβ T cells (**C**) or IL-17a^{+} cells among CD4 T

390 cells (**D**) in jejunum samples (dots) from SPF, GF, Jackson (JAX) microbiota transfer, and Jackson

391 microbiota + SFB transfer (x axis) into WT and GATA4$^{\Delta IEC}$ mice (color, as in A). **** *P*<0.0001, **

392 *P*<0.01, t-test. *N*= 4-6 mice/group.

393 (**E**) Frequency (y axis) of RORγt^{+} FOXP3^{-} cells among transferred CD45.1^{+} CD4^{+} Vβ14^{+} 7B8^{+} T cells

394 in the MLN (dots) draining the jejunum and ileum of CD45.2 WT and GATA4$^{\Delta IEC}$ mice (color, as

395 in A), three days after transfer. **** *P*<0.0001 , *** *P*<0.001, ANOVA with Tukey multiple

396 comparison test. *N*= 5 mice/group.

397 (**F**) Number (y axis) of transferred T cells (as in E) in the jejunum LP(dots) 9 days after transfer

398 into WT and GATA4$^{\Delta IEC}$ mice (color, as in A). ** *P*<0.01, t-test. *N*= 4 mice/group.

399 (**G**) Representative plot showing frequency of cells with vβ14+ TCR among transferred CD4+

400 CD45.1+ 7B8+ T cells in in the jejunum of WT (top) or GATA4$^{\Delta IEC}$ (bottom) mice.

401 (**H**) Rat SFB loads (y axis), as measured by qPCR relative to host DNA, in the jejunum (dots) of
402 monocolonized WT and GATA4$^{\Delta IEC}$ mice (color, as in A).
403 (**I** and **J**) Frequencies (y axis) of IFNγ^{+} cells among CD8αβ T cells (**I**) and of IL-17^{+} cells among CD4
404 T cells (**J**) in jejunum (dots) of GATA4$^{\Delta IEC}$ mice monocolonized with rat or mouse SFB (x axis).
405 **** *P*<0.0001, t-test. *N*= 5 mice/group.
406
407 **Figure 2 supplemental. GATA4 controls the colonization of regional mucosal associated**
408 **bacteria to shape tissue immunity.**
409 (**A**) Relative frequencies (x axis, $\log_2$ fold change) in GATA4$^{\Delta IEC}$ versus WT mice of different types
410 of microbes (dots), and their statistical significance (y axis, $-\log_{10}$ of the FDR-adjusted *P* value),
411 based on 16S rRNA sequencing of the luminal contents and mucosal scrapings of the jejunum
412 (left set) and ileum (right set) of WT and GATA4$^{\Delta IEC}$ mice. SFB are classified as *Candidatus*
413 *arthromitus*.
414 (**B** and **C**) Frequencies (y axis) of IFNγ^{+} cells among CD8αβ T cells (**B**) and of IL-17a^{+} cells among
415 CD4 T cells (**C**) in the jejunum (dots) of WT (blue) and GATA4$^{\Delta IEC}$ (red) mice with different
416 microbiota (x axis), *i.e.,* SPF, GF, or ASF, WT microbiota (no SFB), WT microbiota (with SFB),
417 GATA4$^{\Delta IEC}$ microbiota, or SFB. **** *P*<0.0001, *** *P*<0.001, ** *P*<0.01, t-test. *N*= 3-6
418 mice/group.
419 (**D**) *Mucispirillum schaedleri* load (y axis), as measured by qPCR, relative to host DNA in the
420 jejunum after ASF transfer to WT (blue) and GATA4$^{\Delta IEC}$ (red) mice. * *P*<0.05, Mann-Whitney.
421 *N*=4-5 mice/group.
422

423 **Figure 3. GATA4 regulates regionalization of retinol metabolism and IgA to limit SFB**
424 **colonization in the proximal intestine.**

425 (**A**) FISH staining of SFB from jejunal (left) and ileal (right) tissue of monocolonized IgA deficient
426 (IgA-/-) (bottom) and littermate control (IgA+/-) (top) mice. SFB is stained red with Cy5-
427 conjugated SFB-specific 16S probes. The DAPI counterstain is blue.

428 (**B**) SFB load (y axis), as measured by qPCR, relative to host DNA in mucosal scrapings (dots)
429 from the jejunum (violet) and ileum (light pink) of control (IgA+/-) mice and the jejunum
430 (fuchsia) of IgA deficient (IgA-/-) mice (x axis) mice. ** $P<0.01$, * $P<0.05$, Kruskal-Wallis with
431 Dunn multiple comparison test. N= 7-8 mice/group.

432 (**C**) Number of $CD45^+$ lin^- IgA^+ $B220^-$ plasma cells (y axis), as determined by flow cytometry, in
433 tissue (dots) taken from the jejunum (dark colors) and ileum (light colors) of WT (blues) and
434 $GATA4^{\Delta IEC}$ (reds) mice. *** $P<0.001$, ** $P<0.01$, ANOVA with Tukey multiple comparison test.
435 N= 5 mice/group.

436 (**D**) Concentration of IgA (y axis), as determined by enzyme-linked immunoassay (ELISA), in the
437 contents (dots) of the jejunum (left) or ileum (right) of WT and $GATA4^{\Delta IEC}$ mice (color, as in C).
438 **** $P<0.0001$, t-test.

439 (**E**) Frequency of IgA coated bacteria (gated SYTOBC+) after staining of RAG-/- feces with
440 supernatant from WT and $GATA4^{\Delta IEC}$ jejunal contents, followed with labelled anti-IgA APC
441 antibody . *** $P<0.001$, t-test. N= 4-5 mice/group.

442 (**F**) Heatmap of z-scored expression (color) of epithelial, region-specific, GATA4-regulated genes
443 (rows) overlapping the KEGG retinol metabolism pathway, determined from RNA-seq of

444 epithelial cell samples (columns) from the jejunum and ileum of WT (left) and GATA4$^{\Delta IEC}$ (right)
445 mice.
446 (**G**) Top, representative histogram of ALDH expression in jejunal epithelial cells of WT (blue) and
447 GATA4$^{\Delta IEC}$ (red) mice revealed by ALDEFLUOR staining. WT epithelial cells treated with ALDH
448 inhibitor are shown as negative control for background fluorescence. Bottom, summary plots
449 show the normalized geometric mean fluorescence intensity (gMFI, y axis) of ALDEFLUOR
450 staining in samples (dots) of epithelial cells from the jejunum and ileum of WT and GATA4$^{\Delta IEC}$
451 mice (color, as in C). **** $P<0.0001$, ANOVA with Tukey multiple comparison test. N= 6
452 mice/group.
453 (**H**) SFB loads (y axis), as measured by qPCR relative to host DNA, in jejunal mucosal scrapings of
454 GF WT mice fed a control or vitamin A deficient diet for 4 weeks and subsequently colonized
455 with SFB. * $P<0.05$ Mann Whitney test N= 5 mice/group.
456 (**I**) Total IgA (y axis) in the jejunal contents of WT, GATA4$^{\Delta IEC}$ vehicle-treated, and GATA4$^{\Delta IEC}$ RA-
457 treated mice (x axis, color) after 14 days. *** $P<0.001$, ** $P<0.01$, * $P<0.05$, ANOVA with Tukey
458 multiple comparison test. N= 4 mice/group.
459 (**J**) SFB loads (y axis), as measured by qPCR relative to host DNA, in jejunal mucosal scrapings in
460 PBS-treated WT or GATA4$^{\Delta IEC}$ mice, and IgA-supplemented GATA4$^{\Delta IEC}$ mice (x axis, color). ***
461 $P<0.001$, ANOVA with Tukey multiple comparison test. N= 5-7 mice/group.
462
463 **Figure 3 Supplemental. GATA4 regulates regionalization of retinol metabolism and B cell**
464 **responses**

465 (**A**) FISH staining of SFB in jejunal (left) and ileal (right) tissue of monocolonized B-cell deficient
466 (Jh-/-) and littermate control (Jh+/-) mice. Red indicates SFB stained with Cy5-conjugated SFB-
467 specific 16s probes. Blue is DAPI counterstain.
468 (**B**) SFB load (y axis), as measured by qPCR relative to host DNA, in mucosal scrapings (dots)
469 from the jejunum (light blue) of control (Jh+/-) mice and the jejunum (violet) and ileum (dark
470 purple) of B-cell deficient (Jh-/-) mice. *** $P<0.001$, ** $P<0.01$, Kruskal-Wallis with Dunn
471 multiple comparison test. N= 7-8 mice/group.
472 (**C**) Concentration of IgA (y axis)), as measured by ELISA, in culture supernatant (dots) of tissue
473 explants from the jejunum (dark colors) and ileum (light colors) of WT (blues) and GATA4$^{\Delta IEC}$
474 (reds) mice after 24 hours in culture. **** $P<0.0001$, ANOVA with Tukey multiple comparison
475 test. N= 6 mice/group.
476 (**D**)Concentration of IgA (y axis), as measured by ELISA, in the contents (dots) of the jejunum
477 (left) or ileum (right) of GF WT (blues) and GATA4$^{\Delta IEC}$ (reds) mice. ** $p<0.01$, t-test.
478 (**E**)Intensity (y axis), of IgA bacterial coating as described in Fig. 3E. MFI, mean fluorescence
479 intensity. * $P<0.05$, t-test. N= 4-5 mice/group.
480 (**F**) Pathways (y axis) in the KEGG database that are significantly enriched (enrichment score, x
481 axis) in epithelial, region-specific, GATA4-regulated genes.
482 (**G**) Schema (left) for isolation of sIgA from luminal contents of WT mice, and FACS plots (right)
483 showing the frequency of IgA-coated bacteria after staining feces of RAG-/- mice with isolated
484 sIgA.
485 (**H**) Experimental schema of IgA gavage and SFB colonization experiment in GF WT and
486 GATA4$^{\Delta IEC}$ mice.

487 (**I**)SFB loads (y axis), as measured by qPCR, in ileal mucosal scrapings (dots) in PBS-treated WT
488 or GATA4$^{\Delta IEC}$ mice, and in IgA-supplemented GATA4$^{\Delta IEC}$ mice (x axis, color). * *P*<0.05, ANOVA
489 with Tukey multiple comparison test. *N*= 5-7 mice/group.
490
491 **Figure 4. Loss of intestinal regionalization impacts host susceptibility to enteropathogenic**
492 **infection by *Citrobacter rodentium*.**
493 (**A**) *C. rodentium* load (y axis), measured by qPCR relative to host DNA, in distinct intestinal
494 segments (x axis) in WT (blue) and GATA4$^{\Delta IEC}$ (red) mice. **** *P*<0.0001, ** *P*<0.01, Mann-
495 Whitney test. *N*= 13 mice/group.
496 (**B**) Bacterial loads (y axis) of wild-type *C. rodentium* (purple, orange) and the ΔEAE mutant
497 (black), measured by qPCR relative to host DNA, in distinct intestinal segments (x axis) of GF WT
498 (purple) or GATA4$^{\Delta IEC}$ (orange, black) mice. * *P*<0.05, Kruskal-Wallis with Dunn multiple
499 comparison test, GF WT + WT *C. rodentium* vs. GF GATA4$^{\Delta IEC}$ + WT *C. rodentium*. *N*= 7-9
500 mice/group.
501 (**C**) Representative FACS plots of IFN-γ (x axis) vs. IL-17a (y axis) in PMA ionomycin stimulated
502 CD4+ T cells that were isolated from the LP of infected or uninfected mice (left) or mice 10 days
503 after infection by *C. rodentium* (right).
504 (**D**) Representative immunohistochemistry (IHC) staining of each intestinal region (panel title)
505 from WT (top) and GATA4$^{\Delta IEC}$ (bottom) mice 10 days after *C. rodentium* infection.
506 (**E**) CFUs of *C. rodentium* translocation (y axis) to MLN, liver, and spleen (x axis) 10 days after
507 infection of WT (blue) and GATA4$^{\Delta IEC}$ (red) mice . **** *P*<0.0001, *** *P*<0.001, * *P*<0.05, Mann-
508 Whitney test. *N*= 8-10 mice/group.

509 (**F**) Percent survival (y axis) of WT (blue) and GATA4$^{\Delta IEC}$ (red) mice 0–15 days (x axis) post *C.*

510 *rodentium* infection. ** *P*<0.01 Mantel-Cox test. *N*= 8-9 mice/ group.

511

512 **Figure 5. Loss of GATA4 expression is associated with lipid metabolic dysfunction and**

513 **increased IL-17 signaling in celiac disease.**

514 (**A**) Normalized and batch corrected levels of GATA4 expression in duodenal biopsies from

515 healthy controls (blue), active celiac disease patients (ACeD, in red), and gluten free diet celiac

516 patients (GFD, in yellow). *** *P*<0.001, Wilcoxon test.

517 (**B**) Representative IHC staining for GATA4 in healthy control, active celiac disease patients, and

518 gluten free diet celiac patients.

519 (**C**) Bar plot shows the percentages (y-axis) of human-mouse homologous genes specific to

520 GATA4-hi or GATA4-lo (x-axis) individuals, that are also either jejunum (purple) or ileum-like

521 (yellow) specific and GATA4-regulated, or neither (gray). *** $P < 10^{-51}$, ** *P* <10-6, NS not

522 significant, Fisher's exact test.

523 (**D**) Bar plots show the pathways (y axis) and significance ($-\log_{10}$ of adjusted *P*, x axis) of the

524 most significantly enriched gene ontologies (biological processes) in the 250 genes that are

525 upregulated in both GATA-hi individuals and WT mouse jejunum.

526 (**E**) Single-sample gene set enrichment analysis (ssGSEA) scores (y axis) for the retinol

527 metabolism (left) and IL-17 downstream signaling (right) pathways in GATA4-hi and GATA4-lo (x

528 axis) individuals (dots) from control (blue), ACeD (red), and GFD (yellow) patient groups. **** *P*

529 <0.0001 Wilcoxon rank test.

530 (**F**) Left, heatmap displays the scaled effect size (color) of the absence or presence of relevant 5
531 bacteria (see Fig. S4I) (rows) on GATA4 expression and on the ssGSEA scores of metabolic and
532 immune pathways (columns), while accounting for sex, age and technical covariates.
533 Significance annotated in red, · *P* <0.1, * *P* <0.05, ** *P* <0.01 t-test. Right, bar plot showing the
534 numbers (x axis) of ACeD samples in which the bacteria (y axis) were detected (orange) or not
535 (gray).
536 (**G**) Box plots show GATA4 expression (log_2 CPM, y axis, left panel) and ssGSEA scores (y axis,
537 other panels) for the retinol metabolism (center) and IL17 downstream signaling (right)
538 pathways in ACeD patients (dots), grouped (x axis) by the absence (blue) or presence (orange)
539 of bacteria from the Actinobacillus genus. * *P* <0.05, Wilcoxon test.
540 **Figure 4 supplemental. Loss of GATA4 is associated with lipid metabolic defects, mucosal**
541 **associated bacteria, and increased IL-17 signaling in celiac disease patients.**
542 **(**A) *Gata4* expression (y axis) in GATA4-hi and GATA4-lo groups (x axis) in individuals (dots) from
543 control (blue), ACeD (red), and GFD (orange) patients.
544 (**B**) Euler diagram represents the overlaps of 4 sets in a common universe of 11,657 mouse-
545 human transcriptional homologs: GATA4-regulated jejunum-specific (purple) and ileum-like–
546 specific (yellow) genes, and GATA4-hi specific (right, gray) and GATA4-lo ACeD specific genes
547 (left, gray). Enrichment significance as annotated, Fisher's exact test.
548 (**C**) Retinol pathway ssGSEA scores (y axis) in individuals (dots) from control, ACeD, or GFD
549 patient groups (x axis, color, as in A). **** *P* < 0.0001, NS not significant, Wilcoxon rank test.

550 (**D**) Scatter plot of GATA4 normalized expression (y axis) and ssGSEA scores (x axis) for the
551 retinol metabolism pathway in individual samples (dots) from control (blue) and ACeD (red)
552 patient groups. Annotated with Pearson correlation coefficient and p-value.
553 (**E**) IL-17 downstream signaling pathway ssGSEA scores, analogous to C.
554 (**F**) Scatter plot of GATA4 normalized expression (y axis) and ssGSEA scores (x axis) for the IL-17
555 downstream signaling pathway, analogous to D.
556 (**G** and **H**) Box plots of absolute abundances (y axis) of mucosal 16S (**G**) and Neisseria **(H**) in
557 biopsies (dots) from control (blue) or ACeD (red) (x axis) patients.
558 (**I**) Principal component analysis (PCA) biplot shows PC2 (x axis) and PC3 (y axis) of each
559 individual (dot) from control (blue) or ACeD (red) patient groups; arrows indicate the loadings
560 of the 5 bacteria most associated with PC2 and PC3.
561 (**J**) Numbers (x axis) of control or ACeD patients (y axis) with detectable (orange) or
562 undetectable (gray) levels of the indicated bacteria.
563 (**K)** Heatmap shows the scaled effect size (color) of each of the bacteria from J (x axis) on GATA4
564 expression and ssGSEA scores of the indicated pathways in controls, compared to ACeD
565 patients. · P <0.1, * P <0.05, t-test.
566
567
568
569
570
571

572

573

574

575

576

577

578

579

580

581

582

583 **Materials and Methods**

584 **Mice**

585 7-12 week old mice were used for experiments, co-housed in specific pathogen-free
586 conditions, and kept *Helicobacter hepaticus*, murine norovirus free at the University of Chicago.
587 Some mice were also housed in gnotobiotic isolators and routinely checked for sterility by
588 culture and 16S PCR or kept SFB monocolonized at the University of Chicago Gnotobiotic
589 Research Animal Facility. GATA4fl/fl villin-cre SPF mice were previously generated in the CD1
590 background and obtained from the Matzinger laboratory (*19*).This line was rederived GF for this
591 study and backcrossed for 10 generations to C57BL/6J background for T cell transfers. C57BL/6J,
592 B6-Tg(Tcra, Tcrb)2Litt/J SFB TCRtg, B6.SJL-*Ptprc*^{a} *Pepc*^{b}/BoyJ, B.6129S7-Rag1^{tm1mom}/J were
593 obtained from the Jackson Laboratory. CD-1 IGS mice were obtained from Charles River

594 Laboratories. B cell deficient mice deficient for IgH J segment locus (Jh) were obtained from Dr.
595 Bendelac at the University of Chicago and generated on a C57BL/6 background using Cas9 with
596 the protospacers GCTACTGGTACTTCGATGTC and GCCATTCTTACCTGAGGAGA. IgA deficient mice
597 where the Sα (IgA switch region) and C1α (first exon) were deleted were obtained from Dr.
598 Bendelac and generated on a C57BL/6 background using Cas9 with the protospacers
599 AAGCGGCCACAACGTGGAGG and TCAAGTGACCCAGTGATAAT. Jh and IgA deficient mice were
600 rederived GF at Taconic Biosciences. Littermate controls of GATA4, Jh, and IgA were used for all
601 experiments in this study. Mice were fed a standard chow diet, vitamin A control diet (Harlan
602 TD.91280), or vitamin A deficient diet (Harlan TD. 86143). Animal husbandry and experimental
603 procedures were performed in accordance with Public Health Service policy and approved by
604 the University of Chicago Institutional Animal Care and Use Committees.

605 **Patients**

606 A duodenal biopsy was obtained from 166 individuals undergoing upper gastrointestinal
607 endoscopy at the University of Chicago and at Mayo Clinic as previously reported (*38*). There
608 were 64 control patients, 56 untreated patients with active celiac disease, and 46 patients
609 treated with a gluten free diet. All control patients underwent endoscopies for issues unrelated
610 to celiac disease and had normal intestinal histology, no family history of celiac disease, and no
611 significant levels of anti-TG2 antibodies in the serum. Patients with active celiac disease
612 contained positive anti-TG2 antibodies and small intestinal enteropathies with increased IEL
613 infiltration, crypt hyperplasia, and villous atrophy according to the accepted diagnostic
614 guidelines (*39*). The subjects signed an informed consent as provided by the Institutional
615 Review Board of each institution (IRB-12623B for the University of Chicago, and IRB-1491-03 for

616 the Mayo Clinic). DNA and RNA were isolated from each biopsy as described previously (*38*)
617 using the AllPrep DNA/RNA mini kit (Qiagen).

618 **Isolation of intestinal epithelial cells (IEC), intraepithelial lymphocytes (IEL), and lamina**
619 **propria (LP) cells**

620 The segments of the intestine were excised as follows to isolate cells for flow cytometry:
621 duodenum was taken 12 cm from the stomach, jejunum 12 cm from the middle, and ileum 12
622 cm from the cecum. Any leftover segments were discarded. The entire colon was taken after
623 the cecum to the rectum. To isolate IEL and IECs, Peyer's patches were first removed from the
624 small intestine, and then the segments were opened longitudinally and washed briefly in
625 phosphate-buffered saline (PBS). Epithelial cells, including IELs and LP cells, were isolated as
626 previously described (*40*) using ethylenediaminetetraacetic acid (EDTA) containing calcium-free
627 media and collagenase VIII (Sigma-Aldrich, C2139), respectively. The IEL and LP compartments
628 were then subjected to a 40% percoll density gradient centrifugation step to remove dead cells
629 and debris as previously described (*38*). The IEL and LP cells were then counted on a
630 hemocytometer.

631 **Cytokine stimulation**

632 Up to $2x10^6$ cells were collected and resuspended in RPMI 1640 media with 10% fetal
633 bovine serum (FBS) and cultured in 48-well plates in the presence of 750 ng/ml of ionomycin,
634 50 ng/ml of Phorbol 12-myristate 13-acetate (Sigma-Aldrich), and golgi-stop (BD). The cells
635 were incubated for 2 hours at 37C with 5% CO_2. After stimulation, the reaction was quenched
636 with ice cold FACS buffer, and the cells were subsequently stained with antibodies for flow
637 cytometry.

638 **Flow cytometry**

639 The cells were first stained with FC block (CD16/32) to block nonspecific binding and
640 then were stained with dead dye to exclude dead cells (Aqua, ThermoFisher or Zombie NIR,
641 Biolegend) for 15 min at 4 °C, followed by staining with cell surface markers for 20 min at 4 °C.
642 For intracellular cytokine staining, the BD cytofix/cytoperm kit was used, and cells were
643 incubated with the antibodies for 40 min at 4 °C. For intracellular transcription factors, the
644 Foxp3 eBioscience kit was used according to the manufacturer's instructions. The antibodies
645 used are indicated in Table (1). For ALDH staining of IECs, the ALDEFLUOR kit was used
646 (StemCell Technologies), following the manufacturer's protocol. All cells were gated FSC, SSC,
647 singlets, and live cells. IECs were gated CD45- EpCAM+. 100,000 IECs from the jejunum or ileum
648 were sorted with Aria Fusion (BD Biosciences) into RLT buffer (Qiagen) with β-mercaptoethanol
649 for downstream sequencing analysis. IgA plasma cells were gated as described previously (*41*),
650 i.e. EpCAM-, CD45+/dim, lineage negative (Ter119, F4/80, CD3, Ly6G, NK1.1, CD19), IgA+, B220-
651 . CD8αβ IELs were gated TCRβ+, CD4-, CD8α+, CD8β+. CD4 LP T cells were gated TCRβ+, CD4+,
652 CD8α-. The following antibodies and clones were purchased from Biolegend: CD45 Pacific Blue
653 (30-F11), CD4 BV785 (GK1.5), CD4 BV605 (GK1.5), IL10 PE-Cy7 (JES5-16E3), CD45.1 Pacific Blue
654 (A20), Tbet PE (4B10), CD44 PE-Cy7 (IM7), CD62L PE (MEL-14), Epcam PerCP-Cy5.5 (G8.8), CD19
655 FITC (1D3/CD19), NK1.1 BV605 (PK136), CD11C BV605 (N418), TER119 BV605 (TER-119), F4/80
656 BV605 (BM8), CD3ε BV605 (145-2C11), Ly6G BV605 (1A8), B220 PE-Cy7 (RA3-6B2). The
657 following antibodies and clones were purchased from BD: CD8β BUV395 (H35-17.2), CD8α
658 PerCP-Cy5.5 (53-6.7), NK1.1 PE-CF594 (PK136), TCRβ BUV737 (H57-597), TCRβ BV711 (H57-
659 597), CD3ε BUV737 (145-2C11), IFN-g APC (XMG1.2), CD45.2 BUV395 (104), vβ14 TCR FITC (14-

660 2), RORγt BV786 (Q31-37). The following antibodies and clones were purchased from Thermo
661 Fisher: TCRgd FITC (eBioGL3), IL17a PE (ebio17B7), FOXP3 eFluor450 (FJK-16s), FOXP3 FITC (FJK-
662 16s), FOXP3 PE-Cy7 (FJK-16s), IgA PE (mA-6E1). The cells were run on the LSRFortessa X-20 Flow
663 Cytometer (BD Biosciences) or the Cytek Aurora and data were analyzed using FlowJo software
664 (Treestar).

665 **DNA isolation**

666 For mucosal scrapings for DNA isolation, 5 cm of tissue proximal to the middle of the
667 intestine was taken for the jejunum, and 5 cm from the ileocecal valve was taken for the ileum.
668 The entire colon was used for mucosal scrapings. The tissue was excised, opened longitudinally,
669 scraped with a glass slide, transferred to 2 ml screw cap tube containing 0.1 mm glass beads
670 (Bio-spec), and snap frozen on dry ice. For luminal content, 50-100 mg of content was taken
671 from as close to the middle of the jejunum as possible and from the last 5–7 cm of the ileum.
672 Homogenization was performed after adding 1 ml of inhibitex buffer (Qiagen) using the Bead
673 Ruptor Elite bead mill homogenizer (Omni, 19040E) on speed 6 for 3 min. DNA was then
674 extracted using the QIAmp Fast DNA stool mini kit (Qiagen) following the manufacturer's
675 protocol with the optional high temp (95 °C) lysis step. DNA concentration was determined
676 using the nanodrop UV spectrophotometer (ThermoFisher).

677 **RNA isolation**

678 For eventual RNA purification, 1 cm of tissue was excised from the beginning of the
679 duodenum, the middle of the jejunum, the end of the ileum, and the center of the colon and
680 preserved in RNAprotect (Qiagen) overnight at 4 °C and then transferred to –80 °C for long-
681 term storage. The tissue was transferred to 600 µl of RLT buffer containing β-mercaptoethanol

682 (Qiagen) and homogenized for 30 sec with a hand held rotor (Tissue-Tearor, BioSpec). RNA was
683 purified using RNAeasy plus mini kit (Qiagen) following the manufacturer's protocol with the
684 optional on column DNase digest (Qiagen).

685 **qPCR**

686 RNA was first reverse-transcribed to cDNA using GoScript Reverse Transcriptase kit
687 (Promega) following the manufacturers protocol. For qPCR, 10 ng of cDNA or 20 ng of DNA
688 from mucosal scrapings and content was used. TB green Advantage qPCR Premix (Takara) was
689 used, and the target gene was quantified and normalized to the housekeeping gene as
690 described previously (*38*) using $1000*2^{-(Ct\ target-\ Ct\ housekeeping)}$ formula. For host gene expression,
691 the target gene was normalized to GAPDH. For bacterial load, the target gene was normalized
692 to either host DNA as described previously (*42*) with primers specific for host argininosuccinate
693 lyase (ASL) gene or universal 16S primers. The qPCR was performed on the LightCycler 480
694 System (Roche). The primer pairs and DNA sequences are included in Table (1).

695 **Histology**

696 The tissue was collected in the same manner as for RNA, placed in cassettes and fixed in
697 10% formalin for H&E staining or Carnoy solution (ThermoFisher) for fluorescent *in situ*
698 hybridization (FISH) staining overnight at room temperature. Cassettes were transferred to 70%
699 ethanol for formalin or 100% ethanol after Carnoy fixation to wash out the fixative. The tissue
700 was embedded in paraffin, and slides were cut at 5 µm thickness. The H&E staining was
701 performed by the Human Tissue Resource Center at the University of Chicago. For FISH staining,
702 the paraffin was first removed by running the slides through four 3-min incubations in xylene
703 and four 3-min incubations in 100% ethanol. The slides were then moved to a polypropylene

704 slide container and filled with hybridization solution containing the diluted 16S probe (0.9M
705 NaCl, 20mM Tris-HCL pH 7.5, 0.1% SDS with 0.2ng of probe specific for SFB 16S or universal 16S)
706 (*43*). The 16S probes used are included in Table (1). The slides were incubated overnight at 50
707 °C in the dark. The slides were washed three times with the hybridization buffer, briefly rinsed
708 in H_2O, and then mounted with Prolong diamond antifade with DAPI (ThermoFisher). The slides
709 were scanned with the CRi Pannoramic SCAN 40x Whole Slide Scanner at the University of
710 Chicago Integrated Light Microscopy core.

711 **ELISA**

712 To quantify luminal IgA levels, content was collected from the jejunum and ileum and
713 weighed in 2-ml bead beating tubes containing 0.1mm glass beads. After adding 1 ml of 1X cell
714 lysis buffer with protease inhibitors (Cell Signaling Technologies), the content was homogenized
715 on a vortex for 5 min. The debris were pelleted at 13000 rpm for 10 min, and the supernatant
716 was collected for ELISA. For tissue explants, 1 cm of tissue was excised and opened
717 longitudinally, washed in PBS, and placed in complete RPMI at 37 °C for 24 h. The culture
718 supernatant was collected and used for ELISA. The supernatant was diluted in 1X assay diluent
719 A (ThermoFisher), and the dilution in the middle of the standard curve was used to quantify IgA
720 levels. IgA mouse uncoated ELISA kit (ThermoFisher) was used following the manufacturer's
721 protocol, and absorbance was read at 450 nm. The amounts of IgA were back calculated to the
722 original sample and normalized relative to the weight of the content or to ml of culture
723 supernatant.

724 **Luminal IgA isolation and *in vivo* treatment**

725 To isolate luminal polyclonal sIgA from the intestine, luminal content was pooled from
726 the small intestine, large intestine, and cecum from SFB+ 8–12 week old WT CD1 mice (Charles
727 river). Content was transferred to falcons containing 1X Tris-Buffered Saline, 0.1% Tween 20
728 (TBST) buffer with proteinase inhibitor (Roche). Falcons were then vortexed for 5 min on max
729 speed and centrifuged for 10 min at 5000 rpm. The supernatant was collected and spun again
730 two times to further remove bacteria and debris. Pierce Protein L Magnetic Beads (Thermo
731 Scientific) were added and incubated for 1 h at room temperature while shaking. After 1 h,
732 beads selectively bound to IgA through kappa light chain were separated from the supernatant
733 with EasySep magnetic stand (StemCell technologies). Supernatant was discarded and beads
734 washed 3 times. IgA was separated from the beads with Pierce IgG Elution Buffer pH 2.0
735 (ThermoFisher). Elution buffer was incubated with the beads for 10 min at room temperature
736 on a shaker. Tris-HCl 1M pH 8.5 was added to neutralize the solution. IgA protein concentration
737 was measured using NanoDrop. The isolated IgA preparation was then filtered with 0.22 µm
738 sterile syringe filter unit and protease inhibitors were added (ThermoFisher). The IgA
739 preparation was kept up to one week at 4 °C. When IgA were administered to the mice by
740 gavage, the isolated IgA preparation was further concentrated with Amicon Ultra-4 Centrifugal
741 Filter Units (MilliporeSigma) until 250–350µg/0.1ml final concentration was achieved. To
742 optimize the treatment protocol to restore luminal IgA levels *in vivo* to GATA4$^{\Delta IEC}$ mice, we first
743 treated RAG-/- mice with 250 µg of the IgA preparation. After 1 h, we assessed the frequency of
744 IgA+ bacteria in small intestinal contents by flow cytometry as described below, and noted 10–
745 20% of bacteria were IgA+ after gavage. The bacterial coating was transient due to intestinal
746 flow and undetectable after two hours. Therefore, continuous IgA gavages were necessary to

747 sustain luminal IgA and bacterial coating in the small intestine. To administer the IgA
748 preparation and analyze SFB colonization, GF, WT and GATA4$^{\Delta IEC}$ mice were gavaged with 100 µl
749 of IgA or PBS. After 1 h, the mice were colonized with SFB and gavaged again with IgA or PBS.
750 Three more gavages were performed at 2-h intervals. The mice were euthanized 24 h after the
751 gavage of SFB, and the regionalization of SFB load was assessed in jejunal and ileal mucosal
752 scrapings by qPCR.

753 **Bacterial staining with luminal IgA**

754 Luminal content was taken from WT, GATA4$^{\Delta IEC}$, or RAG-/- mice and resuspended in 1X
755 PBS with protease inhibitors at a concentration of 0.1mg/µl, vortexed for 5 min, and spun at
756 8000 rpm for 5 min. Three fecal pellets from RAG-/- mice were homogenized and pelleted. The
757 bacterial pellet was resuspended in 50µl PBS and combined with 50µl of luminal supernatant
758 containing IgA. The IgA was incubated with the bacteria for 1 h at 4 °C. The bacteria were then
759 washed, pelleted, and stained with SYTO BC (ThermoFisher) diluted 1:5000 and anti-IgA APC
760 (Southern Biotech) diluted 1:200 for 30 min. Bacteria were gated on FSC, SSC, SYTOBC+, and
761 IgA+.

762 ***Citrobacter rodentium* infections**

763 *C. rodentium* strains DBS100, DBS120 *pler-lux*, or DBS100 ΔEAE were grown at 37 °C in
764 Luria broth under agitation (*12, 44*). The cultures were diluted 100X and grew to log phase until
765 the OD^{600nm} reached 0.75. For gavage, 200 µl of bacteria were used, which gave a dose of
766 2.5X10^{9} CFU/mouse. Mice were separated into cages based on genotype for infections, and
767 male mice were used for survival studies. DBS100 or DBS100 ΔEAE strains were given to GF
768 mice and DBS120 *pler-lux* was given to SPF mice. The DBS120 strain has a genomic kanamycin

769 resistance cassette inserted through Tn5. To determine CFUs of DBS120, 2 fecal pellets/mouse
770 were resuspended in 1 ml of PBS and plated on MacConkey agar containing 50 µg/ml of
771 kanamycin. The CFU/mg feces concentration was determined as: (#CFU counted*Dilution
772 factor/(vol plated in ml))/mg feces. To determine the amount of bacterial translocation, the
773 MLN, liver, and spleen were aseptically dissected, weighed, and homogenized with the Tissue-
774 Tearor rotor (BioSpec) in 500 µl of PBS. Then 200 µl of homogenate was plated on MacConkey
775 agar containing 50µg/ml of kanamycin.

776 **Microbial transfers**

777 To colonize mice with SFB (*44*) or rat SFB (*12*), 3–4 fresh fecal pellets from SFB
778 monocolonized mice were homogenized in 1 ml of PBS, vortexed for 3 min, and spun at 300 g
779 to remove large debris. Then 200 µl of the homogenate were gavaged to recipient mice. When
780 possible, SFB donor pellets were taken from monocolonized Jh or IgA deficient mice, which
781 harbor 10-fold higher levels of SFB. To colonize mice with SFB-free microbiota, C57BL6 mice
782 from Jackson Labs, which lack SFB in the microbiota, were used as donor mice. Small intestinal
783 and cecal contents were pooled for one donor mouse homogenized in PBS, and gavaged to
784 recipient mice with or without SFB supplemented. For WT and GATA4$^{\Delta IEC}$ microbiota transfer to
785 GF WT and GATA4$^{\Delta IEC}$ hosts, jejunal content was pooled from two donor SPF GATA4$^{\Delta IEC}$ mice or
786 littermate WT mice. Colonization of ASF strains (Taconic) was performed as described
787 previously (*45*) and gavaged to recipient WT and GATA4$^{\Delta IEC}$ mice. For all microbial transfers,
788 mice were colonized at 4 weeks of age and analyzed at 8 weeks.

789 **Vitamin A deficient diet**

790 GF C57BL6 mice were placed on control (Harlan TD.91280) or vitamin A deficient (Harlan
791 TD. 86143) diets from 4 to 8 weeks of age. At 8 weeks, mice were monocolonized with SFB for
792 one week as described above, and the amount of SFB in jejunal mucosal scrapings was
793 quantified by qPCR.

794 **SFB TCRtg adoptive transfer**

795 Naïve SFB TCRtg Vβ8 CD4 T cells were isolated from LNs and spleen of congenically
796 marked CD45.1 Vβ8+/- female mice using the naïve CD4 T cell isolation kit (Miltenyi), and $2x10^5$
797 cells/100µl mouse were injected retroorbitally into CD45.2 WT and $GATA4^{\Delta IEC}$. Three days after
798 transfer, the mice were euthanized to assess T cell priming and activation in the jejunal and ileal
799 draining MLN as described previously (Esterházy et al., 2019). To assess T cell expansion in the
800 LP of the jejunum nine days after transfer, 50,000 cells were injected/mouse.

801 **16S sequencing libraries and analysis**

802 The following 16S methods that were used were adapted from a paper currently in
803 press (*46*). Extracted DNA was amplified, barcoded, and sequenced as described previously (*36,*
804 *47, 48*). Briefly, amplification of the variable 4 (V4) region of the 16S rRNA gene was performed
805 in 20 uL duplicate reactions: 8 uL of 2.5X 5Prime Hotstart Mastermix (VWR, Radnor, PA, USA), 1
806 uL of 20X Evagreen (VWR), 2 uL each of 5 uM forward and reverse primers (519F, barcoded
807 806R, IDT, CoralVille, IA, USA), variable input volumes of extracted DNA template and nuclease
808 free water. Total DNA input (determined by NanoDrop) was limited to 400 ng to prevent
809 inefficient amplification. Amplification reactions were monitored on a CFX96 RT-PCR machine
810 (Bio-Rad Laboratories, Hercules, CA, USA), and samples were removed in the late exponential
811 phase to minimize chimera formation and non-specific amplification (1,4,5). Amplification

812 cycling conditions were as follows: 94 °C for 3 min, up to 50 cycles of 94 °C for 45 s, 54 °C for 60
813 s, and 72 °C for 90 s. Successfully amplified duplicates were pooled together and quantified
814 with KAPA library quantification kit (Roche, Basel, Switzerland), and then all samples were
815 combined at equimolar concentrations with up to 96 samples per library. AMPureXP beads
816 (Beckman Coulter, Brea, CA, USA) were used to clean up and concentrate libraries before final
817 library quantification with a High Sensitivity D1000 Tapestation Chip (Agilent, Santa Clara, CA,
818 USA). Illumina MiSeq sequencing was performed with a 2x300bp reagent kit by Fulgent
819 Genetics (Temple City, CA, USA). Raw reads were demultiplexed by Fulgent Genetics.
820 Demultiplexed forward and reverse reads were processed with QIIME 2 2020.2 (*49*). Loading of
821 sequence data was performed with the demux plugin followed by quality filtering and denoising
822 with the dada2 plugin (*50*). Dada2 trimming parameters were set to the base pair where the
823 average quality score dropped below thirty. All samples were rarefied to the lowest read depth
824 present in all samples (48,305 reads) to decrease biases from variable sequencing depth across
825 samples (*51*). The q2-feature-classifier was then used to assign taxonomy to amplicon sequence
826 variants (ASV) with the Silva 132 99% OTUs references (*52, 53*). Resulting read count tables
827 were used for downstream analyses in IPython notebooks.

828 **Absolute abundance**

829 The total microbial load (bacteria and archaea) of each sample and the absolute
830 abundance of each taxon in individual samples was determined as described previously (*36,*
831 *47*). Briefly, the Bio-Rad QX200 droplet dPCR system (Bio-Rad Laboratories) was utilized to
832 measure the 16S concentration in each sample with the following reaction components: 1X
833 QX200 EvaGreen Supermix (Bio-Rad), 500 nM forward primer, and 500 nM reverse primer (Uni

834 16S 519F, Uni 16S 806R) and thermocycling conditions: 95 °C for 5 min, 40 cycles of 95 °C for 30
835 s, 52 °C for 30 s, and 68 °C for 30 s, followed by a dye stabilization step of 4 °C for 5 min and 90
836 °C for 5 min. The final concentration of 16S rRNA gene copies in each sample was corrected for
837 dilutions and normalized to the extracted sample total DNA measurement from NanoDrop. We
838 used total DNA levels as a proxy for tissue mass in biopsy samples. For each sample, the input-
839 DNA-normalized total microbial load from dPCR was multiplied by each amplicon sequence
840 variant's (ASV) relative abundance to determine the absolute abundance of each ASV.

841 **Poisson quality filtering**

842 Two separate quality-filtering steps based on Poisson statistics were used to determine
843 the statistical confidence in the measured values. First, a 95% confidence interval was
844 calculated from the repeated measures of water blanks. Samples with a total microbial load
845 below the upper bound of this confidence interval were removed from further analysis.

846 Second, the limit of detection (LOD) in terms of relative abundance was determined for
847 each sample. Sequencing can be divided into two separate Poisson sampling steps. First, an
848 aliquot of sample is taken from the extracted sample and input into the library amplification
849 reaction. The LOD of the library amplification step (LOD_{amp}) was determined by multiplying the
850 total microbial load from dPCR by the input volume into the library amplification reaction and
851 then finding the relative abundance corresponding to an input of three copies. Poisson statistics
852 tells us that the likelihood of sampling one or more copies with an average input of three copies
853 is 95%. The second Poisson sampling step in sequencing arises from the number of reads
854 generated from the amplified library. Here, the relative abundance at which 95% confidence of
855 detection is observed was previously shown to follow a negative exponential curve with respect

856 to the total read depth: $LOD_{seq} = 7.115 * read_depth^{-0.115}$ (*36*) The minimum LOD =
857 $min(LOD_{amp}, LOD_{seq})$ of the two described LODs was then determined for each sample. For each
858 sample, the abundance of any ASV with a relative abundance below the LOD was set to zero.
859 After this filtering step, data tables for each taxonomic level were generated.

860 **Statistical analysis and correlations**

861 Group comparisons were analyzed using the non-parametric Kruskal-Wallis rank sum
862 test with Benjamini–Hochberg multiple hypothesis testing correction using *SciPy.stats Kruskal*
863 function and *statsmodels.stats.multitest multipletests* function with the *fdr_bh* option.

864 **Purified IECs RNA-seq**

865 To perform RNAseq on IECs, 100,000 EPCAM+ CD45- cells were cell sorted from the
866 jejunum and ileum of WT and $GATA4^{\Delta IEC}$ mice. Three independent cell sorting experiments were
867 performed and the libraries and sequencing were done on the same batch with 8 mice per
868 group. The SMART-Seq v4 Ultra Low Input RNA Kit (TaKaRa) was used to generated amplified
869 cDNA, using 7500 pg of RNA input. The cDNA was generated and purified according to the
870 manufacturer's specifications, and cDNA was amplified 12 cycles. The Nextera XT DNA Library
871 Preparation Kit (Illumina) was used to generate the RNA-seq libraries, with an input of 125 pg
872 cDNA, according to the manufacturer's specifications. Subsequently, the libraries were
873 multiplexed and sequenced at a depth of 20 million reads per sample (50bp SR) on a
874 HiSeq4000.

875 **Whole tissue RNA-seq**

876 Whole tissue biopsies stored at –80 °C were thawed on ice and transferred to Starstedt
877 tubes containing 350uL RLT Plus supplemented with 1% 2-mercaptoethanol and equal

878 quantities of 1.0 mm and 0.5 mm zirconium oxide beads (Next Advance). Biopsies were bead
879 beat 3 times for 1 min at a setting of 9 on a Bullet Blender 24, with 1 min of cooling on ice
880 between each beating. Lysates were processed using the AllPrep DNA/RNA/miRNA Universal Kit
881 (Qiagen). To generate sample libraries, 500 ng of purified RNA was used as input in the TruSeq
882 Stranded mRNA Library Prep kit (Illumina) according to manufacturer's specifications. Libraries
883 were multiplexed and sequenced at a depth of 20 million reads per sample (50 bp SR) on a
884 HiSeq4000.

885 **Mouse RNA-seq data processing to obtain raw counts**

886 Mouse RNA-seq raw data were processed using a standard workflow based on the
887 GENPIPES framework (*54*). Specifically, the "stringtie" type "rnaseq" pipeline was used. Reads
888 were first trimmed using Trimmomatic software (*55*). Trimmed reads were aligned to the
889 Mus_musculus.GRCm38 mouse reference genome using the STAR aligner (*56*) following a two-
890 pass mapping protocol. Alignments were then sorted and filtered for duplicates using
891 Picard(sort, markduplicates) ("Picard Toolkit" Broad Institute
892 http://broadinstitute.github.io/picard/; Broad Institute). Gene-level read counts for
893 downstream processing were calculated from spliced alignments using HTseq count (*57*).

894 **RNA-seq data analysis: quality control filtering and normalization**

895 All statistical analyses of the mouse RNA-seq data were performed using R(v4.0.3).The
896 raw count matrices contained the transcript counts, with columns representing samples and
897 rows indexed by genes. Genes expressed (*i.e.*, having at least two counts) in fewer than two
898 samples were removed. The resulting matrices will be referred to as the count matrices. The
899 variance stabilizing transformation from DESeq2 R package (*58*) was applied to normalize the

900 count matrices. Batch effects were removed using the *removeBatchEffect()* function from the
901 limma R package (*59*).

902 **RNA-seq data analysis: differential expression and gene set enrichment**

903 A negative binomial generalized linear model in the DESeq2 package was used to test
904 whether genes were differentially expressed (DE) between sample groups. The analysis takes
905 the count matrix as an input and estimates size factors and dispersions prior to fitting the
906 model. Wald statistics were used to determine the significance and the log2-fold change (LFC)
907 of the fit for each gene. We used the Benjamini-Hochberg method for reducing the false
908 discovery rate (FDR), with a cutoff of <0.05 for identifying differentially expressed genes for
909 further analysis. *P* values reported below are FDR adjusted. The LFCs and FDR-adjusted *P* values
910 were given as input to the *fgsea()* function from the fgsea R package (*60*), which implements a
911 preranked gene set enrichment analysis. The rankings of the genes were based on the FDR-
912 adjusted *P* values. The mouse KEGG pathway database (mmuKegg) (*61*) and/or the Gene
913 Ontology, Biological Processes (GO-BP) database (*62, 63*) were the gene sets used in the
914 enrichment analysis. Enriched pathways (*i.e.*, *P* <0.05) were collapsed to independent pathways
915 to avoid repetition, using the fgsea *collapsePathways()* function (*60*).

916 *(i) Identifying region-specific GATA4-regulated genes*

917 We determined this set of DE genes by grouping together "ileum-like" samples, *i.e.*, WT
918 ileum and GATA$^{\Delta IEC}$ jejunum samples, and comparing them with WT Jejunum samples. This was
919 done separately for the tissue and EC RNA-seq analyses. A threshold of >0.25 for the absolute
920 value of the LFC was used to filter the very high number of DE genes in the tissue RNA-seq data,
921 whereas no LFC threshold was used for the EC data.

922 *(ii) Microbiota dependent and independent genes*

923 We determined the influence of microbiota on the region-specific GATA4-regulated
924 genes by first systematically comparing jejunum tissue samples to identify: (1) DE genes
925 between GATA$^{\Delta IEC}$ SPF in comparison to WT SPF, (2) DE genes between GATA$^{\Delta IEC}$ GF in
926 comparison to WT GF, and (3) genes whose expression indicated a significant interaction
927 between genotype (*i.e.*, KO vs WT) and microbiota (*i.e.*, GF vs SPF).

928 (a) Microbiota dependent genes: Genes that were strongly DE (*i.e.*, P <0.01, |LFC| > 0.6)
929 in Group 1 but not DE (*i.e.*, P >0.2) in Group 2, or vice versa, were deemed microbiota
930 dependent. Further, genes in group 3 that had a strong, significant interaction term (*i.e.*, P
931 <0.01, |LFC| > 0.9) were also included.

932 (b) Microbiota independent genes: Genes that were strongly DE (i.e., P <0.01, |LFC| >
933 0.6) in both groups were deemed microbiota independent. Further, genes in group 3 that also
934 had a weak or insignificant interaction term (i.e., P >0.2 or |LFC| < 0.9) were also added to the
935 list of microbiota independent genes.

936 Both these sets of genes were intersected with the region-specific, GATA4-regulated genes,
937 resulting in region-specific, GATA4-regulated genes that are either microbiota dependent or
938 independent.

939 *(iii) Genes regulated by retinoic acid*

940 We determined this set of DE genes by grouping together RA-treated, GATA$^{\Delta IEC}$ jejunum
941 and vehicle-treated WT jejunum samples, and comparing them to vehicle-treated GATA$^{\Delta IEC}$
942 jejunum samples. Due to the small number (4) of samples per group, many genes were dropped
943 due to the FDR threshold. To get a better view of the overall trends, DE genes were instead

944 determined in this step by thresholding on the unadjusted p-value (P <0.05). As a sanity check,
945 we observed that using the same P value threshold to compare just vehicle-treated WT jejunum
946 samples with vehicle-treated GATA$^{\Delta IEC}$ jejunum samples resulted in a set of DE genes similar to
947 the region-specific GATA4-regulated DE genes identified in the first analysis.

948 **RNA-seq data analysis: data visualization**

949 *(i) Principal components plots*

950 The principal components analysis (PCA) was done using the *pca()* function from the
951 PCAtools R package (Blighe K, Lun A, https://github.com/kevinblighe/PCAtools). The top 500
952 genes selected by highest row variance in the centered (across rows) normalized count matrix
953 were used to calculate the principal components. Using the full set of genes to calculate the
954 PCA shows the same patterns but with a larger spread amongst sample groups.

955 *(ii) Heatmaps*

956 Heatmaps were plotted using the *Heatmap()* function from the ComplexHeatmap R
957 package (*64*) . The z-scored (i.e., across rows) normalized expression values were used to plot
958 the heatmaps.

959 **Annotation of GATA4-targeted genes**

960 A previously published and publicly available table of annotated GATA4 ChIP-Seq peaks
961 (*11*) was used to annotate GATA4 targets among the DE genes reported.

962 **Generation of IL-17 and IFNγ gene modules**

963 We curated an IL-17 and IFNg modules, which consisted of known gene pathways from
964 established databases and publications. The IL-17 module consisted of genes encompassing the
965 following pathways from mmuKEGG: 04657, 04659 and msigDB: m6335, m19422, m298, m300,

966 m461, m460, m39560, m8578, m8581, m8579, m8927, m8928, and the following papers (*12,*
967 *18*).
968 The IFNg module consisted of genes encompassing the following pathways from
969 msigDB: m22085, m5972, m5970, m4551, m9583 , m39363, m161, m6305, m6313, m6696,
970 m6695, m6689, m6688, m6523, m6522, m6513, m6512, m1898, m2913, m8662, m8657,
971 m5913, from the Gene Ontology database: GO:0034341, from the Reactome database: R-
972 913531, and the following papers (*65, 66*).
973 **Human celiac RNA-seq analysis**
974 Adaptors and low-quality bases were trimmed using Trim Galore (v 0.4.4). Resulting
975 reads were then aligned to the human reference sequence Ensembl GRCh38 release 87 using
976 Kallisto (*67*). Next, the derived pseudo counts were normalized into log2 counts per million
977 reads (CPM) using the voom function from the limma package (v3.46.0) (*59*). To evaluate the
978 transcriptomic changes associated to GATA4 de-regulation in the small intestine and in the
979 context of celiac disease we defined two contrast groups: GATA4 low, to reflect loss of
980 regionalization, and GATA4 hi, as a normal jejunum tissue. To define these two groups, we used
981 the normalized expression of GATA4 were then used to rank all control, ACeD and GFD samples
982 to define two contrast groups of samples. First, by grouping the top 30% GATA4 expressing
983 samples into the GATA4 hi group (ACeD n =6, Control n= 18, GFD n= 18) and, second by
984 considering the bottom 30% of GATA4 expression only from ACeD samples, (n=15) which we
985 defined as the GATA4 low group. Next, to define a universe set for the enrichment analysis we
986 generated a set of 11,657 homologous genes from the human and mouse cohort. We then
987 tested transcriptome wide for significant differences in expression across the GATA4 hi and

988 GATA4 low groups of samples, by using a linear model while accounting for sex, age, batch and
989 technical covariates and correcting for multiple testing. Over-enrichment analysis of gene
990 ontologies biological processes were performed using the enrichGO function from the
991 clusterProfiler (v3.0.4) (*68*). All statistical analyses within the human cohort and the overlaps
992 with the mouse GATA4$^{\Delta IEC}$ dataset were performed using R (v4.0.3).

993 Using on the log2(CPM) expression values, we calculated single sample Gene Set
994 Enrichment scores (ssGSEA) for the Retinol pathway, IL17 downstream genes and the hallmark
995 (*69*), using the R Bioconductor Package GSVA (*70*). These ssGSEA scores where then used to test
996 association between GATA hi vs GATA lo contrasts groups, as well as its association with
997 presence or absence of bacteria. Associations where evaluated using a linear model while
998 accounting for sex, age and technical covariates.

999

1000

1001

1002

1003

1004

1005

1006

1007 **Table 1**

Reagent or Resource	**Source**	**Identifier**
Antibodies		
CD45 Pac Blue (30-F11)	Biolegend	Cat#103126
TCRgd FITC (eBioGL3)	Thermo Fisher	Cat#11-5711-82

CD4 BV785 (GK1.5)	Biolegend	Cat#100453
CD4 BV605 (GK1.5)	Biolegend	Cat#100451
CD8β BUV395 (H35-17.2)	BD	Cat#740278
CD8α PerCp/Cy5.5 (53-6.7)	BD	Cat#551162
NK1.1 PE-CF594 (PK136)	BD	Cat#562864
TCRβ BUV737 (H57-597)	BD	Cat#612821
TCRβ BV711 (H57-597)	BD	Cat#563135
CD3ε BUV737 (145-2C11)	BD	Cat#612771
IFNγ APC (XMG1.2)	BD	Cat#554413
IL10 PEcy7 (JES5-16E3)	Biolegend	Cat#505026
IL17a PE (ebio17B7)	Thermo Fisher	Cat#12-7177-81
CD45.1 Pac Blue (A20)	Biolegend	Cat#110722
CD45.2 BUV395 (104)	BD	Cat#553772
vβ14 TCR FITC (14-2)	BD	Cat#553258
RORγt BV786 (Q31-37)	BD	Cat#564723
FOXP3 eflour450 (FJK-16s)	Thermo Fisher	Cat#48-5773-82
FOXP3 FITC (FJK-16s)	Thermo Fisher	Cat#11-5773-82
FOXP3 PE-cy7 (FJK-16s)	Thermo Fisher	Cat#25-5773-82
Tbet PE (4B10)	Biolegend	Cat#644810
CD44 PE-CY7 (IM7)	Biolegend	Cat#103030
CD62L PE (MEL-14)	Biolegend	Cat#104408
Epcam PerCp/Cy5.5 (G8.8)	Biolegend	Cat#118220
CD19 FITC (1D3/CD19)	Biolegend	Cat#152404
NK1.1 BV605 (PK136)	Biolegend	Cat#108753
CD11C BV605 (N418)	Biolegend	Cat#117334
TER119 BV605 (TER-119)	Biolegend	Cat#116239
F4/80 BV605 (BM8)	Biolegend	Cat#123133
CD3ε BV605 (145-2C11)	Biolegend	Cat#100351
Ly6g BV605 (1A8)	Biolegend	Cat#127639
B220 PE-cy7 (RA3-6B2)	Biolegend	Cat#103222
IgA PE (mA-6E1)	Thermo Fisher	Cat#12-4204-81
IgA AF647 goat polyclonal	Southern Biotech	Cat#1040-31
Biological Samples		
Fetal Bovine Serum	Biowest	Cat#S01520; Lot#A11504E
Normal Goat Serum	JacksonImmunoResearch	Cat#005-000-121
Chemicals, Peptides, and		

Recombinant Proteins		
EDTA, 0.5M, pH8.0	Corning	Cat#46-034-CI
1M MgCl2	Thermo Fisher	Cat#AM9530G
Cytiva Percoll™ Centrifugation Media	GE Healthcare	Cat#45-001-747
RNAprotect Tissue Reagent	Qiagen	Cat#76106
2-Mercaptoethanol (BME)	Sigma-Aldrich	Cat#M7154
Phorbol Myristate Acetate	Sigma-Aldrich	Cat#P1585
Ionomycin Calcium Salt from Streptomyces conglobatus	Sigma-Aldrich	Cat#10634
BD GolgiStop Protein Transport Inhibitor	BD	Cat#554724
Ethanol 200 Proof	Decon Labs Inc	Cat#DSP-MD 43
Inhibitex Buffer	Qiagen	Cat#19593
Nuclease-free Water	Ambion	Cat#AM9932
Carnoy Solution	Ricca Chemical	Cat#R18510004C
10% Formalin Solution	Thermo Fisher	Cat#SF98-4
RPMI 1640 with L-Glutamine	Corning	Cat#MT-10043CV
Collagenase from Clostridium histolyticum	Sigma-Aldrich	Cat#C2139-500MG
1M TRIS-HCL pH 7.5	Thermo Fisher	Cat#15567027
10% SDS solution	Thermo Fisher	Cat#15553027
Sodium chloride	Sigma-Aldrich	Cat#S9888
Critical Commercial Assays		
SytoBC	Thermo Fisher	Cat#S34855
LIVE/DEAD® Fixable Aqua Dead Cell Stain Kit	Thermo Fisher	Cat#L34966
Zombie NIR™ Fixable Viability Kit	Biolegend	Cat#423106
BD Cytofix/Cytoperm Plus Fixation/Permeabilization Solution Kit	BD	Cat#554714
eBioscience™ Foxp3 / Transcription Factor Staining Buffer Set	Thermo Fisher	Cat#00-5523-00
QIAamp Fast DNA Stool Mini Kit	Qiagen	Cat#51604
RNeasy Plus Mini Kit	Qiagen	Cat#74136
RNeasy Micro Kit	Qiagen	Cat#74004
GoScript Reverse Transcriptase Kit	Promega	Cat#A5001
SYBR Advantage qPCR Premix	Clontech	Cat#639676

ProLong™ Diamond Antifade Mountant with DAPI	Thermo Fisher	Cat#P36962
Pierce™ Protein L Magnetic Beads	Thermo Fisher	Cat#88850
Pierce™ IgG Elution Buffer, pH 2.0	Thermo Fisher	Cat#21028
1.0 M Tris HCl pH 8.5	VWR	Cat#76236-402
ALDEFLUOR™ Kit	Stemcell technologies	Cat#01700
IgA mouse ELISA	Thermo Fisher	Cat#88-50450-86
QX200 droplet digital PCR system	BioRad	Cat#1864001
QX200 ddPCR EvaGreen Supermix	BioRad	Cat#1864034
5Prime Hotstart Mastermix	QuantaBio	Cat#2200410
Evagreen	Biotium	Cat#31000
KAPA library quantification kit	Roche	Cat#07960140001
AmpureXP beads	Beckman Coulter	Cat#A63880
Oligonucleotides		
GAPDH Forward 5′-AGGTCGGTGTGAACGGATTTG-3′	(*38*)	N/A
GAPDH Reverse 5′-TGTAGACCATGTAGTTGAGGTCA-3′	(*38*)	N/A
IFNγ Forward 5′-ATGAACGCTACACACTGCATC-3′	(*38*)	N/A
IFNγ Reverse 3′-TCTAGGCTTTCAATGACTGTGC-5′	(*38*)	N/A
ASL Forward 5’-TCTTCGTTAGCTGGCAACTCACCT-3’	(*71*)	N/A
ASL Reverse 5’-ATGACCCAGCAGCTAAGCAGATCA-3’	(*71*)	N/A
Uni 16S 340F 5’-ACTCCTACGGGAGGCAGCAGT-3’	(*42*)	N/A
Uni 16S 514R 5’-ATTACCGCGGCTGCTGGC-3’	(*42*)	N/A
SFB 736F 5’-GACGCTGAGGCATGAGAGCAT-3’	(*42*)	N/A
SFB 844R 5’-GACGGCACGGATTGTTATTCA-3’	(*42*)	N/A
ASF457F 5’ -TGCAAGAATGAAACTCAAAGGAAT-3’	(*72*)	N/A

ASF457R 5'-TAAGGTTCTTCGGTTAGCATCGA-3'	(*72*)	N/A
C. rodentium espBF 5'-ATGCCGCAGATGAGACAGTTG-3'	(*73*)	N/A
C. rodentium espBR 5'-CGTCAGCAGCCTTTTCAGCTA-3'	(*73*)	N/A
EUB338 5'-Alexa546-GCTGCCTCCCGTAGGAGT-3'	(*74*)	N/A
EUB338 5'-Cy3-GCTGCCTCCCGTAGGAGT-3'	(*74*)	N/A
SFB1008 5'-Alexa488-GCGAGCTTCCCTCATTACAAGG-3'	(*42*)	N/A
SFB1008 5'-Cy5-GCGAGCTTCCCTCATTACAAGG-3'	(*42*)	N/A
Uni 16S 519F 5'-CAGCMGCCGCGGTAA	(*36*)	N/A
Uni 16S 806R 5'-GGACTACHVGGGTWTCTAAT	(*36*)	N/A
Other		
Bead ruptor elite bead mill homogenizer	Omni International	Cat#19-040E
Tissue-Tearor	Biospec Products	Cat#985370-XL
Glass Beads	Biospec Products	Cat#11079101
LightCycler® 480 System	Roche	N/A
LSRFortessa™ X-20 Flow Cytometer	BD	N/A
Cytek® Aurora	Cytek	N/A
CRi Pannoramic SCAN 40x Whole Slide Scanner	3DHistech	N/A
EasyEights™ EasySep™ Magnet	Stemcell technologies	Cat#18103
CFX96 Touch RT-PCR detection system	BioRad	Cat#1855196
Software and Algorithms		
QIIME2: 2020.2.0	(*49*)	
Silva 132 SSURef NR99	(*53*)	
Python 3.7.6	https://www.python.org/	
Scipy 1.4.1	https://www.scipy.org/	
Statsmodels 0.10.1	https://www.statsmodels.org/stable/index.html	

Numpy 1.18.1	https://numpy.org/	
Pandas 1.0.3	https://pandas.pydata.org/	
GraphPad Prism 8	https://www.graphpad.com	
FlowJo	https://www.flowjo.com	

Acknowledgements:

We would like to thank the patients and their family members, as well as the University of Chicago Celiac Disease Center, for supporting our research. We would like to thank Betty Theriault, Kristin Kolar, and the entire Gnotobiotic Research Animal Facility at the University of Chicago for their help. We also thank the Human Tissue Resource Center, the Integrated Light Microscopy Core Facility, Flow Cytometry Facility, and the Genomics Facility at the University of Chicago. We would like to thank Kenya Honda for providing rat SFB, and Gabriel Nuñez for ΔEAE

1027 *C. rodentium*. We would like to thank Ivaylo Ivanov for providing SFB TCR transgenic mice and
1028 for helpful discussions. We would like to thank Toufic Mayassi, and Cezary Ciszewski for their
1029 thoughtful discussions and insightful comments. Finally, we thank Valerie Abadie for help
1030 editing the manuscript. This work was supported by the National Institutes of Health via T32
1031 AI007090 to Z.M.E., T32 GM007281 to D.G.S., U01AI109695 for A.M., U01 AI125250 to A.B.,
1032 R01 AI144094 to A.B., R01DK103761 to N.S., R01 DK067180 to B.J., and the Digestive Diseases
1033 Research Core Center P30 DK42086 at the University of Chicago to B.J.
1034

1035 **Author Contributions:**

1036 Z.M.E. and B.J. conceived the study and designed the experiments. N.S., A.M. and P.M, based
1037 on their previously published work and studies performed in gnotobiotic mice, put forward the
1038 concept that absence of GATA4 in epithelial cells was driving inflammatory immune responses
1039 in the jejunum in a microbiota-dependent manner. Z.M.E., W.L., A.K., D.G.S., J.D.E., and J.T.B.
1040 performed the experiments. Z.M.E., W.L., J.J.S, R.A.G., J.T.B., analyzed and interpreted the data.
1041 J.J.S., R.A.G., performed the RNAseq analysis, J.T.B., performed the 16S sequencing and
1042 analysis. D.G.S. and V.D. created the RNAseq libraries, and S.G. performed the DNA alignments.
1043 P.M. supervised the initial GATA4 transcriptomic analysis. I.L.T contributed to and L.B.B
1044 supervised RNA seq analysis of human celiac disease patients, and S.J.R. supervised the overall
1045 RNAseq analysis. R.F.I. supervised the 16S analysis. A.B. supervised the IgA experiments. Z.M.E,
1046 S.J.R, B.J. wrote the manuscript. W.L., J.J.S, J.T.B., R.F.I., A.M., N.S., and A.B. edited the
1047 manuscript. R.F.I., A.M., A.B., N.S., and B.J. acquired funds to support the work; and B.J.
1048 directed the study.

1049

1050 **Declaration of Interests:**

1051 The authors declare no competing interests

1052

1053

1054

1055

1056

1057

1058

1059

1060

1061

1062

1063
1064
1065 **References**

1066 1. A. M. Mowat, W. W. Agace, Regional specialization within the intestinal immune system.
1067 *Nature reviews. Immunology* **14**, 667-685 (2014).
1068 2. Ivanov, II *et al.*, Induction of intestinal Th17 cells by segmented filamentous bacteria.
1069 *Cell* **139**, 485-498 (2009).
1070 3. Z. Shi *et al.*, Segmented Filamentous Bacteria Prevent and Cure Rotavirus Infection. *Cell*
1071 **179**, 644-658.e613 (2019).
1072 4. D. Esterházy *et al.*, Compartmentalized gut lymph node drainage dictates adaptive
1073 immune responses. *Nature* **569**, 126-130 (2019).
1074 5. P. M. Smith *et al.*, The microbial metabolites, short-chain fatty acids, regulate colonic
1075 Treg cell homeostasis. *Science (New York, N.Y.)* **341**, 569-573 (2013).

1076 6. M. E. Johansson *et al.*, The inner of the two Muc2 mucin-dependent mucus layers in
1077 colon is devoid of bacteria. *Proceedings of the National Academy of Sciences of the*
1078 *United States of America* **105**, 15064-15069 (2008).
1079 7. T. Bosse *et al.*, Gata4 is essential for the maintenance of jejunal-ileal identities in the
1080 adult mouse small intestine. *Mol Cell Biol* **26**, 9060-9070 (2006).
1081 8. M. A. Battle *et al.*, GATA4 is essential for jejunal function in mice. *Gastroenterology* **135**,
1082 1676-1686.e1671 (2008).
1083 9. K. Breuer *et al.*, InnateDB: systems biology of innate immunity and beyond--recent
1084 updates and continuing curation. *Nucleic Acids Res* **41**, D1228-1233 (2013).
1085 10. B. E. Aronson *et al.*, GATA4 represses an ileal program of gene expression in the
1086 proximal small intestine by inhibiting the acetylation of histone H3, lysine 27. *Biochimica*
1087 *et biophysica acta* **1839**, 1273-1282 (2014).
1088 11. C. A. Thompson *et al.*, GATA4 Is Sufficient to Establish Jejunal Versus Ileal Identity in the
1089 Small Intestine. *Cellular and molecular gastroenterology and hepatology* **3**, 422-446
1090 (2017).
1091 12. K. Atarashi *et al.*, Th17 Cell Induction by Adhesion of Microbes to Intestinal Epithelial
1092 Cells. *Cell* **163**, 367-380 (2015).
1093 13. Y. Umesaki, H. Setoyama, S. Matsumoto, A. Imaoka, K. Itoh, Differential roles of
1094 segmented filamentous bacteria and clostridia in development of the intestinal immune
1095 system. *Infect Immun* **67**, 3504-3511 (1999).
1096 14. V. Gaboriau-Routhiau *et al.*, The key role of segmented filamentous bacteria in the
1097 coordinated maturation of gut helper T cell responses. *Immunity* **31**, 677-689 (2009).
1098 15. Y. Yang *et al.*, Focused specificity of intestinal TH17 cells towards commensal bacterial
1099 antigens. *Nature* **510**, 152-156 (2014).
1100 16. T. Prakash *et al.*, Complete genome sequences of rat and mouse segmented filamentous
1101 bacteria, a potent inducer of th17 cell differentiation. *Cell host & microbe* **10**, 273-284
1102 (2011).
1103 17. G. P. Donaldson *et al.*, Gut microbiota utilize immunoglobulin A for mucosal
1104 colonization. *Science (New York, N.Y.)* **360**, 795-800 (2018).
1105 18. P. Kumar *et al.*, Intestinal Interleukin-17 Receptor Signaling Mediates Reciprocal Control
1106 of the Gut Microbiota and Autoimmune Inflammation. *Immunity* **44**, 659-671 (2016).
1107 19. N. Shulzhenko *et al.*, Crosstalk between B lymphocytes, microbiota and the intestinal
1108 epithelium governs immunity versus metabolism in the gut. *Nature medicine* **17**, 1585-
1109 1593 (2011).
1110 20. A. Goncalves *et al.*, Fat-soluble vitamin intestinal absorption: absorption sites in the
1111 intestine and interactions for absorption. *Food Chem* **172**, 155-160 (2015).
1112 21. F. M. Herr *et al.*, Intestinal vitamin A metabolism: coordinate distribution of enzymes
1113 and CRBP(II). *Journal of lipid research* **34**, 1545-1554 (1993).
1114 22. J. R. Mora *et al.*, Generation of gut-homing IgA-secreting B cells by intestinal dendritic
1115 cells. *Science (New York, N.Y.)* **314**, 1157-1160 (2006).
1116 23. H. B. Jijon *et al.*, Intestinal epithelial cell-specific RARα depletion results in aberrant
1117 epithelial cell homeostasis and underdeveloped immune system. *Mucosal Immunol* **11**,
1118 703-715 (2018).

1119 24. S. Gattu *et al.*, Epithelial retinoic acid receptor β regulates serum amyloid A expression
1120 and vitamin A-dependent intestinal immunity. *Proceedings of the National Academy of*
1121 *Sciences* **116**, 10911 (2019).
1122 25. B. Jabri, L. M. Sollid, Tissue-mediated control of immunopathology in coeliac disease.
1123 *Nature reviews. Immunology* **9**, 858-870 (2009).
1124 26. I. Monteleone *et al.*, Characterization of IL-17A-producing cells in celiac disease mucosa.
1125 *Journal of immunology (Baltimore, Md. : 1950)* **184**, 2211-2218 (2010).
1126 27. A. Castellanos-Rubio *et al.*, TH17 (and TH1) signatures of intestinal biopsies of CD
1127 patients in response to gliadin. *Autoimmunity* **42**, 69-73 (2009).
1128 28. M. Bodd *et al.*, HLA-DQ2-restricted gluten-reactive T cells produce IL-21 but not IL-17 or
1129 IL-22. *Mucosal Immunol* **3**, 594-601 (2010).
1130 29. T. Mayassi *et al.*, Chronic Inflammation Permanently Reshapes Tissue-Resident
1131 Immunity in Celiac Disease. *Cell* **176**, 967-981.e919 (2019).
1132 30. P. Brar *et al.*, Change in lipid profile in celiac disease: beneficial effect of gluten-free diet.
1133 *Am J Med* **119**, 786-790 (2006).
1134 31. A. C. Bledsoe *et al.*, Micronutrient Deficiencies Are Common in Contemporary Celiac
1135 Disease Despite Lack of Overt Malabsorption Symptoms. *Mayo Clin Proc* **94**, 1253-1260
1136 (2019).
1137 32. A. Caminero *et al.*, Duodenal bacterial proteolytic activity determines sensitivity to
1138 dietary antigen through protease-activated receptor-2. *Nat Commun* **10**, 1198 (2019).
1139 33. M. C. Cenit, M. Olivares, P. Codoñer-Franch, Y. Sanz, Intestinal Microbiota and Celiac
1140 Disease: Cause, Consequence or Co-Evolution? *Nutrients* **7**, 6900-6923 (2015).
1141 34. M. Vuoristo, Y. A. Kesäniemi, H. Gylling, T. A. Miettinen, Metabolism of cholesterol and
1142 apolipoprotein B in celiac disease. *Metabolism* **42**, 1386-1391 (1993).
1143 35. M. Xu *et al.*, c-MAF-dependent regulatory T cells mediate immunological tolerance to a
1144 gut pathobiont. *Nature* **554**, 373-377 (2018).
1145 36. J. T. Barlow, S. R. Bogatyrev, R. F. Ismagilov, A quantitative sequencing framework for
1146 absolute abundance measurements of mucosal and lumenal microbial communities. *Nat*
1147 *Commun* **11**, 2590 (2020).
1148 37. V. D'Argenio *et al.*, Metagenomics Reveals Dysbiosis and a Potentially Pathogenic N.
1149 flavescens Strain in Duodenum of Adult Celiac Patients. *Am J Gastroenterol* **111**, 879-890
1150 (2016).
1151 38. V. Abadie *et al.*, IL-15, gluten and HLA-DQ8 drive tissue destruction in coeliac disease.
1152 *Nature* **578**, 600-604 (2020).
1153 39. S. Husby *et al.*, European Society for Pediatric Gastroenterology, Hepatology, and
1154 Nutrition Guidelines for the Diagnosis of Coeliac Disease. *Journal of Pediatric*
1155 *Gastroenterology and Nutrition* **54**, (2012).
1156 40. L. Lefrançois, N. Lycke, Isolation of mouse small intestinal intraepithelial lymphocytes,
1157 Peyer's patch, and lamina propria cells. *Curr Protoc Immunol* **Chapter 3**, Unit 3.19
1158 (2001).
1159 41. J. J. Bunker *et al.*, Innate and Adaptive Humoral Responses Coat Distinct Commensal
1160 Bacteria with Immunoglobulin A. *Immunity* **43**, 541-553 (2015).
1161 42. T. Sano *et al.*, An IL-23R/IL-22 Circuit Regulates Epithelial Serum Amyloid A to Promote
1162 Local Effector Th17 Responses. *Cell* **163**, 381-393 (2015).

1163 43. Z. M. Earley *et al.*, Burn Injury Alters the Intestinal Microbiome and Increases Gut
1164 Permeability and Bacterial Translocation. *PloS one* **10**, e0129996 (2015).
1165 44. J. M. Pickard *et al.*, Rapid fucosylation of intestinal epithelium sustains host-commensal
1166 symbiosis in sickness. *Nature* **514**, 638-641 (2014).
1167 45. M. J. Molloy *et al.*, Intraluminal containment of commensal outgrowth in the gut during
1168 infection-induced dysbiosis. *Cell host & microbe* **14**, 318-328 (2013).
1169 46. J.T. Barlow *et. al*, Quantitative sequencing clarifies the role of disruptor taxa, oral
1170 microbiota, and strict anaerobes in the human small-intestine microbiome. *Microbiome*,
1171 (2021).
1172 47. S. R. Bogatyrev, R. F. Ismagilov, Quantitative microbiome profiling in lumenal and tissue
1173 samples with broad coverage and dynamic range via a single-step 16S rRNA gene DNA
1174 copy quantification and amplicon barcoding. *bioRxiv*, 2020.2001.2022.914705 (2020).
1175 48. S. R. Bogatyrev, J. C. Rolando, R. F. Ismagilov, Self-reinoculation with fecal flora changes
1176 microbiota density and composition leading to an altered bile-acid profile in the mouse
1177 small intestine. *Microbiome* **8**, 19 (2020).
1178 49. E. Bolyen *et al.*, Reproducible, interactive, scalable and extensible microbiome data
1179 science using QIIME 2. *Nat Biotechnol* **37**, 852-857 (2019).
1180 50. B. J. Callahan *et al.*, DADA2: High-resolution sample inference from Illumina amplicon
1181 data. *Nat Methods* **13**, 581-583 (2016).
1182 51. S. Weiss *et al.*, Normalization and microbial differential abundance strategies depend
1183 upon data characteristics. *Microbiome* **5**, 27 (2017).
1184 52. N. A. Bokulich *et al.*, Optimizing taxonomic classification of marker-gene amplicon
1185 sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* **6**, 90 (2018).
1186 53. C. Quast *et al.*, The SILVA ribosomal RNA gene database project: improved data
1187 processing and web-based tools. *Nucleic Acids Res* **41**, D590-596 (2013).
1188 54. M. Bourgey *et al.*, GenPipes: an open-source framework for distributed and scalable
1189 genomic analyses. *Gigascience* **8**, (2019).
1190 55. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina
1191 sequence data. *Bioinformatics* **30**, 2114-2120 (2014).
1192 56. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21
1193 (2013).
1194 57. S. Anders, P. T. Pyl, W. Huber, HTSeq--a Python framework to work with high-
1195 throughput sequencing data. *Bioinformatics* **31**, 166-169 (2015).
1196 58. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for
1197 RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
1198 59. M. E. Ritchie *et al.*, limma powers differential expression analyses for RNA-sequencing
1199 and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).
1200 60. G. Korotkevich *et al.*, Fast gene set enrichment analysis. *bioRxiv*, 060012 (2021).
1201 61. R. H. Waterston *et al.*, Initial sequencing and comparative analysis of the mouse
1202 genome. *Nature* **420**, 520-562 (2002).
1203 62. M. Ashburner *et al.*, Gene ontology: tool for the unification of biology. The Gene
1204 Ontology Consortium. *Nat Genet* **25**, 25-29 (2000).
1205 63. The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res* **49**, D325-d334
1206 (2021).

1207 64. Z. Gu, R. Eils, M. Schlesner, Complex heatmaps reveal patterns and correlations in
1208 multidimensional genomic data. *Bioinformatics* **32**, 2847-2849 (2016).
1209 65. S. Y. Liu, D. J. Sanchez, R. Aliyari, S. Lu, G. Cheng, Systematic identification of type I and
1210 type II interferon-induced antiviral factors. *Proceedings of the National Academy of*
1211 *Sciences of the United States of America* **109**, 4239-4244 (2012).
1212 66. J. L. Benci *et al.*, Tumor Interferon Signaling Regulates a Multigenic Resistance Program
1213 to Immune Checkpoint Blockade. *Cell* **167**, 1540-1554.e1512 (2016).
1214 67. N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq
1215 quantification. *Nat Biotechnol* **34**, 525-527 (2016).
1216 68. G. Yu, L. G. Wang, Y. Han, Q. Y. He, clusterProfiler: an R package for comparing biological
1217 themes among gene clusters. *Omics* **16**, 284-287 (2012).
1218 69. A. Liberzon *et al.*, The Molecular Signatures Database (MSigDB) hallmark gene set
1219 collection. *Cell Syst* **1**, 417-425 (2015).
1220 70. S. Hänzelmann, R. Castelo, J. Guinney, GSVA: gene set variation analysis for microarray
1221 and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).
1222 71. C. S. N. Klose *et al.*, Differentiation of type 1 ILCs from a common progenitor to all
1223 helper-like innate lymphoid cell lineages. *Cell* **157**, 340-356 (2014).
1224 72. A. Deloris Alexander *et al.*, Quantitative PCR assays for mouse enteric flora reveal strain-
1225 dependent differences in composition that are influenced by the microenvironment.
1226 *Mamm Genome* **17**, 1093-1104 (2006).
1227 73. S. Sagaidak, A. Taibi, B. Wen, E. M. Comelli, Development of a real-time PCR assay for
1228 quantification of Citrobacter rodentium. *J Microbiol Methods* **126**, 76-77 (2016).
1229 74. R. I. Amann *et al.*, Combination of 16S rRNA-targeted oligonucleotide probes with flow
1230 cytometry for analyzing mixed microbial populations. *Appl Environ Microbiol* **56**, 1919-
1231 1925 (1990).
1232

Figure 1. Intestinal epithelial cells control microbiota-dependent and independent regionalization of tissue immunity through GATA4

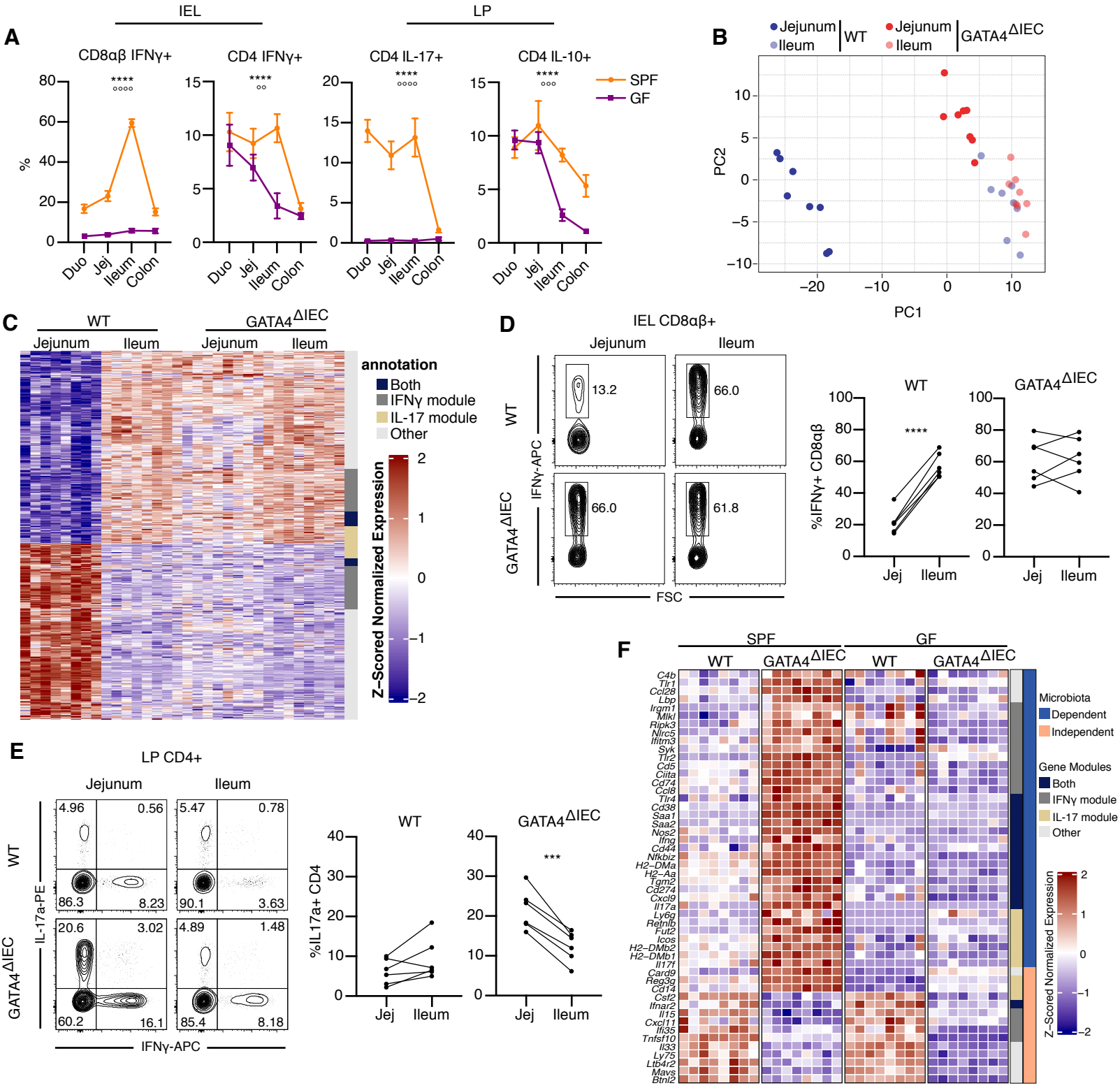

Figure 2. GATA4 controls the colonization of segmented filamentous bacteria to shape tissue immunity

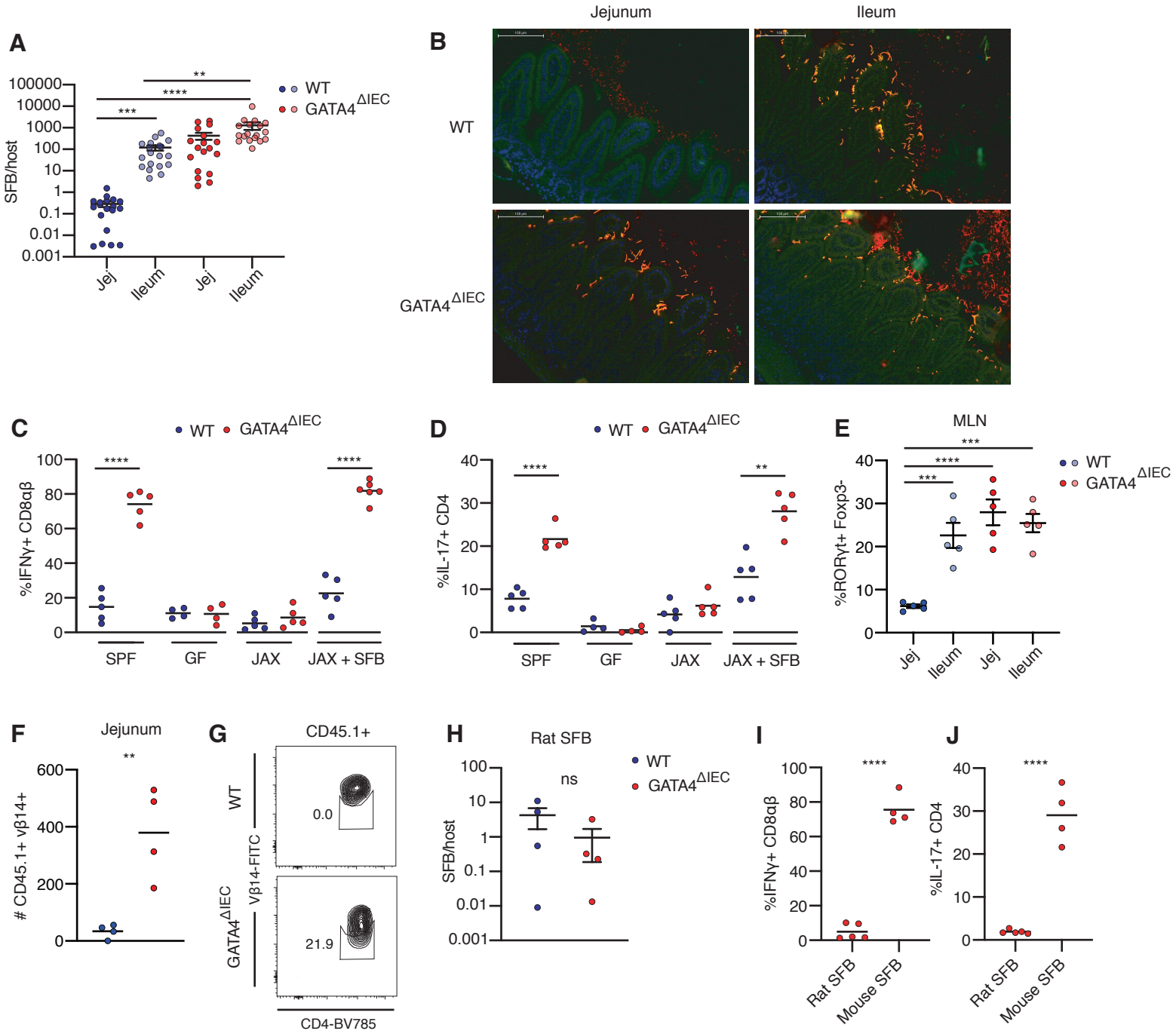

Figure 3. GATA4 regulates regionalization of retinol metabolism and IgA to limit SFB colonization in the proximal intestine

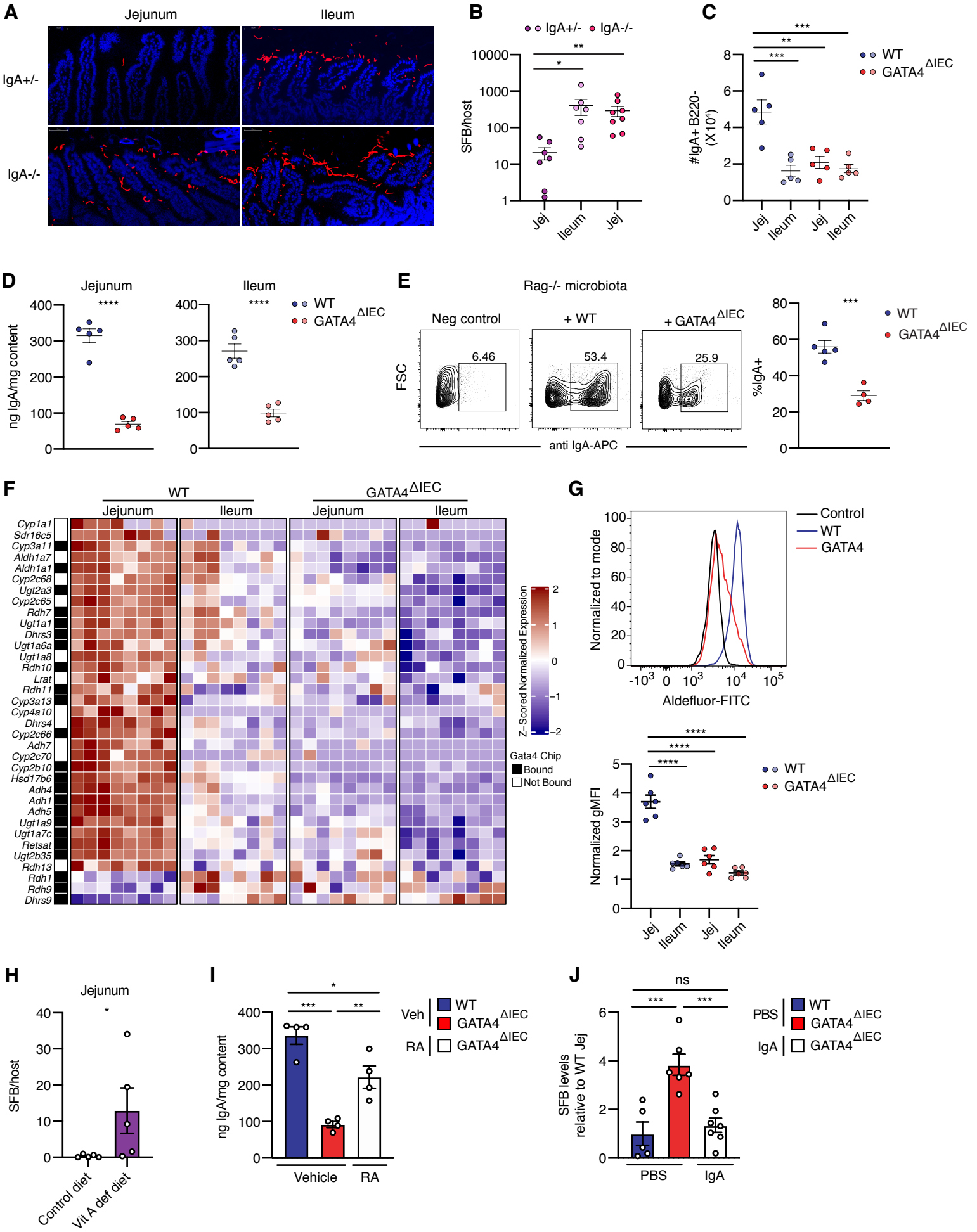

Figure 4. Loss of intestinal regionalization impacts host susceptibility to enteropathogenic infection by *Citrobacter rodentium*

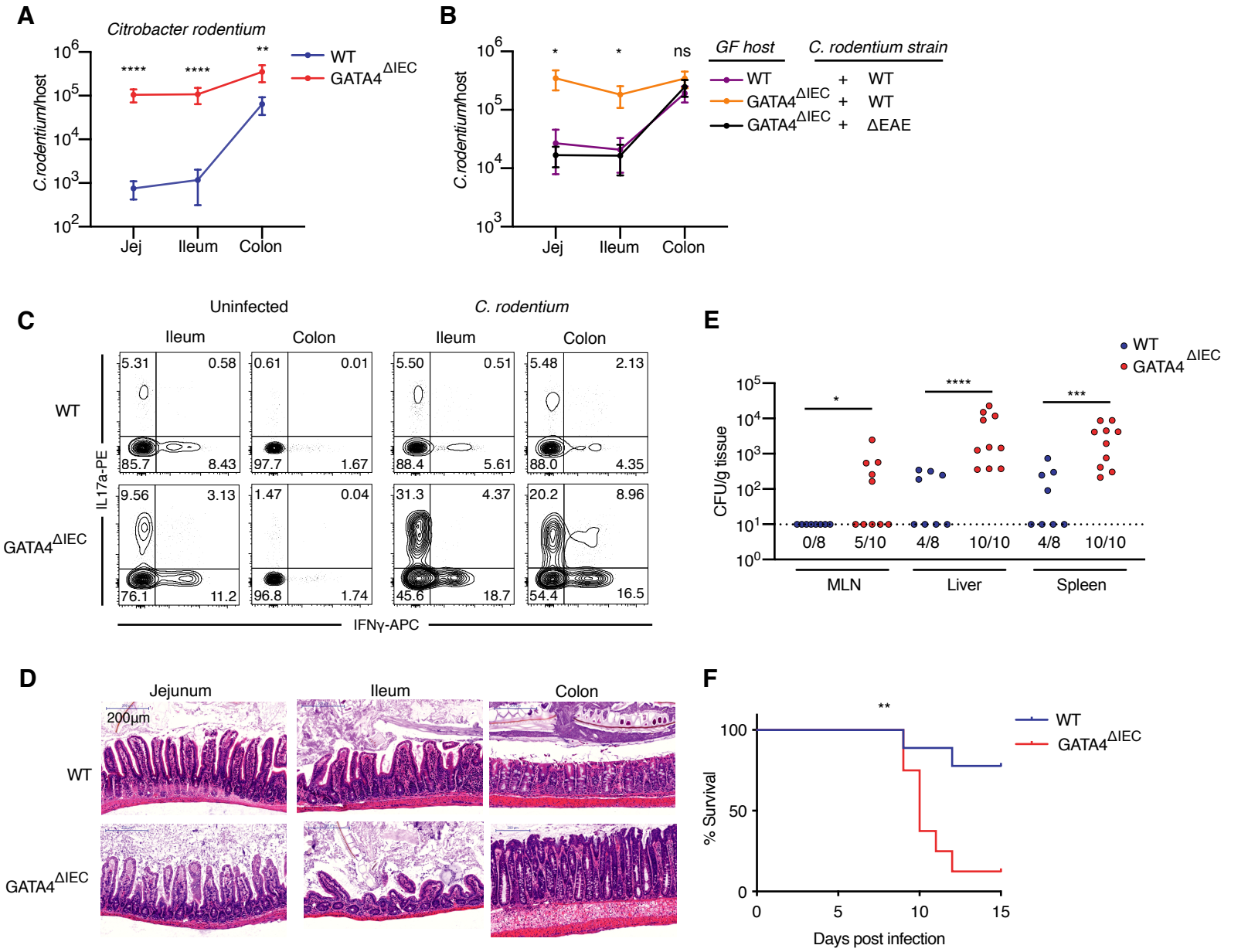

Figure 5. Loss of GATA4 expression is associated with lipid metabolic dysfunction and increased IL-17 signaling in celiac disease

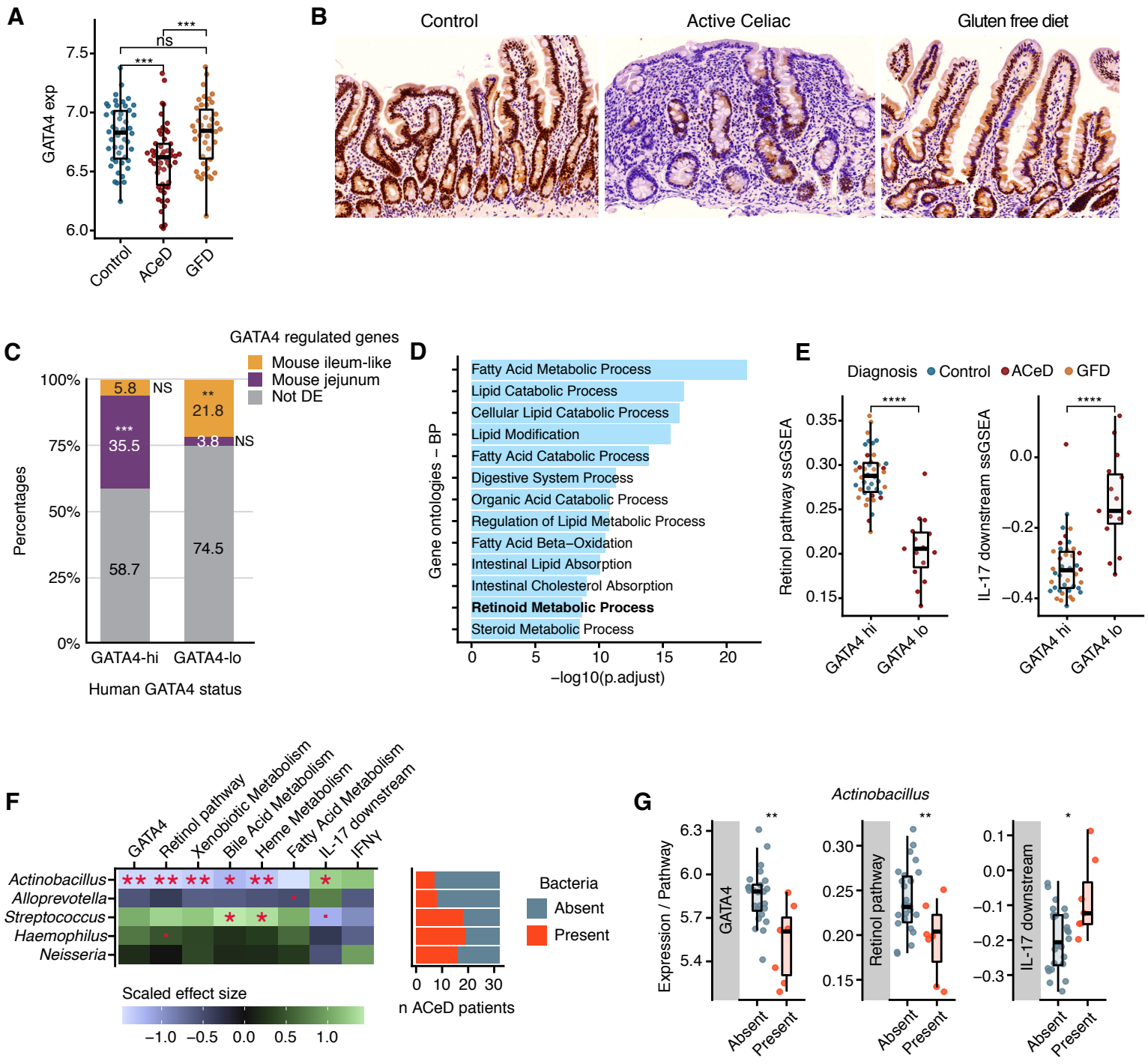

Figure S1.GATA4 controls regionalization of tissue metabolism and immunity

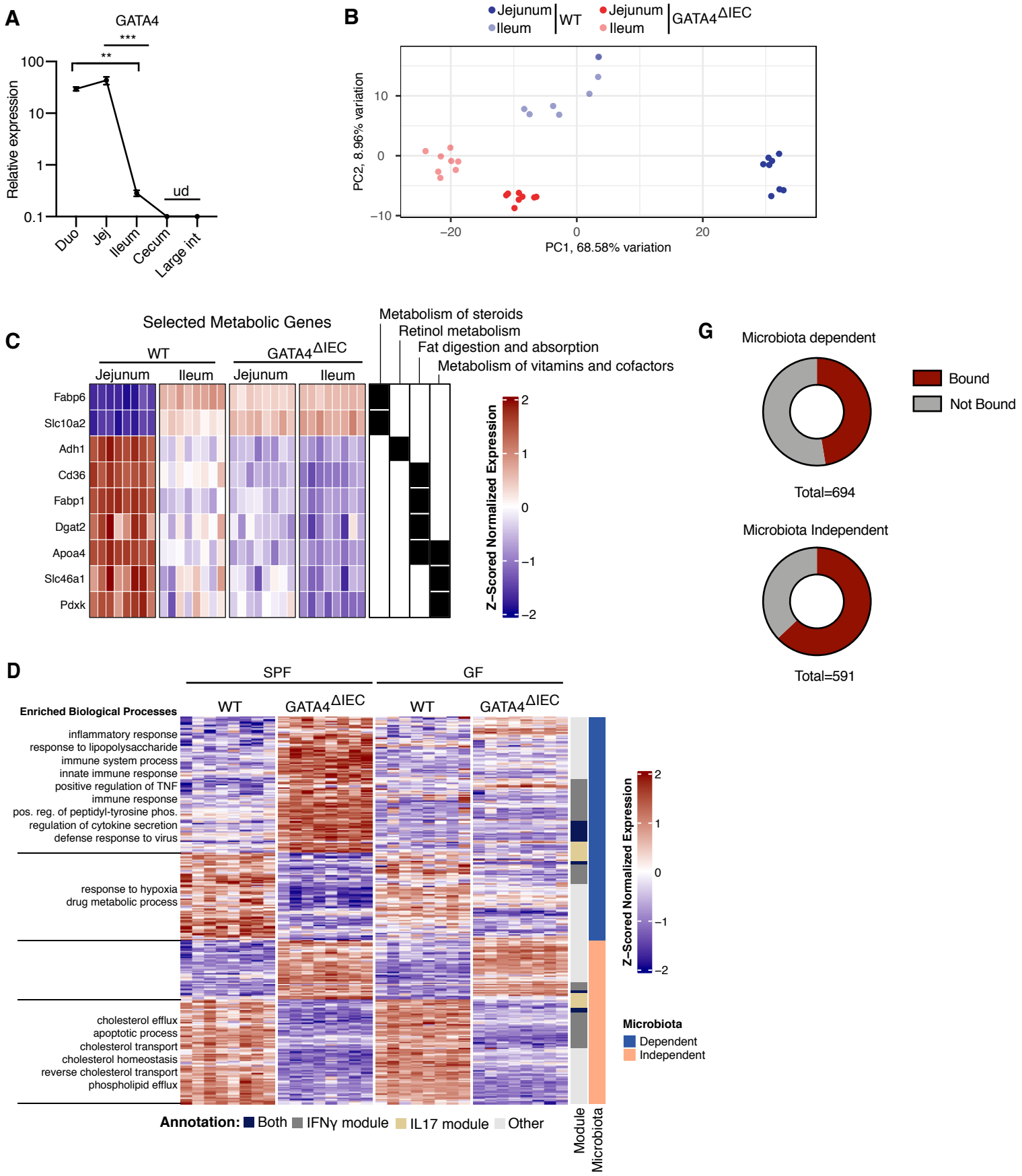

Figure S2. GATA4 controls the colonization of regional mucosal associated bacteria to shape tissue immunity

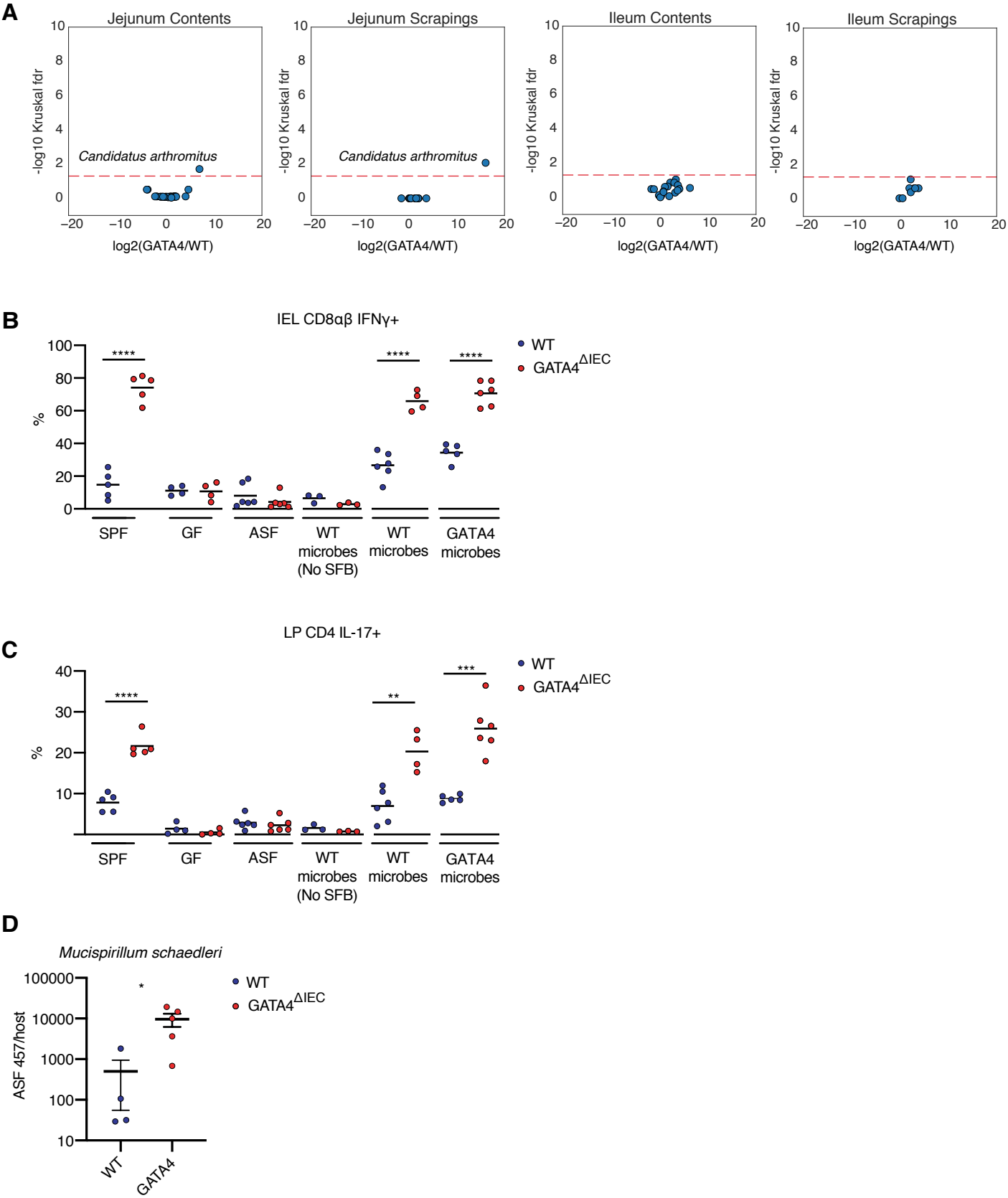

Figure S3. GATA4 regulates regionalization of retinol metabolism and B cell responses

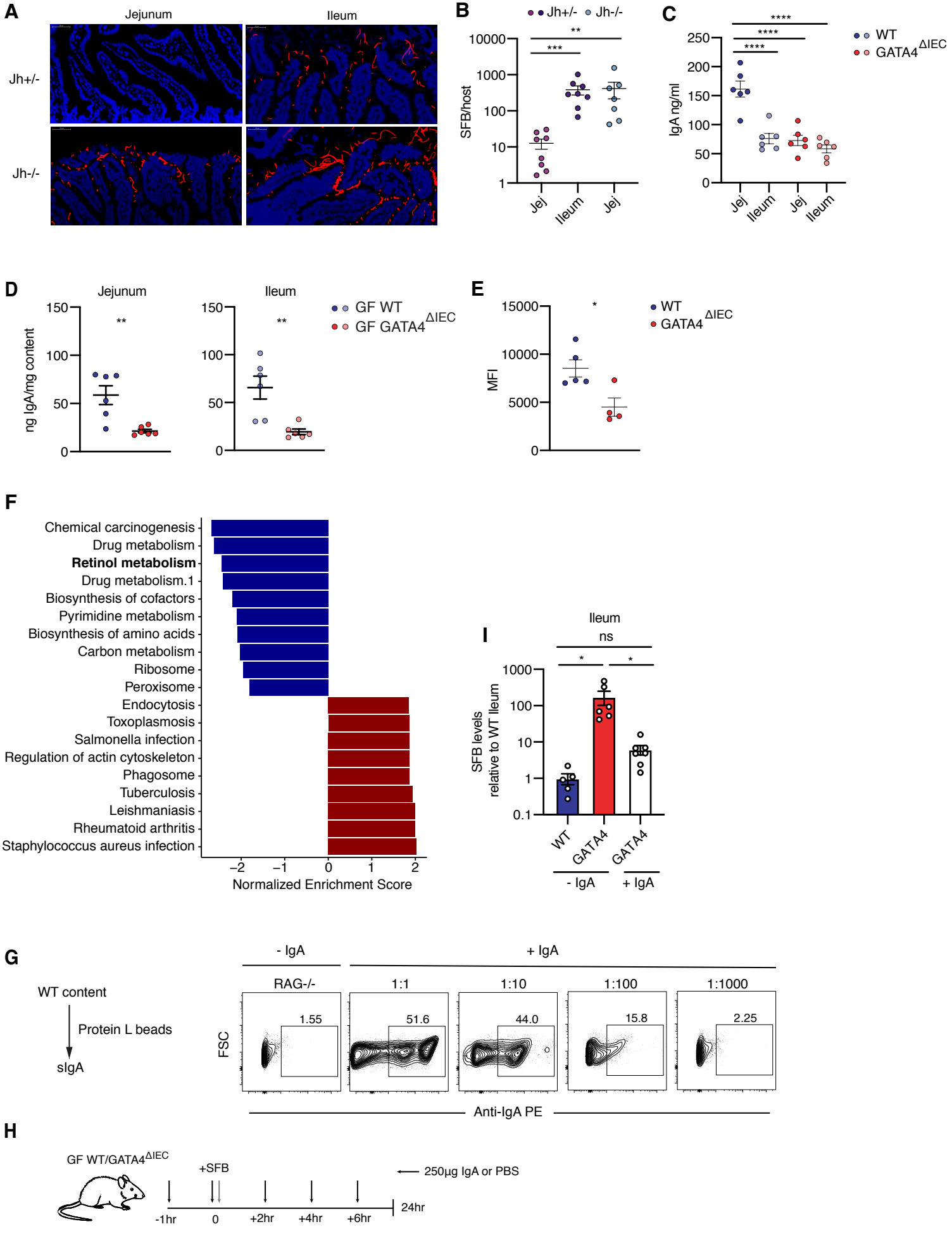
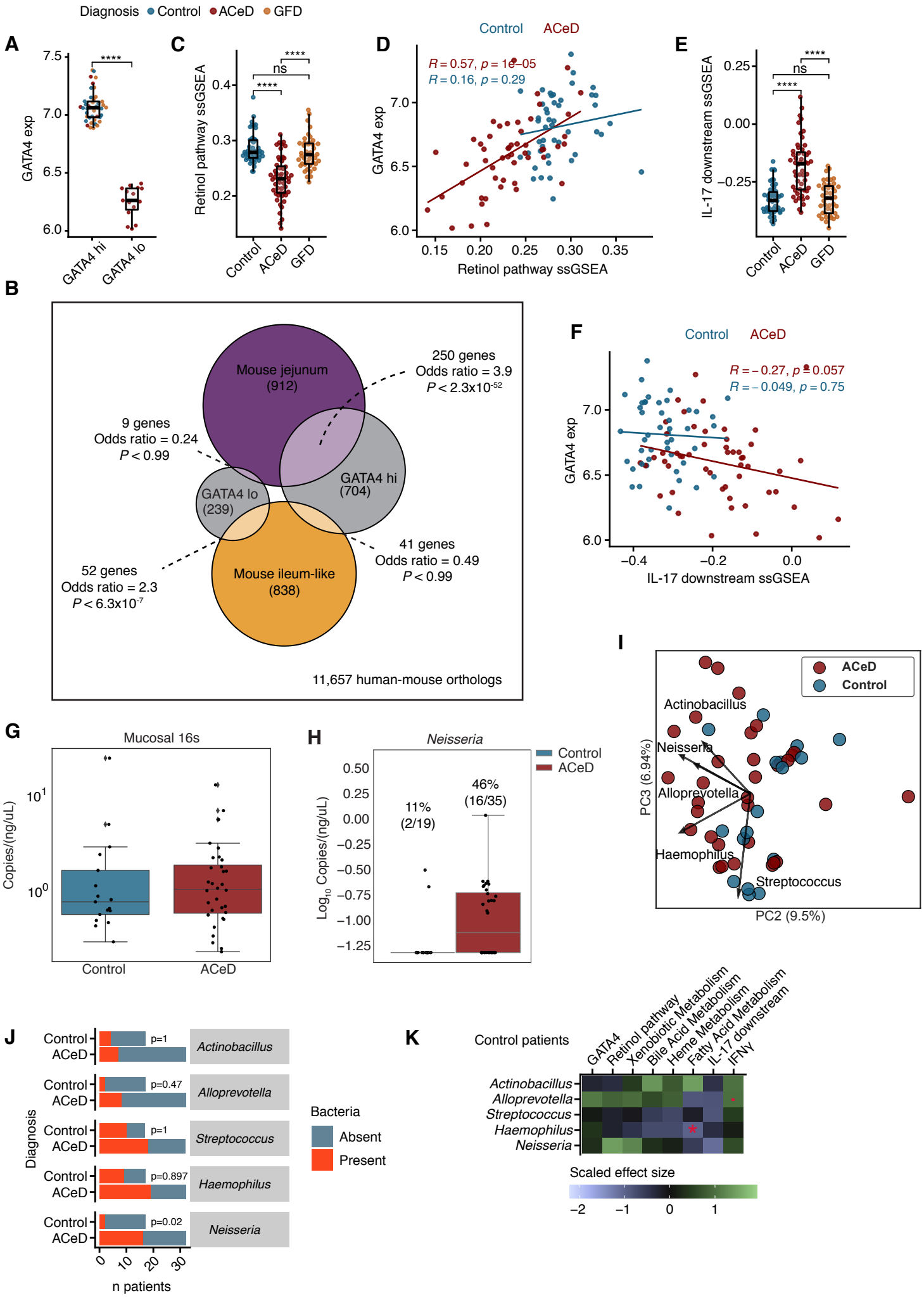
Figure S4. Loss of GATA4 is associated with lipid metabolic defects, mucosal associated bacteria, and increased IL-17 signaling in celiac disease patients