Children Developing Celiac Disease Have a Distinct and Proinflammatory Gut Microbiota in the First 5 Years of Life

Qian Huang[1], Yi Yang[2], Vladimir Tolstikov[3], Michael A. Kiebish[3], Jonas F Ludvigsson[4,5], Noah W. Palm[2], Johnny Ludvigsson[6], Emrah Altindis[1]

Affiliations:

[1] Boston College Biology Department, Chestnut Hill, MA 02467, USA

[2] Department of Immunobiology, Yale University School of Medicine, New Haven, CT 06510, USA.

[3] BERG, LLC, Framingham, MA, USA.

[4] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

[5] Department of Paediatrics, Örebro University Hospital, Sweden

[6] Crown Princess Victoria's Children's Hospital, Region Östergötland, Division of Pediatrics, Linköping University, Linköping, SE 58185, Sweden.

Correspondence to: Emrah Altindis, Boston College Biology Department, Higgins Hall, 140 Commonwealth Avenue Chestnut Hill, MA 02467. E-mail: altindis@bc.edu

ABSTRACT

Objective: Celiac disease (CD) is an immune-mediated disease characterized by small intestinal inflammation. Although CD is associated with HLA-DQ2 and HLA-DQ8 haplotypes, genetics alone cannot explain the increasing incidence of CD. The main objective of this study was to determine the role of the gut microbiota in CD pathogenesis in the first five years of life.

Design: We conducted a longitudinal study using fecal samples representing three developmental phases of the gut microbiota (ages 1, 2.5 and 5 years). These samples were obtained from 16 children who developed CD and 16 HLA/breastfeeding-matched controls in Sweden born 1997-1999. We used comparative 16S sequencing combined with functional analysis, immunoglobulin A (IgA) separation-based sequencing (IgA-seq), and plasma metabolomics to determine a microbial link to CD pathogenesis.

Results: We identified a distinct gut microbiota composition in CD progressors (children who develop CD during or after their gut microbiota were sampled) in each developmental phase. Pathogenesis and inflammation-related microbial pathways were enriched in CD progressors samples. Moreover, CD progressors had significantly more IgA coated bacteria compared to healthy controls and the IgA targets were significantly different. Among others, *Coprococcus comes*, *Bacteroides finegoldii* and *Faecalibacterium prausnitzii* were the unique top targets of the IgA response in CD progressors. Proinflammatory and pathogenesis related metabolic pathways were enriched in CD progressors. Further, we identified inflammatory metabolites, particularly microbiota-derived taurodeoxycholic acid (TDCA) as increased in CD progressors.

Conclusion: Our study defines an inflammatory gut microbiota for the CD progressors in terms of its composition, function, IgA response and related metabolites. The inflammatory nature of CD gut microbiota during development is potentially related to the onset of the disease and

52 targeting inflammatory bacteria in this critical window could affect the pathogenesis and
53 prognosis of CD. Targeting inflammation related bacteria in the early stages of CD development
54 could be a preventative tool.

55 **Significance of this study**

56 **What is already known on this subject?**

57 • Celiac Disease (CD) is a gluten induced immune-mediated disease in genetically
58 predisposed individuals.

59 • CD incidence is increasing worldwide which genetics alone cannot explain. Previous
60 studies have shown that the gut microbiota of CD patients differ from that of healthy
61 populations. However, the role of the microbiome in CD pathogenesis and its role in
62 chronic inflammation is yet be established.

63 **What are the new findings?**

64 • In a prospective longitudinal study in children using samples representing all three phases
65 of gut microbiota development (ages 1, 2.5 and 5), we identified significant differences in
66 the composition and function of gut microbiota at each phase. Pathogenesis and
67 inflammation-related functions are enriched in the gut microbiome of CD progressors.

68 • We applied IgA-sequencing to identify inflammatory bacteria in both healthy subjects
69 and CD progressors. Flow Cytometry analysis identified more IgA coated bacteria at ages
70 1 and 5 in CD progressors, indicating an early inflammatory response. CD bacterial IgA
71 targets also differed significantly from healthy controls.

72 • We analyzed plasma metabolites obtained at age 5. The CD plasma metabolome was
73 significantly different from healthy controls. Particularly, proinflammatory plasma

74 metabolites, including microbiota-derived taurodeoxycholic acid (TDCA) and isobutyryl-
75 L-carnitine, were increased two-fold in CD progressors.

76 **How might it impact clinical practice in the foreseeable future?**

77 • Our results establish a link between gut microbiota composition and chronic
78 inflammation in CD during child development. The highly IgA-coated bacteria identified
79 in IgA sequencing and inflammatory bacteria potentially contribute to CD pathogenesis.
80 Targeting these bacteria in the early stages of CD development could be a preventative
81 tool.

82 • TDCA is a microbiota-derived proinflammatory metabolite increased two-fold in CD
83 progressors. Increased TDCA levels may be used as a predictive/diagnostic tool in
84 genetically predisposed subjects. Moreover, targeting TDCA-producing bacteria (e.g.,
85 Clostridium XIVa species) could potentially help to control the intestinal inflammation in
86 CD.

87 • Developing anti-inflammatory probiotics/prebiotics might be viable therapeutics for
88 altering microbiota composition in children genetically predisposed for CD. These
89 microbes/compounds may also complement a gluten-free diet in patients that continue to
90 experience persistent CD symptoms.

91

92

93

94

95 **INTRODUCTION**

96 Celiac disease (CD) is a gluten-induced autoimmune disorder that is predicted to affect 1
97 in 100 individuals worldwide[1]. The adaptive autoimmune response in CD is characterized by
98 gluten-specific CD4+ T cell and antibodies against gluten gliadin peptide and the enzyme tissue
99 transglutaminase (tTG)[2] responsible for deamidating the gliadin peptide[1]. This biochemical
100 reaction increases the immunogenicity of gliadin peptides[3]. Antigen-presenting cells (APCs)
101 present gliadin peptides to T cells and cause mucosal lesions in the small intestine[4]. Almost all
102 CD patients possess HLA-DQ2 or HLA-DQ8. Although 20%- 40% of the population in Europe
103 and the USA carries these alleles, only 1% of individuals develop the disease[5]. These findings
104 suggest that the presence of HLA-DQ2 or HLA-DQ8 genes are necessary but not sufficient for
105 the development of CD and thus environmental factors also play a role in disease onset[4]. King et
106 al. recently showed that the incidence of CD to be increasing by 7.5% per year in the last
107 decades[6]. Furthermore, even among twins, the concordance of CD is not 100 %[7 8].

108 This evidence indicates that CD is a multifactorial disease and environmental factors play
109 a role in CD onset. Various environmental factors are implicated in CD development[9], but the
110 roles of these environmental factors in CD progression remain largely unknown. Gut
111 microbiome studies observe an altered microbial[10-12] and metabolite composition[13-15] in both
112 infant and adult CD patients but have not identified any causal link[16-19]. Immunoglobulin A (IgA)
113 is the most abundant antibody isotype at mucosal surfaces and is a major mediator of intestinal
114 immunity in humans[20]. IgA-sequencing (IgA-seq) combines bacterial flow cytometry with high-
115 throughput sequencing to identify distinct subsets of highly IgA coated (IgA+) and non-coated
116 microbiota (IgA-)[21-24]. It was previously shown in a mice model that IgA+ microbiota could
117 induce more severe colitis than IgA- microbiota[25]. Similarly, IgA-seq identified *Escherichia coli*

118 as an inflammatory bacterium enriched in Crohn's disease-patients with spondyloarthritis[26].
119 Therefore, we hypothesized that the pathogenic bacteria and some immunoregulatory
120 commensals involved in CD onset would be highly coated with IgA (IgA+) while most of the
121 commensals would not be coated (IgA-). We also hypothesized that the immune response would
122 have specific targets in CD progressors that differ from the control targets.
123 In this study, we assessed the composition and function of the gut microbiota in a
124 prospective, longitudinal cohort of 32 children matched for human leukocyte antigen (HLA)
125 genotype and breastfeeding duration (n=16/group). we focused on samples obtained at ages 1,
126 2.5 and 5 because these samples represent the three stages of gut microbiota development[27]. We
127 then identified the functional pathways enriched in CD progressors, determined the targets of the
128 IgA in the gut microbiota using IgA-seq. Lastly, we compared the plasma metabolome and
129 identified significant differences. Our findings demonstrate that children who go on to develop
130 CD have significant alterations in their gut microbiome years before diagnosis. CD-associated
131 gut microbiota are enriched in inflammatory- and pathogenicity-related bacteria, as well as
132 microbial functions and metabolites that potentially contribute to chronic inflammation in CD.
133
134 **EXPERIMENTAL PROCEDURE:**
135 **Human Fecal Samples**
136 The fecal samples were obtained from subjects in the All Babies in Southeast Sweden
137 (ABIS) cohort. ABIS study was ethically approved by the Research Ethics Committees of the
138 Faculty of Health Science at Linköping University, Sweden (Ref. 1997/96287 and 2003/03-092)
139 and the Medical Faculty of Lund University, Sweden (Dnr 99227, Dnr 99321). All children born
140 in southeast Sweden between 1st October 1997 and 1st October 1999 were recruited. Informed

141 consent from the parents was obtained. Fresh fecal samples were collected either at home or at
142 the clinic. Samples collected at home were stored at -20 °C with freeze clamps, mailed to the
143 WellBaby Clinic and stored dry at −80 °C. The questionnaire was completed by the parents to
144 collect participants' health information including, but not limited to, breast feeding duration,
145 antibiotic use, gluten exposure time, and more. In total 68 fecal samples were collected for the
146 analysis (10 at 1 year old, 32 at 2.5 years old and 26 at 5 years old).
147
148 **IgA+ and IgA- Bacteria Separation and Fecal IgA Flow Cytometry**
149 IgA+ and IgA- bacteria separation was performed as previously describe[25]. Briefly,
150 human fecal bacteria were stained with Anti-human IgA PE (clone IS11-8E10 Miltenyi Biotec)
151 followed by Magnetic Activated Cell Separation (MACS) or flow cytometric analysis.
152 Additional details are provided in the Supplemental Methods.
153
154 **Flow Cytometric analysis.**
155 Bacterial cells were isolated from fecal samples as described in the Supplemental
156 Methods and analyzed by flow cytometry using a BD FACSAria™ IIIu cell sorter (Becton-
157 Dickinson) as previously described[21].
158
159 **16S rRNA Gene Sequencing and Statistical Analyses**
160 16S rRNA sequencing of the V4 region sequencing for all bacteria samples were
161 performed on the Miseq platform with barcoded primers. Microbial diversity and statistical
162 analyses were performed with *data 2*, *phyloseq 2*, *vegan*, *edge,* and *PICRUSt2*. Additional details
163 are provided in the Supplemental Methods.

164

165 **Plasma Metabolomics and Metabolite pathway Analyses**

166 Preparation of plasma samples for metabolomics analysis and subsequent plasma
167 metabolomics and metabolite pathway analyses were performed as previously described[28-30].
168 Additional details are provided in the Supplemental Methods.

169

170

171 **RESULTS**

172 **Study Cohort**

173 All Babies in Southeast Sweden (ABIS) is a prospective population-based study that
174 established a large biobank of biological specimens obtained longitudinally at birth and ages 1,
175 2.5, and 5 [31]. To determine the role of gut microbiota in CD pathogenesis, we used ABIS
176 samples selecting a sub cohort of 32 individuals born 1997-1999 in Sweden. We chose 16
177 subjects who developed CD but were not diagnosed with any other autoimmune disease as of
178 December 2017. We matched this group with 16 healthy controls based on their HLA-risk class
179 distribution and breastfeeding duration (**Table S1**). The diagnosis of CD for 11 individuals
180 occurred after the age of 5, one subject was diagnosed at age 1.8 while the other four were
181 diagnosed between the ages of 2.5 and 5. Swedish National Patient Register[32] was used for
182 verifying diagnosis of CD according to international classification of disease (ICD) code-10
183 K90.0. In total, we used 68 longitudinal stool samples (**Table S1**). Although we did not match
184 subjects for other parameters, timing of gluten exposure, delivery method, breastfeeding duration,
185 family history of CD, infections and use of antibiotics were comparable between groups (**Table**
186 **S1**).

187 **Celiac Disease Progressors Have a Distinct Gut Microbiota Composition**

188 By sequencing the V4 region of the 16S rRNA gene[25], we identified 661 operational
189 taxonomic units (OTUs) (**Table S2**). Consistent with previous studies[33], gut microbiome alpha
190 diversity increased until age 2.5 and remained stable up to age 5 in both groups. However, alpha
191 diversity was significantly higher for the CD progressors at age 1, indicating a more diverse
192 microbiota composition (**Figure 1A**). Beta diversity was comparable between CD and healthy
193 controls in each phase (**Figure 1A**). Non-metric multidimensional scaling (NMDS) plots show a
194 trend of separation of gut microbiome composition between CD and healthy control individuals
195 at ages 1 and 2.5 (**Figure 1B**).

196 Relative abundance analyses of microbial taxa uncovered strong differences at both
197 phylum and genus levels between the CD progression and healthy microbiota, with the largest
198 differences occurring in the first year (**Figure 1C**). CD progressors had higher levels of
199 Firmicutes than controls, while Verrucomicrobia was only identified in control samples. Relative
200 abundances of the genera Prevotella, Romoboutsia, Roseburia, Rumminococcus, Ruminococcus2
201 (Ruminococcus of family Lachospiraceae), Streptococcus, and Veillonella were higher in the CD
202 progressors (**Fig 1C, D**). The Acinetobacter genus was highly enriched in controls but was
203 absent in CD progressors at age 1. However, these differences dissipated over time at both the
204 genus and phylum levels.

205 When we analyzed the differences at the level of OTUs, we identified 10 OTUs that are
206 significantly (FDR≤0.05, p<0.05) different at age 1. Clostridium XVIII, *Ruminococcus bromii*,
207 *Bifidobacterium dentium*, and *Clostridium XIVa sciendens* were highly enriched in CD
208 progressors while Enterococcus was highly enriched in control samples (**Figure S1A, Figure**
209 **1D**). We identified more significant differences in OTU level at ages 2.5 and 5. 133 OTUs were

210 different at age 2.5 and 112 OTUs were different at age 5 (**Figure 1E**). The most significantly
211 enriched OTUs in CD samples were *Dialister* and *Gemmiger* at age 2.5. On the other hand,
212 *Bacteroides eggerthii* and *Methanobrevibacter*, and *Ruminococcus2* were highly enriched in
213 healthy subjects (**Fig S1B**). Likewise, *Phascolarctobacterium faecium* and *Dialister* and
214 *Ruminococcus* were enriched in CD samples while *Prevotella* and *Holdemanella* were enriched
215 in healthy samples at age 5 (**Fig S1C**). These results demonstrate that the gut microbiota in CD
216 progressors are significantly different from healthy controls in the first 5 years of life.

217 **CD progressors Have More Bacteria Coated with IgA Indicating an Inflammatory**
218 **Gut Microbiota Composition**

219 To test our initial IgA hypothesis, we used a modified method of IgA-sequencing[1]
220 (**Figure S2A**). PCA analysis showed a clear separation between IgA+ and IgA- bacteria at all
221 ages both in control and CD samples (**Figure 2A).** We also identified an overall separation for
222 all samples (**Figure S2B)**. We confirmed this finding using flow cytometry (**Figure S2C)**. The
223 flow cytometry analysis revealed that the number of IgA+ bacteria was increased from a least
224 squares (LS) mean of 4.57% at age 1 to 8.88% at age 2.5 and maintained at 6.03% at age 5 in
225 controls. On the other hand, IgA+ bacteria was already LS mean 11.05% at age 1 in the CD
226 progressors, indicating a two fold increase compared to controls. It was 8.5 % at age 2.5,
227 comparable to control samples. At age 5, there was a two-fold increase compared to the controls
228 and 12.8 % of the bacteria was IgA+ in CD progressors (**Figure 2B**). These results reveal that
229 CD progressors have more IgA+ bacteria especially at age 5 ($p=0.026$). When we removed the
230 five CD progressors who developed CD before age 5 from the analysis, we had the same
231 significant result (**Figure S2D)**. This result indicates a more pathogenic gut microbiota
232 composition and a more inflammatory environment for CD progressors. Moreover, we also show

233 that only a small fraction of the microbiota are coated by IgA during development (~ 5-8 %) in
234 healthy controls and it is increased in the disease state (8.5-12 %).

235 **A Specific IgA Response to Bacteria Develops After Age 1**

236 Because there are very few reports on the IgA response in early human gut microbiota
237 development[23], we first focused on the results obtained from the healthy children. We did not
238 observe any difference between IgA+ and IgA- samples in the control group at age 1 (**Figure**
239 **S3A**). This result suggests that the IgA response does not target specific bacteria in this early
240 stage of development. Consistent with the flow cytometry analysis, we identified 113 OTUs at
241 age 2.5 and 43 OTUs at age 5 that were significantly different between IgA+ and IgA- samples
242 in healthy controls (**Figure S3A, Table S2**). The top targets of IgA in healthy subjects were
243 OTUs *Clostridium IV*, *Bifidobacterium* and *Bacteroides clarus* at age 2.5 and *Clostridium IV*,
244 *Gemmiger* and *Elizabethkingia* at age 5.

245 Meanwhile we identified only one OTU, Brucella, that was different between IgA+ and
246 IgA- samples in CD progressors at age 1. Notably, 121 and 41 OTUs were significantly different
247 at ages 2.5 and 5, respectively (FDR≤0.05, P<0.05; **Figure S3A**). CD progressors shared similar
248 top IgA targets with those in healthy subjects such as *Clostridium IV* at ages 2.5 and 5 and
249 *Gemmiger* at age 5. On the other hand, CD progressors showed unique targets of the IgA+
250 including *Coprococcus comes*, *Bacteroides finegoldii* and Methanobrevibacter at age 2.5 and
251 *Faecalibacterium prausnitzii*, *Clostridium XlVa* and *Streptococcus* at age 5.

252 **IgA Response Targets Are Comprised of Different Bacteria in CD progressors**

253 Consistent with the presorting data, the alpha diversity increased at age 2.5 and remained
254 stable in both control and CD IgA- groups (**Figure 2C)**. Likewise, beta diversity was comparable
255 between CD IgA- and control IgA- samples (**Figure 2C**). There was a separation of IgA- gut

256 microbiome composition between CD and healthy control individuals at ages 1 and 2.5 (**Figure**
257 **2D**). The first year IgA- microbiome composition was most different at the phylum and genus
258 levels between groups as observed in the presorting samples for the same time point. These
259 observed differences were less pronounced in the IgA+ samples and mostly dissipated over time
260 (**Figure 2E**). At age 2.5, IgA-seq identified 144 different OTUs between control IgA- samples
261 and CD IgA- samples. Likewise, we identified 167 different OTUs between control IgA+
262 samples and CD IgA+ samples (FDR<0.05, p<0.05, **Figure S3B)**. At age 5, 71 OTUs were
263 different between control IgA- and CD IgA- samples. Additionally, 112 different OTUs were
264 identified for CD IgA+ and control IgA+ samples (**Figure S3B**). The top differential targets of
265 the immune system were *Clostridium IV*, *Bacteroides finegoldii*, and *Dislister propionicifaciens*
266 at age 2.5 and *Enterobacter*, *Blautia*, and *Enterobacteriaceae* at age 5 in CD progressors.
267 In addition to the differences caused by altered gut microbiota composition, we also
268 identified 72 OTUs at age 2.5 and 45 OTUs at age 5 in which abundances were the same in the
269 gut microbiota of CD and healthy samples (presorting) but differentially targeted by the immune
270 system (**Table S3)**. For example, *Lachnospiraceae*, *Bacterioides finegoldii* (**Figure S3C, 1E**),
271 and *Bacteroides vulgatus* OTUs at age 2.5 and Enterobacter, Blautia (**Figure S3C**), and
272 Barnesiella OTUs at age 5 were coated with IgA in CD groups but not in controls. Overall, these
273 results indicate that not only gut microbiota composition, but also the IgA response to microbiota,
274 is altered in CD progressors.
275 **Pathogenesis and Inflammation Related Functions Are Enriched in CD Progressors'**
276 **Gut Microbiota**
277 Phylogenetic Investigation of Communities by Reconstruction of Unobserved States
278 (PICRUSt) analysis[34] is designed to estimate the functional metagenome of gut bacteria using

279 16S rRNA data (**Figure S4A**). Combining PICRUSt with Kyoto Encyclopedia of Genes and
280 Genomes (KEGG) metabolic pathway analysis, we identified 71 different metabolic pathways
281 that differed between CD and control samples at age 1 (FDR<0.05, **Figure 3A**). Among these
282 pathways, N-glycan biosynthesis, penicillin and cephalosporin biosynthesis, beta-Lactam
283 resistance and bacterial chemotaxis pathways were the top pathways enriched in CD progressors
284 (**Figure S5A**). Interestingly, most of these pathways are involved in bacterial pathogenesis[35 36] or
285 shaping the composition of microbiota[37 38]. We also identified 9 pathways showing strong trends
286 of difference at ages 2.5 and 5. For example, styrene degradation, lysine degradation, fatty acid
287 metabolism and glutathione metabolism were decreased in CD progressors at age 2.5 ($P \leq 0.05$)
288 (**Figure S5B**). Meanwhile, retinol metabolism, steroid hormone biosynthesis, and
289 glycosaminoglycan degradation pathways were increased in CD progressors at age 5 ($P<0.05$)
290 (**Figure S5C**). We also identified interesting correlations between the most abundant 20 OTUs
291 and most different 20 metabolic pathways (**Table S4).** Among these correlated OTUs, some
292 were identified at the species level at each phase (**Figure S6A-C**)
293 We also used PICRUSt to examine the functional pathways comparing IgA+ to IgA-
294 microbiota (**Figure 3B, S4B, Table S5)**. We identified styrene degradation pathway enriched in
295 IgA+ population at age 1. Further, we identified 31 different functional pathways at age 2.5 and
296 23 functional pathways at age 5 (FDR<0.05) in the healthy subjects. The top enriched pathways
297 in healthy IgA+ population were styrene degradation, chloroalkane and chloroalkene degradation,
298 and toluene degradation at the ages of 2.5 and 5 years old. Analyzing CD samples, we identified
299 7 pathways at age 1, 28 pathways at age 2.5 and 13 pathways at age 5 enriched in IgA+
300 microbiota (FDR<0.05). The most significantly enriched pathways in CD IgA+ microbiota were
301 beta- alanine metabolism, chloroalkane and chloroalkene degradation, and styrene degradation

302 for all three phases. At age 1, CD progressors had more metabolic pathways predicted to be
303 enriched in the IgA+ samples compared to healthy control. For example, pathogenic pathways,
304 bacterial invasion of epithelial cells and beta-alanine metabolism were identified in CD IgA+
305 microbiota population but was absent in control IgA+ at age 1.

306 **Plasma Metabolomics Analysis Reveals an Inflammatory Metabolic Profile for CD**
307 **Progressors**

308 In order to determine the early markers of CD progression in the plasma metabolome and
309 its link to gut microbiota, we applied a targeted plasma metabolomics analysis. We used 10 CD
310 and 9 control plasma samples obtained at age 5. Three subjects in the CD group were diagnosed
311 before age 5. In total, we identified 386 metabolites. Partial least squares-discriminant analysis
312 (PLS-DA) showed a clear separation of the plasma metabolites between CD-progressors and
313 healthy control groups (**Figure 4A).** Volcano plots show the most significantly altered
314 metabolites (**Figure 4B, Table S6**). We identified a clear separation between these groups and
315 19 out of 387 metabolites were significantly different ($p < 0.05$, **Table S6**) between the two
316 groups. The top three most altered metabolites ($p < 0.01$) were TDCA, Glucono-D-lactone and
317 Isobutyryl-L-carnitine. All three top metabolites were increased in CD samples (**Figure 4C**). The
318 most altered metabolite, TDCA, is a conjugated bile acid that was shown to be
319 proinflammatory[39]. TDCA is mainly produced by gut microbes, particularly by Clostridium
320 XIVa and Clostridium XI, with 7-α-dehydroxylation of taurocholic acid and cholic acid[40]. This
321 result is consistent with our microbiota analysis since we identified several Clostridium XIVa
322 OTUs that were significantly more abundant in CD samples, especially at age 5 (**Figure 4D**).
323 The heat map shows 50 of the most altered metabolites and indicates a strong signature in the
324 plasma metabolome in which 19 metabolites were significantly altered ($p<0.05$) (**Figure 4E**).

325 We used Pathway Analysis to determine the functions related to these metabolites (**Figure 4F,**
326 **Table S4**). Indeed, pentose phosphate pathway (PPP), lysine degradation, and glycolipid
327 metabolism were the most significantly altered pathways.

328 **DISCUSSION**

329 Recent studies have demonstrated strong associations between the gut microbiota and the
330 pathogenesis of autoimmune diseases[41]. Studies of the gut microbiome in CD have demonstrated
331 intestinal dysbiosis in CD patients[10 11 16-19 42-48]. However, the majority of these studies were
332 performed using adult samples with diagnosed disease and none of them used a longitudinal
333 approach as this study. Human gut microbiota development is divided into three phases; a
334 developmental phase (months 3-14), a transitional phase (months 15-30), and a stable phase
335 (months 31-46)[27]. Because recent data show that most childhood CD cases will develop by age
336 5years[49], we analyzed samples representing all of these critical phases. We report, for what we
337 believe is the first time, that there are significant differences in microbiome composition and
338 function at each developmental phase in CD progressors.

339 In this study, we first showed that alpha diversity was significantly higher in the CD
340 progressors compared to healthy controls at age 1 (**Figure 1A**). This was an unexpected finding
341 since alpha diversity is reported to be lower in children developing Type 1 Diabetes indicating a
342 difference between two immune-mediated diseases[50]. Consistent with previous reports[43], we
343 identified the proportion of phylum Firmicutes higher in CD progressors at age 1. (**Figure 1C**).
344 Bacterial proteases of species mostly classified within the Firmicutes phylum are involved in
345 gluten metabolism and this might be a link to CD[51]. Additionally, we observed that phylum
346 Verrucomicrobia was only identified in control group at age 1. Akkermansia is the only genus of
347 the Verrucomicrobia phylum identified in gut microbiota[52]. In particular, *Akkermansia*

348 *muciniphila* has a key role in maintaining the integrity and the function of the mucus barrier and
349 it is inversely associated with several diseases[53]. Further, we identified bacterial species highly
350 enriched in CD progressors at age 1 including *Ruminococcus bromii*, *Bifidobacterium dentium*,
351 and *Clostridium XIVa sciendens*. A previous study reported that the abundance of *R. bromii* was
352 greatly reduced in CD patients when gluten free (GF) diet was introduced[54]. Consistent with our
353 findings, other studies have shown that the abundance of *B. dentium* was increased in the CD
354 patients[19]. Clostridium XIVa genus is responsible for producing the proinflammatory metabolite
355 TDCA. As subjects aged, the differences in the gut microbiota was decreased at the phylum level
356 but significantly increased in the OTU level, which is more informative about massive alterations
357 in the microbiota of CD progressors.

358 Intestinal IgA plays a crucial role in defending against pathogenic microorganisms and in
359 maintaining gut microbiome homeostasis[55]. Interestingly, IgA-deficient patients are more
360 susceptible to variety of pathologies, including CD[56]. Planer et al described mucosal IgA
361 responses progression during two postnatal years in healthy US twins[23]. They showed that (i)
362 IgA coated bacteria is affected by age and host genetics and (ii) IgA response is determined by
363 "intrinsic" properties of gut microbiota community members. We used a similar approach to
364 investigate the gut immune development towards healthy and CD states, we initially focused on
365 the development of the IgA response during the gut microbiota maturation. At age 1, we did not
366 identify any OTUs that were significantly different in IgA- and IgA+ samples, suggesting that
367 the intestinal IgA response is not mature enough to target specific bacteria in the gut. However,
368 we identified one OTU in the pathogenic Brucella genus that was highly coated with IgA in CD
369 progressors. Further studies using distinct cohorts will be needed to verify this result, but
370 Brucella species are well characterized pathogens[57] and might be related to increased

371 inflammation and IgA responses. Flow cytometry analyses showed that the IgA response is
372 highly selective and only a small fraction of the gut microbiota is highly coated with IgA in the
373 first five years of life. More importantly, the percentage of IgA+ bacteria was higher in CD
374 progressors compared with healthy controls at ages 1 and age 5. While a reduction of secretory
375 IgA (sIgA) using infant (4-6 months) fecal samples in CD progressors[16] was reported previously,
376 we did not observe such a defect in our cohort but identified a two fold increase in the number of
377 IgA coated bacteria in CD progressors especially at age 5.

378 Our analysis revealed 144 OTUs at age 2.5 and 167 OTUs at age 5 years old that were
379 highly IgA coated in the CD progressors. Among them *Coprococcus comes*, *Bacteroides*
380 *finegoldii* at age 2.5 and *Faecalibacterium prausnitzii* and *Clostridium_XlVa* at age 5 were the
381 main targets of the mucosal immune response in CD progressors. Among these bacteria,
382 *Coprococcus comes* has been recently identified as the main IgA target in the human colon[58].
383 Notably, we identified 72 OTUs at age 2.5 and 45 OTUs at age 5 that are equally abundant in
384 CD progressors and healthy controls, but selectively targeted by IgA in CD progressors. Some of
385 these OTUs are at the species level. For example, *Bacterioides finegoldii*, and *Bacteroides*
386 *vulgatus* at age 2.5 and *Peptostreptococcus stomatis* at age 5 were selectively targeted by IgA in
387 CD progressors but not in controls. *B. vulgatus* was implicated in the development of gut
388 inflammation and a previous report identified this bacterium as enriched in older children with
389 CD[59 60]. In agreement with our results, a pathogenic role for Bacteriodes species is found to be
390 related to the loss of integrity of the intestinal epithelial barrier[60].

391 PICRUSt analyses showed significant differences at all developmental phases, in
392 particular, within the transition period at age 1. Most significant differences were identified in
393 pathways related to bacterial pathogenesis and shaping the composition of microbiota. For

394 example, glutathione metabolism was greatly decreased in CD progressors. Decreased
395 glutathione redox cycle in CD patients is strongly associated with disease development[61]. At age
396 5, PICRUSt predicted that retinol metabolism, steroid hormone biosynthesis, and
397 glycosaminoglycan degradation as over-represented pathways in CD progressors. Retinoic acid
398 is one of the products of retinol metabolism and plays a key role in the intestinal immune
399 response[62]. A previous study showed that retinoic acid mediated inflammatory responses to
400 gluten in CD patients[63]. The increased retinol metabolism and glycosaminoglycan degradation
401 pathways in CD progressors are potentially related to chronic inflammation. Indeed,
402 glycosaminoglycan help to form a protective barrier for the intestinal mucin. The breakdown of
403 glycosaminoglycan is reported to be associated with inflammatory response in intestinal
404 disorders such as IBD[64]. These results suggest that during the developmental phase, the gut
405 microbiota functions in CD progressors were related to shaping the gut microbiota composition.
406 Entering the transition phase, the gut microbiota in CD progressors displayed more
407 proinflammatory and oxidative stress related features. At stable phases, gut microbiota in CD
408 progressors begin to become more involved in functions related to the clinical manifestation of
409 the disease. This longitudinal observation provides insight into the proinflammatory and
410 pathogenic function of gut microbiota in different stages of early CD pathogenesis.
411 Although the hallmark of the CD is intestinal inflammation, the disease affects different
412 tissues. To determine the systemic effects of gut microbiota on different organs, we performed a
413 comparative plasma metabolomics analysis at age 5. In agreement with the gut microbiota
414 analysis, plasma metabolites were significantly altered prior to diagnosis in CD progressors. The
415 top plasma metabolites altered were TDCA and Isobutyryl-L-carnitine (**Figure 4C**) in which
416 both were increased two-fold in CD progressors. TDCA is a conjugated bile acid that is shown to

417 be proinflammatory[39] and is mainly produced by gut bacteria, particularly by Clostridium XIVa
418 and Clostridium XI[40]. This observation suggests that the plasma TDCA detected in our study is
419 secondary to the increased abundance of some Clostridium XIVa species in CD progressors at
420 age 5 (**Fig 4D**). Likewise, Isobutyry-l-carnitine is a member of acylcarnitines. As the byproduct
421 of incomplete beta oxidation, the increased isobutyry-l-carnitine is related to abnormalities in
422 fatty acid metabolism[65] and activates proinflammatory signaling[66]. Pathway analysis for plasma
423 metabolites identified several pathways including pentose phosphate pathway (PPP), lysine
424 degradation, and glycerolipid metabolism. PPP was identified as the most significantly altered
425 pathway and stimulates formation of NADPH for antioxidant, thereby controlling cell
426 inflammation. Thus, plasma metabolites of CD progressors are a component of the inflammatory
427 storm.

428 Currently, the only way to treat CD is strict adherence to a gluten-free (GF) diet, but 20%
429 of patients do not respond to GF diet and continue to have persistent or recurrent symptoms[67].
430 CD permanently reshapes intestinal immunity and alterations in TCRγδ+ intraepithelial
431 lymphocytes in particular may underlie non-responsiveness to the GF diet[68]. Our findings
432 suggest that the inflammatory nature of the CD progressors' gut microbiota is a key component
433 of intestinal inflammation in CD. The proinflammatory factors identified in this study potentially
434 trigger local and systemic inflammation independent of the diet and may explain a failure to
435 respond to GF diet in some patients. Taken together, our findings suggest that the gut microbiota
436 of CD progressors in the first five years of life has profound effects on the inflammatory
437 response and can potentially contribute to onset and progression of CD. These early markers, for
438 example TDCA, have the potential to serve as useful biomarkers for CD diagnosis.

439 Understanding the role of the gut microbiota in chronic inflammation in CD may open novel
440 approaches to understand disease pathogenesis and reveal new preventive and treatment models.
441

442 **Acknowledgements**:

443 We are grateful to all children participating in the ABIS study, and their parents. Thanks
444 also to Ingela Johansson, KEF, Linköping, for her skillful work with the samples, and Åshild
445 Faresjö for register data. The authors also want to thank Hui Pan, Jonathan Dreyfuss (Joslin
446 Diabetes Center Bioinformatic Core) and Sam Minot (Microbiome Research Initiative, Fred
447 Hutchinson Cancer Research Center) for their help with bioinformatic analysis and statistics. The
448 authors would also like to acknowledge Patrick Autissier for the cytometry service (Flow
449 Cytometry Core of Boston College) and Sandra Dedrick (Boston College) for her comments on
450 the text. The ABIS-study has been supported by Swedish Research Council (K2005-72X-11242-
451 11A and K2008-69X-20826-01-4) and the Swedish Child Diabetes Foundation
452 (Barndiabetesfonden), JDRF Wallenberg Foundation (K 98-99D-12813-01A), Medical Research
453 Council of Southeast Sweden (FORSS) and the Swedish Council for Working Life and Social
454 Research (FAS2004–1775) and Östgöta Brandstodsbolag. This work was supported by NIH
455 NIDDK 1K01DK117967-01 and a G. Harold & Leila Y. Mathers Foundation grants to EA.
456

457 **Author contributions:** Q.H and E.A designed research and wrote the paper. Q.H assisted
458 with all experiments, J. L is the Head of the ABIS study and assisted with human fecal sample
459 collection, classification and interpreting the data. Y.Y and N.W.P assisted with IgA-seq
460 experiments and analysis. V.T and M.A.K assisted with plasma metabolomics analysis. J.F.L

assisted with research design and writing. All authors helped the analysis of the data that they contributed to produce and approved the final version of the manuscript.

REFERENCES

1. Lebwohl B, Sanders DS, Green PHR. Coeliac disease. *Lancet* 2018;391(10115):70-81. doi: 10.1016/S0140-6736(17)31796-8 [published Online First: 2017/08/02]
2. Lindfors K, Ciacci C, Kurppa K, et al. Coeliac disease. *Nat Rev Dis Primers* 2019;5(1):3. doi: 10.1038/s41572-018-0054-z [published Online First: 2019/01/12]
3. Tollefsen S, Arentz-Hansen H, Fleckenstein B, et al. HLA-DQ2 and -DQ8 signatures of gluten T cell epitopes in celiac disease. *J Clin Invest* 2006;116(8):2226-36. doi: 10.1172/JCI27620 [published Online First: 2006/08/01]
4. Liu E, Lee HS, Aronsson CA, et al. Risk of pediatric celiac disease according to HLA haplotype and country. *N Engl J Med* 2014;371(1):42-9. doi: 10.1056/NEJMoa1313977 [published Online First: 2014/07/06]
5. Wolters VM, Wijmenga C. Genetic background of celiac disease and its clinical implications. *Am J Gastroenterol* 2008;103(1):190-5. doi: 10.1111/j.1572-0241.2007.01471.x [published Online First: 2008/01/11]
6. King JA, Jeong J, Underwood FE, et al. Incidence of Celiac Disease Is Increasing Over Time: A Systematic Review and Meta-analysis. *Am J Gastroenterol* 2020 doi: 10.14309/ajg.0000000000000523 [published Online First: 2020/02/06]
7. Greco L, Romino R, Coto I, et al. The first large population based twin study of coeliac disease. *Gut* 2002;50(5):624-8. doi: 10.1136/gut.50.5.624 [published Online First: 2002/04/16]
8. Kuja-Halkola R, Lebwohl B, Halfvarson J, et al. Heritability of non-HLA genetics in coeliac disease: a population-based study in 107 000 twins. *Gut* 2016;65(11):1793-98. doi: 10.1136/gutjnl-2016-311713 [published Online First: 2016/05/22]
9. Lionetti E, Catassi C. The Role of Environmental Factors in the Development of Celiac Disease: What Is New? *Diseases* 2015;3(4):282-93. doi: 10.3390/diseases3040282 [published Online First: 2015/10/27]
10. Nistal E, Caminero A, Vivas S, et al. Differences in faecal bacteria populations and faecal bacteria metabolism in healthy adults and celiac disease patients. *Biochimie* 2012;94(8):1724-9. doi: 10.1016/j.biochi.2012.03.025 [published Online First: 2012/05/01]
11. Nistal E, Caminero A, Herran AR, et al. Differences of small intestinal bacteria populations in adults and children with/without celiac disease: effect of age, gluten diet, and disease. *Inflamm Bowel Dis* 2012;18(4):649-56. doi: 10.1002/ibd.21830 [published Online First: 2011/08/10]
12. Sjogren YM, Tomicic S, Lundberg A, et al. Influence of early gut microbiota on the maturation of childhood mucosal and systemic immune responses. *Clin Exp Allergy* 2009;39(12):1842-51. doi: 10.1111/j.1365-2222.2009.03326.x [published Online First: 2009/09/09]
13. Serena G, Yan S, Camhi S, et al. Proinflammatory cytokine interferon-gamma and microbiome-derived metabolites dictate epigenetic switch between forkhead box protein 3 isoforms in coeliac disease. *Clin Exp Immunol* 2017;187(3):490-506. doi: 10.1111/cei.12911 [published Online First: 2016/12/10]
14. De Angelis M, Vannini L, Di Cagno R, et al. Salivary and fecal microbiota and metabolome of celiac children under gluten-free diet. *Int J Food Microbiol* 2016;239:125-32. doi: 10.1016/j.ijfoodmicro.2016.07.025 [published Online First: 2016/07/28]

505 15. Sellitto M, Bai G, Serena G, et al. Proof of concept of microbiome-metabolome analysis and delayed
506 gluten exposure on celiac disease autoimmunity in genetically at-risk infants. *PLoS One*
507 2012;7(3):e33387. doi: 10.1371/journal.pone.0033387 [published Online First: 2012/03/21]
508 16. Olivares M, Walker AW, Capilla A, et al. Gut microbiota trajectory in early life may predict
509 development of celiac disease. *Microbiome* 2018;6(1):36. doi: 10.1186/s40168-018-0415-6
510 [published Online First: 2018/02/21]
511 17. Cheng J, Kalliomaki M, Heilig HG, et al. Duodenal microbiota composition and mucosal homeostasis
512 in pediatric celiac disease. *BMC Gastroenterol* 2013;13:113. doi: 10.1186/1471-230X-13-113
513 [published Online First: 2013/07/13]
514 18. Sanchez E, Donat E, Ribes-Koninckx C, et al. Duodenal-mucosal bacteria associated with celiac
515 disease in children. *Appl Environ Microbiol* 2013;79(18):5472-9. doi: 10.1128/AEM.00869-13
516 [published Online First: 2013/07/10]
517 19. Collado MC, Donat E, Ribes-Koninckx C, et al. Imbalances in faecal and duodenal Bifidobacterium
518 species composition in active and non-active coeliac disease. *BMC Microbiol* 2008;8:232. doi:
519 10.1186/1471-2180-8-232 [published Online First: 2008/12/24]
520 20. Bunker JJ, Bendelac A. IgA Responses to Microbiota. *Immunity* 2018;49(2):211-24. doi:
521 10.1016/j.immuni.2018.08.011 [published Online First: 2018/08/23]
522 21. Wilmore JR, Gaudette BT, Gomez Atria D, et al. Commensal Microbes Induce Serum IgA Responses
523 that Protect against Polymicrobial Sepsis. *Cell Host Microbe* 2018;23(3):302-11 e3. doi:
524 10.1016/j.chom.2018.01.005 [published Online First: 2018/02/27]
525 22. Bunker JJ, Erickson SA, Flynn TM, et al. Natural polyreactive IgA antibodies coat the intestinal
526 microbiota. *Science* 2017;358(6361) doi: 10.1126/science.aan6619 [published Online First:
527 2017/10/04]
528 23. Planer JD, Peng Y, Kau AL, et al. Development of the gut microbiota and mucosal IgA responses in
529 twins and gnotobiotic mice. *Nature* 2016;534(7606):263-6. doi: 10.1038/nature17940
530 [published Online First: 2016/06/10]
531 24. Bunker JJ, Flynn TM, Koval JC, et al. Innate and Adaptive Humoral Responses Coat Distinct
532 Commensal Bacteria with Immunoglobulin A. *Immunity* 2015;43(3):541-53. doi:
533 10.1016/j.immuni.2015.08.007 [published Online First: 2015/09/01]
534 25. Palm NW, de Zoete MR, Cullen TW, et al. Immunoglobulin A coating identifies colitogenic bacteria in
535 inflammatory bowel disease. *Cell* 2014;158(5):1000-10. doi: 10.1016/j.cell.2014.08.006
536 [published Online First: 2014/08/30]
537 26. Viladomiu M, Kivolowitz C, Abdulhamid A, et al. IgA-coated E. coli enriched in Crohn's disease
538 spondyloarthritis promote TH17-dependent inflammation. *Sci Transl Med* 2017;9(376) doi:
539 10.1126/scitranslmed.aaf9655 [published Online First: 2017/02/10]
540 27. Stewart CJ, Ajami NJ, O'Brien JL, et al. Temporal development of the gut microbiome in early
541 childhood from the TEDDY study. *Nature* 2018;562(7728):583-88. doi: 10.1038/s41586-018-
542 0617-x [published Online First: 2018/10/26]
543 28. Baskin AS, Linderman JD, Brychta RJ, et al. Regulation of Human Adipose Tissue Activation,
544 Gallbladder Size, and Bile Acid Metabolism by a beta3-Adrenergic Receptor Agonist. *Diabetes*
545 2018;67(10):2113-25. doi: 10.2337/db18-0462 [published Online First: 2018/07/08]
546 29. Drolet J, Tolstikov V, Williams BA, et al. Integrated Metabolomics Assessment of Human Dried Blood
547 Spots and Urine Strips. *Metabolites* 2017;7(3) doi: 10.3390/metabo7030035 [published Online
548 First: 2017/07/18]
549 30. Tolstikov V, Nikolayev A, Dong S, et al. Metabolomics analysis of metabolic effects of nicotinamide
550 phosphoribosyltransferase (NAMPT) inhibition on human cancer cells. *PLoS One*
551 2014;9(12):e114019. doi: 10.1371/journal.pone.0114019 [published Online First: 2014/12/09]

552 31. Akerman L, Ludvigsson J, Swartling U, et al. Characteristics of the pre-diabetic period in children with
553 high risk of type 1 diabetes recruited from the general Swedish population-The ABIS study.
554 *Diabetes Metab Res Rev* 2017;33(6) doi: 10.1002/dmrr.2900 [published Online First:
555 2017/04/04]
556 32. Ludvigsson JF, Andersson E, Ekbom A, et al. External review and validation of the Swedish national
557 inpatient register. *BMC Public Health* 2011;11:450. doi: 10.1186/1471-2458-11-450 [published
558 Online First: 2011/06/11]
559 33. Koenig JE, Spor A, Scalfone N, et al. Succession of microbial consortia in the developing infant gut
560 microbiome. *Proc Natl Acad Sci U S A* 2011;108 Suppl 1:4578-85. doi: 10.1073/pnas.1000081107
561 [published Online First: 2010/07/30]
562 34. Langille MG, Zaneveld J, Caporaso JG, et al. Predictive functional profiling of microbial communities
563 using 16S rRNA marker gene sequences. *Nat Biotechnol* 2013;31(9):814-21. doi:
564 10.1038/nbt.2676 [published Online First: 2013/08/27]
565 35. Tan FY, Tang CM, Exley RM. Sugar coating: bacterial protein glycosylation and host-microbe
566 interactions. *Trends Biochem Sci* 2015;40(7):342-50. doi: 10.1016/j.tibs.2015.03.016 [published
567 Online First: 2015/05/06]
568 36. Matilla MA, Krell T. The effect of bacterial chemotaxis on host infection and pathogenicity. *FEMS*
569 *Microbiol Rev* 2018;42(1) doi: 10.1093/femsre/fux052 [published Online First: 2017/10/27]
570 37. Aharonowitz Y, Cohen G, Martin JF. Penicillin and cephalosporin biosynthetic genes: structure,
571 organization, regulation, and evolution. *Annu Rev Microbiol* 1992;46:461-95. doi:
572 10.1146/annurev.mi.46.100192.002333 [published Online First: 1992/01/01]
573 38. Moore AM, Ahmadi S, Patel S, et al. Gut resistome development in healthy twin pairs in the first year
574 of life. *Microbiome* 2015;3:27. doi: 10.1186/s40168-015-0090-9 [published Online First:
575 2015/06/27]
576 39. Fujisaka S, Ussar S, Clish C, et al. Antibiotic effects on gut microbiota and metabolism are host
577 dependent. *J Clin Invest* 2016;126(12):4430-43. doi: 10.1172/JCI86674 [published Online First:
578 2016/10/25]
579 40. Ridlon JM, Kang DJ, Hylemon PB. Bile salt biotransformations by human intestinal bacteria. *J Lipid*
580 *Res* 2006;47(2):241-59. doi: 10.1194/jlr.R500013-JLR200 [published Online First: 2005/11/22]
581 41. Marietta E, Horwath I, Balakrishnan B, et al. Role of the intestinal microbiome in autoimmune
582 diseases and its use in treatments. *Cell Immunol* 2019;339:50-58. doi:
583 10.1016/j.cellimm.2018.10.005 [published Online First: 2018/10/28]
584 42. D'Argenio V, Casaburi G, Precone V, et al. Metagenomics Reveals Dysbiosis and a Potentially
585 Pathogenic N. flavescens Strain in Duodenum of Adult Celiac Patients. *Am J Gastroenterol*
586 2016;111(6):879-90. doi: 10.1038/ajg.2016.95 [published Online First: 2016/04/06]
587 43. Wacklin P, Kaukinen K, Tuovinen E, et al. The duodenal microbiota composition of adult celiac
588 disease patients is associated with the clinical manifestation of the disease. *Inflamm Bowel Dis*
589 2013;19(5):934-41. doi: 10.1097/MIB.0b013e31828029a9 [published Online First: 2013/03/13]
590 44. Palma GD, Capilla A, Nova E, et al. Influence of milk-feeding type and genetic risk of developing
591 coeliac disease on intestinal microbiota of infants: the PROFICEL study. *PLoS One*
592 2012;7(2):e30791. doi: 10.1371/journal.pone.0030791 [published Online First: 2012/02/10]
593 45. Collado MC, Donat E, Ribes-Koninckx C, et al. Specific duodenal and faecal bacterial groups
594 associated with paediatric coeliac disease. *J Clin Pathol* 2009;62(3):264-9. doi:
595 10.1136/jcp.2008.061366 [published Online First: 2008/11/11]
596 46. Cenit MC, Olivares M, Codoner-Franch P, et al. Intestinal Microbiota and Celiac Disease: Cause,
597 Consequence or Co-Evolution? *Nutrients* 2015;7(8):6900-23. doi: 10.3390/nu7085314 [published
598 Online First: 2015/08/20]

599 47. Verdu EF, Galipeau HJ, Jabri B. Novel players in coeliac disease pathogenesis: role of the gut
600 microbiota. *Nat Rev Gastroenterol Hepatol* 2015;12(9):497-506. doi: 10.1038/nrgastro.2015.90
601 [published Online First: 2015/06/10]
602 48. Peron G, Nicolai CA, Koons DN. Demographic response to perturbations: the role of compensatory
603 density dependence in a North American duck under variable harvest regulations and changing
604 habitat. *J Anim Ecol* 2012;81(5):960-9. doi: 10.1111/j.1365-2656.2012.01980.x [published Online
605 First: 2012/03/22]
606 49. Lionetti E, Castellaneta S, Francavilla R, et al. Introduction of gluten, HLA status, and the risk of celiac
607 disease in children. *N Engl J Med* 2014;371(14):1295-303. doi: 10.1056/NEJMoa1400697
608 [published Online First: 2014/10/02]
609 50. Kostic AD, Gevers D, Siljander H, et al. The dynamics of the human infant gut microbiome in
610 development and in progression toward type 1 diabetes. *Cell Host Microbe* 2015;17(2):260-73.
611 doi: 10.1016/j.chom.2015.01.001 [published Online First: 2015/02/11]
612 51. Caminero A, Herran AR, Nistal E, et al. Diversity of the cultivable human gut microbiome involved in
613 gluten metabolism: isolation of microorganisms with potential interest for coeliac disease. *FEMS*
614 *Microbiol Ecol* 2014;88(2):309-19. doi: 10.1111/1574-6941.12295 [published Online First:
615 2014/02/07]
616 52. Geerlings SY, Kostopoulos I, de Vos WM, et al. Akkermansia muciniphila in the Human
617 Gastrointestinal Tract: When, Where, and How? *Microorganisms* 2018;6(3) doi:
618 10.3390/microorganisms6030075 [published Online First: 2018/07/26]
619 53. Cani PD, de Vos WM. Next-Generation Beneficial Microbes: The Case of Akkermansia muciniphila.
620 *Front Microbiol* 2017;8:1765. doi: 10.3389/fmicb.2017.01765 [published Online First:
621 2017/10/12]
622 54. Bonder MJ, Tigchelaar EF, Cai X, et al. The influence of a short-term gluten-free diet on the human
623 gut microbiome. *Genome Med* 2016;8(1):45. doi: 10.1186/s13073-016-0295-y [published Online
624 First: 2016/04/23]
625 55. Gutzeit C, Magri G, Cerutti A. Intestinal IgA production and its role in host-microbe interaction.
626 *Immunol Rev* 2014;260(1):76-85. doi: 10.1111/imr.12189 [published Online First: 2014/06/20]
627 56. Chow MA, Lebwohl B, Reilly NR, et al. Immunoglobulin A deficiency in celiac disease. *J Clin*
628 *Gastroenterol* 2012;46(10):850-4. doi: 10.1097/MCG.0b013e31824b2277 [published Online
629 First: 2012/04/06]
630 57. Delpino MV, Fossati CA, Baldi PC. Occurrence and potential diagnostic applications of serological
631 cross-reactivities between Brucella and other alpha-proteobacteria. *Clin Diagn Lab Immunol*
632 2004;11(5):868-73. doi: 10.1128/CDLI.11.5.868-873.2004 [published Online First: 2004/09/11]
633 58. Sterlin D, Fieschi C, Malphettes M, et al. Immune/microbial interface perturbation in human IgA
634 deficiency. *Gut Microbes* 2019;10(3):429-33. doi: 10.1080/19490976.2018.1546520 [published
635 Online First: 2018/11/20]
636 59. Schippa S, Iebba V, Barbato M, et al. A distinctive 'microbial signature' in celiac pediatric patients.
637 *BMC Microbiol* 2010;10:175. doi: 10.1186/1471-2180-10-175 [published Online First:
638 2010/06/23]
639 60. Shiba T, Aiba Y, Ishikawa H, et al. The suppressive effect of bifidobacteria on Bacteroides vulgatus, a
640 putative pathogenic microbe in inflammatory bowel disease. *Microbiol Immunol*
641 2003;47(6):371-8. doi: 10.1111/j.1348-0421.2003.tb03368.x [published Online First:
642 2003/08/09]
643 61. Stojiljkovic V, Pejic S, Kasapovic J, et al. Glutathione redox cycle in small intestinal mucosa and
644 peripheral blood of pediatric celiac disease patients. *An Acad Bras Cienc* 2012;84(1):175-84. doi:
645 10.1590/s0001-37652012000100018 [published Online First: 2012/03/24]

646 62. Mucida D, Park Y, Kim G, et al. Reciprocal TH17 and regulatory T cell differentiation mediated by
647 retinoic acid. *Science* 2007;317(5835):256-60. doi: 10.1126/science.1145697 [published Online
648 First: 2007/06/16]
649 63. DePaolo RW, Abadie V, Tang F, et al. Co-adjuvant effects of retinoic acid and IL-15 induce
650 inflammatory immunity to dietary antigens. *Nature* 2011;471(7337):220-4. doi:
651 10.1038/nature09849 [published Online First: 2011/02/11]
652 64. Winslet MC, Poxon V, Allan A, et al. Mucosal glucosamine synthetase activity in inflammatory bowel
653 disease. *Dig Dis Sci* 1994;39(3):540-4. doi: 10.1007/bf02088339 [published Online First:
654 1994/03/01]
655 65. Vreken P, van Lint AE, Bootsma AH, et al. Quantitative plasma acylcarnitine analysis using
656 electrospray tandem mass spectrometry for the diagnosis of organic acidaemias and fatty acid
657 oxidation defects. *J Inherit Metab Dis* 1999;22(3):302-6. doi: 10.1023/a:1005587617745
658 [published Online First: 1999/06/29]
659 66. Rutkowsky JM, Knotts TA, Ono-Moore KD, et al. Acylcarnitines activate proinflammatory signaling
660 pathways. *Am J Physiol Endocrinol Metab* 2014;306(12):E1378-87. doi:
661 10.1152/ajpendo.00656.2013 [published Online First: 2014/04/25]
662 67. Stasi E, Marafini I, Caruso R, et al. Frequency and Cause of Persistent Symptoms in Celiac Disease
663 Patients on a Long-term Gluten-free Diet. *J Clin Gastroenterol* 2016;50(3):239-43. doi:
664 10.1097/MCG.0000000000000392 [published Online First: 2015/08/19]
665 68. Mayassi T, Ladell K, Gudjonson H, et al. Chronic Inflammation Permanently Reshapes Tissue-Resident
666 Immunity in Celiac Disease. *Cell* 2019;176(5):967-81 e19. doi: 10.1016/j.cell.2018.12.039
667 [published Online First: 2019/02/12]

668

669

670

Figure Legends

Figure 1. The Gut Microbiota of Children Developing CD and Healthy Controls were Different in the First 5 Years of Life.

a. Box plots showing the comparison between CD progressors and healthy controls : the alpha diversity measured by observed (upper panel) and the beta diversity measured by Bray–Curtis dissimilarity (lower panel).

b. Non-metric MultiDimentional Scaling (NMDS) analysis ordination of sample similarity/dissimilarity between CD progressors and healthy controls at age 1, 2.5, and 5 years old. Each circle represents an individual sample.

c. Average relative abundance of bacterial phylum (upper panel) or genera (lower panel) of greater than 1% abundance (proportion) between the gut microbiota of CD progressors and healthy controls at age 1, 2.5, and 5 years old (taxa average relative abundance>1%).

d. Heat map showing the relative abundance of the top OTUs significantly different between CD progressors and healthy controls. Each column represents an individual participant and each raw represents an OTU.

e. Empirical Bayes quasi-likelihood F-tests analysis for the comparisons of gut microbiota OTUs between CD progressors and healthy controls age 1, 2.5, and 5 years old. Frequency: number of OTUs. **a-e**: age 1: n=5/group; age 2.5: n=16/group; age 5: n=13/group.

Figure 2. IgA-based Sorting and 16S sequencing revealed that Gut Microbiota were Differentially Coated by IgA.

24 **a.** Principle component analysis (PCA) of sample similarity/dissimilarity between IgA+
25 and IgA- microbiota in healthy control (left) or CD progressors (right).
26 **b.** Flow cytometry results for IgA coating of fecal bacteria from CD progressors and
27 healthy controls at at age 1, 2.5 and 5. Indicated are mean±SEM. *P<0.05 (Two way
28 ANOVA).
29 **c.** Box plots showing the comparison between CD progressors and healthy controls : the
30 alpha diversity measured by observed for IgA+/IgA- microbiota (upper panel), and the
31 beta diversity measured by Bray–Curtis dissimilarity for IgA+/IgA- microbiota (lower
32 panel) at ages 1, 2.5 , and 5 years old. *P<0.05.
33 **d.** NMDS analysis ordination of IgA+/IgA- microbiota similarity/dissimilarity between
34 CD progressors and healthy controls at ages 1, 2.5, and 5 years old. Each circle represents
35 an individual sample.
36 **e.** Average relative abundance of IgA+/IgA- bacterial phylum (left) or genera (right) of
37 greater than 1% abundance (proportion) between the gut microbiota of CD progressors
38 and healthy controls at ages 1, 2.5, and 5 years old (taxa average relative
39 abundance>1%).
40 **f** Heat map showing the relative abundance of the top OTUs significantly different
41 between IgA+/IgA- CD progressors and healthy controls (OTUs=51 based on p-value).
42 Each column represents an individual participant and each raw represents an OTU.
43 **a-f**: age 1: n=5/group; age 2.5: n=16/group; age 5: n=13/group.
44
45 **Figure 3. The Enriched Microbial Pathways Altered in CD Progressors' Gut**
46 **Microbiota are Significantly Different from Healthy Controls**

a. Heat map of PICRUSt predicted metabolic pathways of CD progressors and healthy controls. Each column represents an individual participant and each raw represents a predicted microbial functional pathway. Color code is shown on the figure.

b. Heat map of predicted metabolic pathways of CD progressors and healthy control obtained from PICRUSt analysis. Each column represents an individual participant and each raw represents a predicted microbial functional pathway. **a-b**: ages 1year old : n=5/group); age 2.5 years old: n=16/group; age 5 years old: n=13/group.

Figure 4. CD Progressors Have a Distinct Plasma Metabolic Profile at Age 5

a. Partial Least Square-Discrimination Analysis (PLS-DA) of plasma metabolites for CD progressors (n=10) and controls (n=9).

b. Volcano plot of plasma metabolites with fold change threshold ($|\log_2 (FC)|>1.2$) and t-tests threshold ($-\log_{10}(p)>0.1$). The red dots represent metabolites above the threshold. Fold changes are $\log_2$ transformed and p values are $\log_{10}$ transformed.

c. Box plots showing three most significantly increased metabolites in CD plasma.

d. Box plots showing the representative Clostridium XIVa bacteria abundance between CD progressors and healthy controls.

e. Heatmap showing 50 of the most altered metabolites

f. The Pathway Analysis (combined results from powerful pathway enrichment analysis with pathway topology analysis) identify the most altered metabolic pathways between CD progressors and healthy controls. Pathway impact value (x) is calculated from pathway topology analysis. p is the original p value calculated from the enrichment analysis and depicted on a logarithmic scale.

Figure S1. Box plots of representative bacteria that are differentially distributed in CD progressors and healthy controls at a. Age 1 year old. b. Age 2.5 years old and c. age 5 years old.

Figure S2. Flow Cytometry Analysis of IgA Coating of Fecal Bacterial from CD Progressors and Health

a. Schematic overview of IgA-based fecal bacteria separation combined with 16S rRNA gene sequencing (IgA-seq) for stool samples from CD progressors and healthy controls. MACS: Magnetic-activated cell separation

b. PCA plot of separation for all samples including presorting and postsorting (after IgA-seq) from both CD progressors and healthy controls in all age groups. Round dot: IgA coated and triangle: IgA uncoated. (Age 1: n=5; Age 2.5: n=16; Age 5: n=13).

c. Representative flow cytometry results of human fecal bacteria from healthy control and CD progressors with anti-IgA. (Age 1: n=5; Age 2.5: n=16; Age 5: n=13).

d. Flow cytometry results for IgA coating of fecal bacteria from CD progressors and healthy controls at at age 1, 2.5 and 5 excluding results from CD progressors with CD diagnosis before 5. (Age 1: n=5; Age 2.5: n=15 and 16; Age 5 n=9 and 13). Indcated are mean±SEM. *P<0.05 (Two way ANOVA).

Figure S3. IgA- Sequencing for CD Progressors and Healthy Controls.

a. Empirical Bayes quasi-likelihood F-tests analysis for the comparisons of IgA coated and non coated gut microbiota OTUs in healthy controls (upper row) and CD progressors (lower row) at ages 1, 2.5, and 5. Frequency: number of OTUs.

93 **b.** Empirical Bayes quasi-likelihood F-tests analysis for the comparisons of IgA coated or
94 non coated gut microbiota OTUs between CD progressors and healthy controls (upper
95 row: age 2.5 years old; lower row: age 5 years old).
96 **c.** Box plots showing representative bacteria in which abundances were similar in the gut
97 microbiota (presorting samples) but differently targeted by IgA (upper row: age 2.5 years
98 old; lower row: age 5 years old).
99 **a-c**: Age 1 : n=5/group; Age 2.5: n=16/group; age 5:n=13/group.
100
101 **Figure S4. PICRUSt predicted genes enriched in pre-IgA seq and post-IgA seq**
102 **samples.**
103 **a.** Heat map of PICRUSt predicted microbial genes enriched in CD progressors or
104 healthy controls Each column represents an individual participants and each raw
105 represents a predicted microbial gene. COLOR CODE?
106 **b.** Heat map of PICRUSt predicted microbial genes enriched in CD progressors or
107 healthy controls after IgA sequencing. Each column represents an individual participants
108 and each raw represents a predicted microbial gene.
109 **a-b**: age 1: n=5/group; age 2.5: n=16/group; age 5: n=13/group.
110
111 **Figure S5. Box plots of representative metabolic pathways that are affected in CD**
112 **progressors. A.** Age 1**, B.** Age 2.5 , **C.** Age 5
113

114 **Figure S6. Heat map showing the correlation of most significantly altered top 20**

115 **OTUs and most significantly altered 20 metabolic pathways at a.** Age 1, **b.** Age 2.5 **,**

116 **c.** Age 5**.** Color key represents the Pearson correlation coeffient.

117

Figure 1 PRESORTING FIGURES: The gut microbiota of children developing celiac disease and healthy controls are different in the first 5 years of life

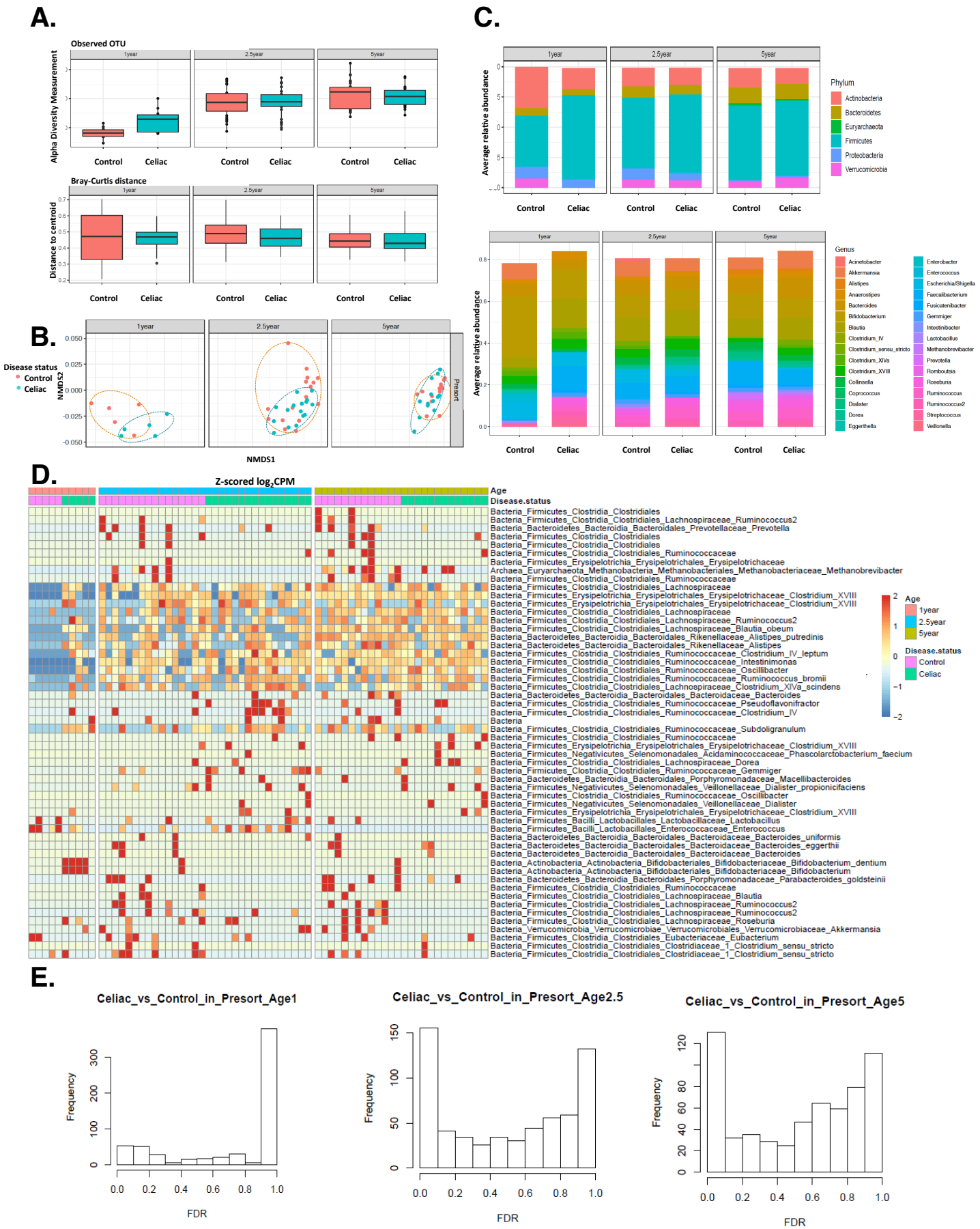


Figure 2 POSTSORTING FIGURES: IgA Seq analysis differences

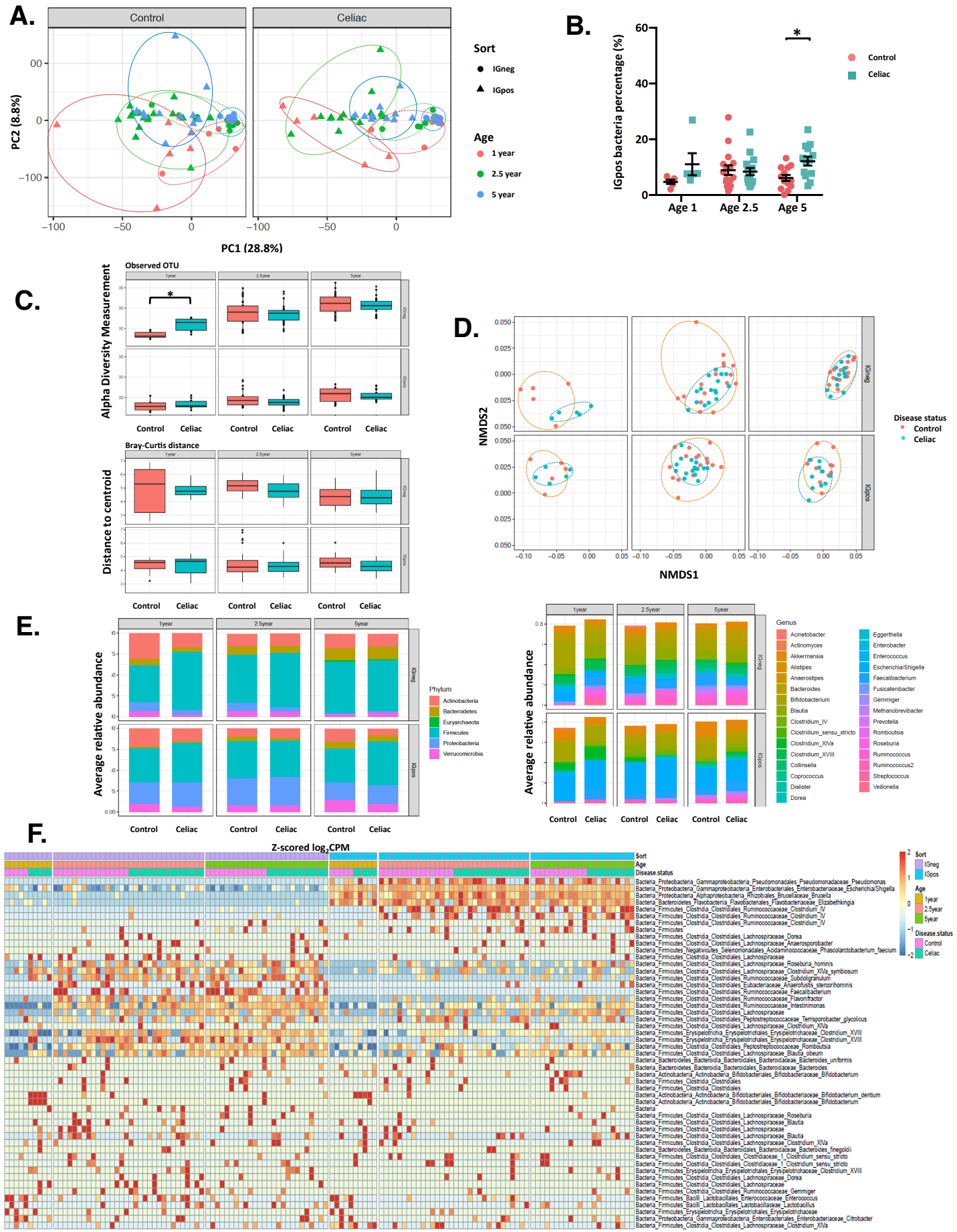


Figure 3 PICRUST

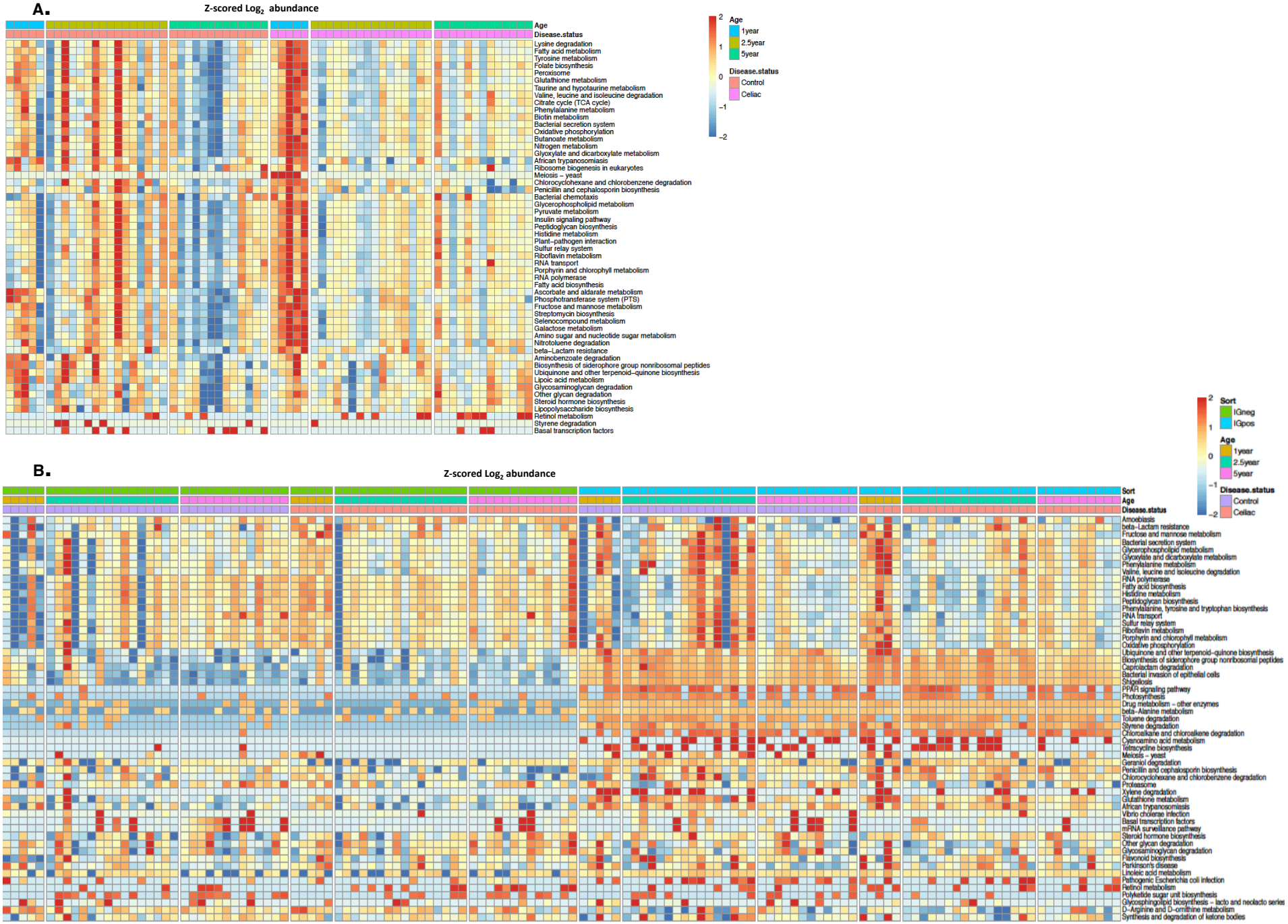
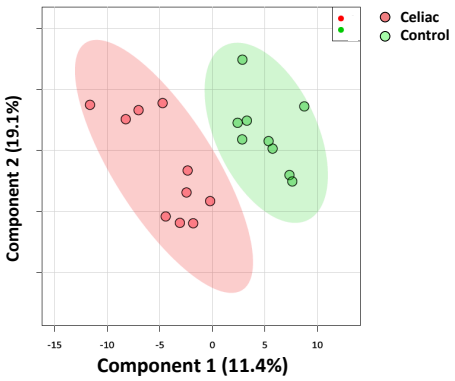
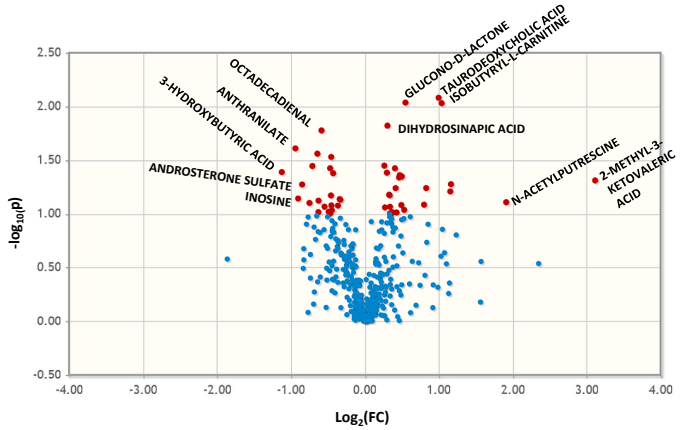


Figure 4 Plasma Metabolomics

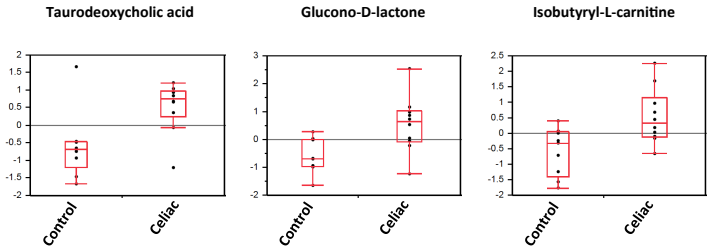
A.



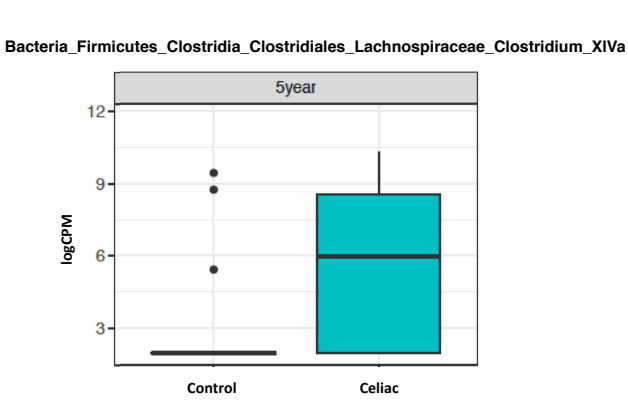
B.



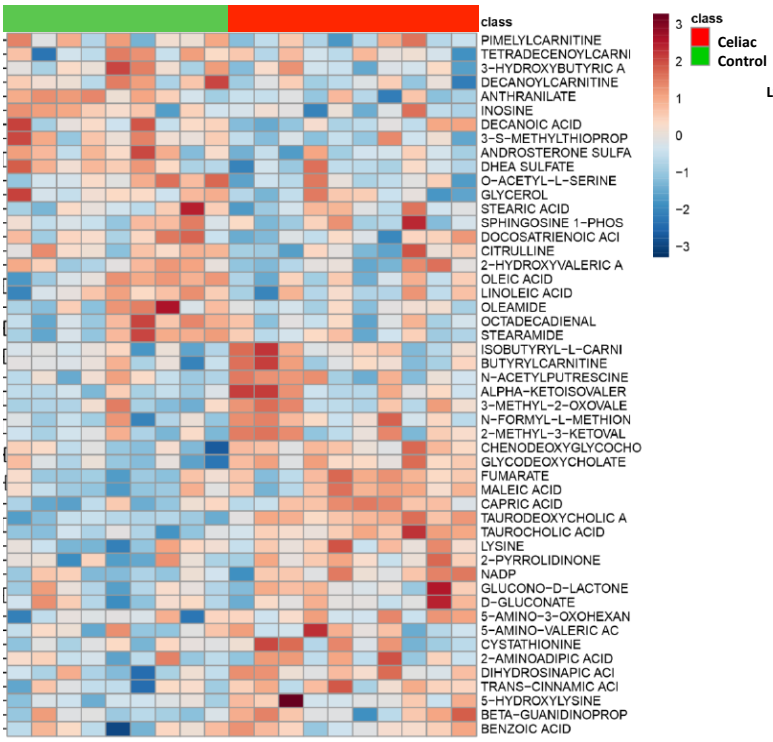
C.



D.



E.



F.

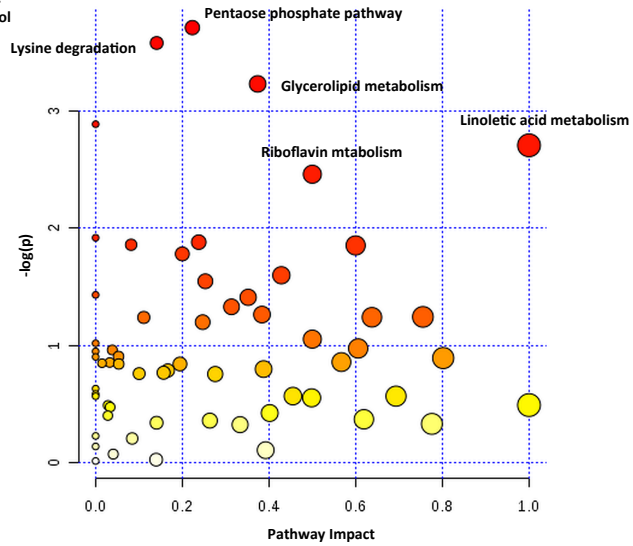
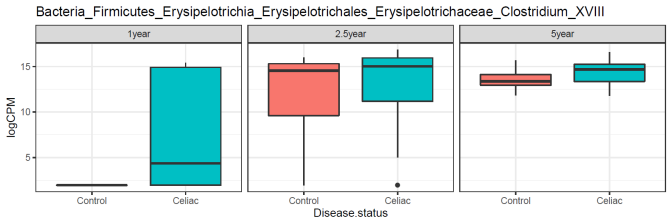
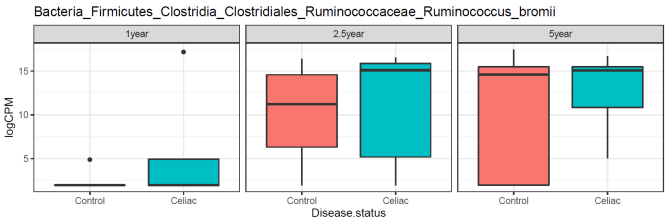
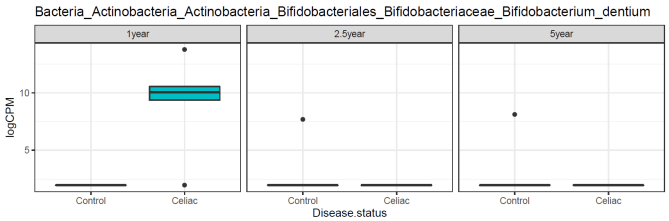


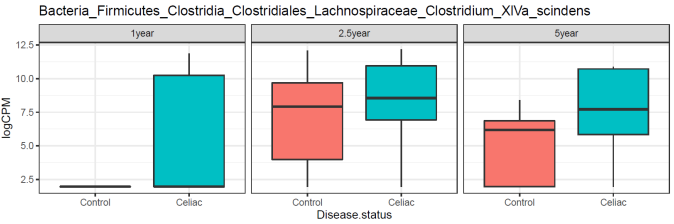
Figure S1. Pre-sorting. Most Significant Differences for PRESORTING

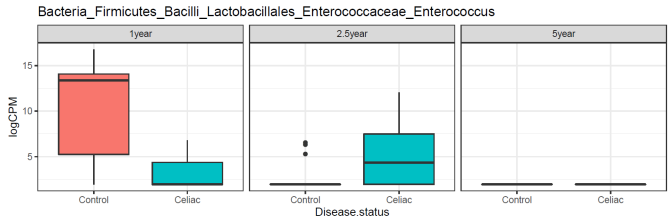
A. Age 1



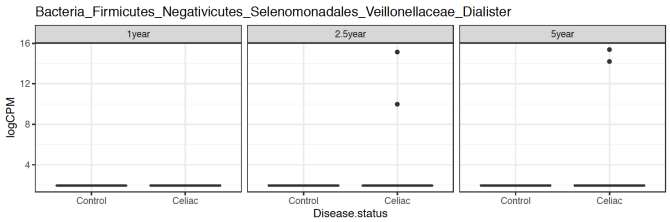


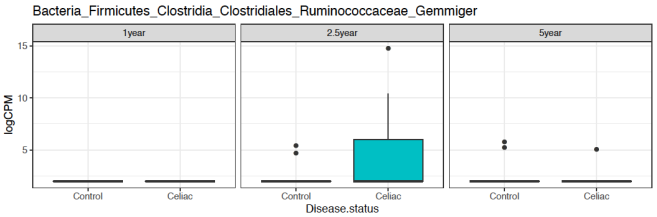


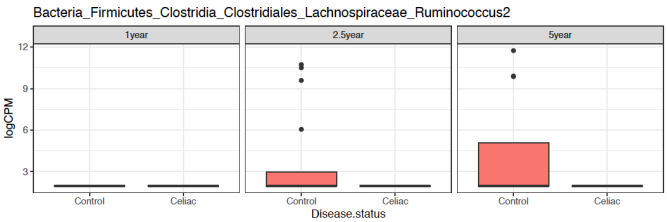


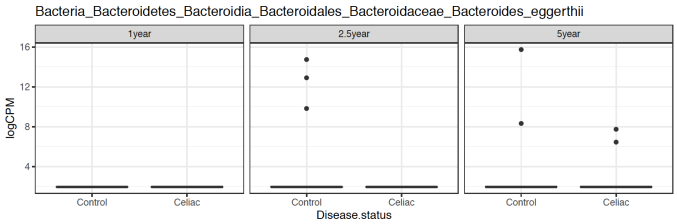


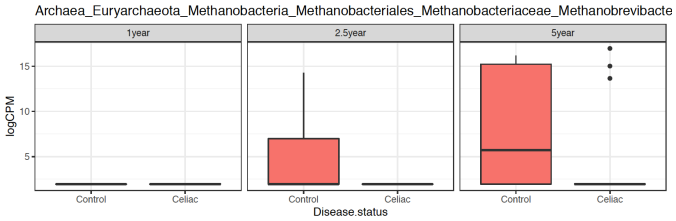
B. Age 2.5



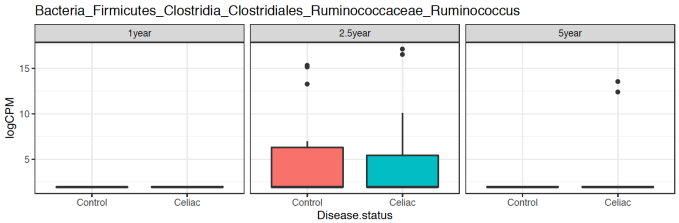


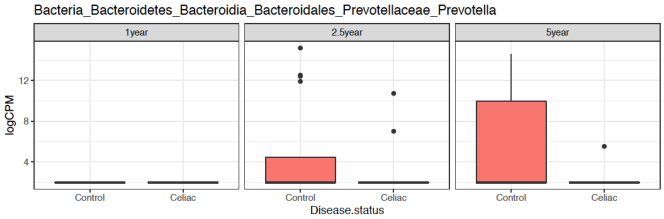


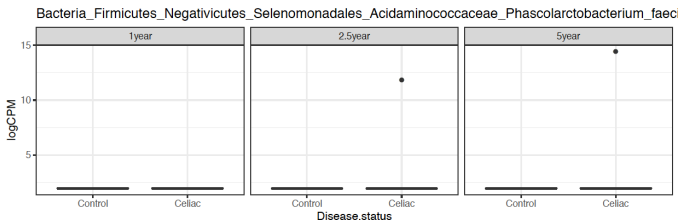




C. Age 5







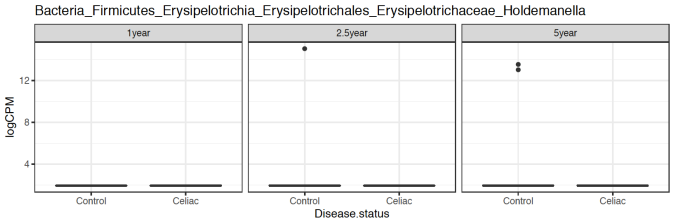


Figure S2. Post-sorting. IgA+ bacteria percentage

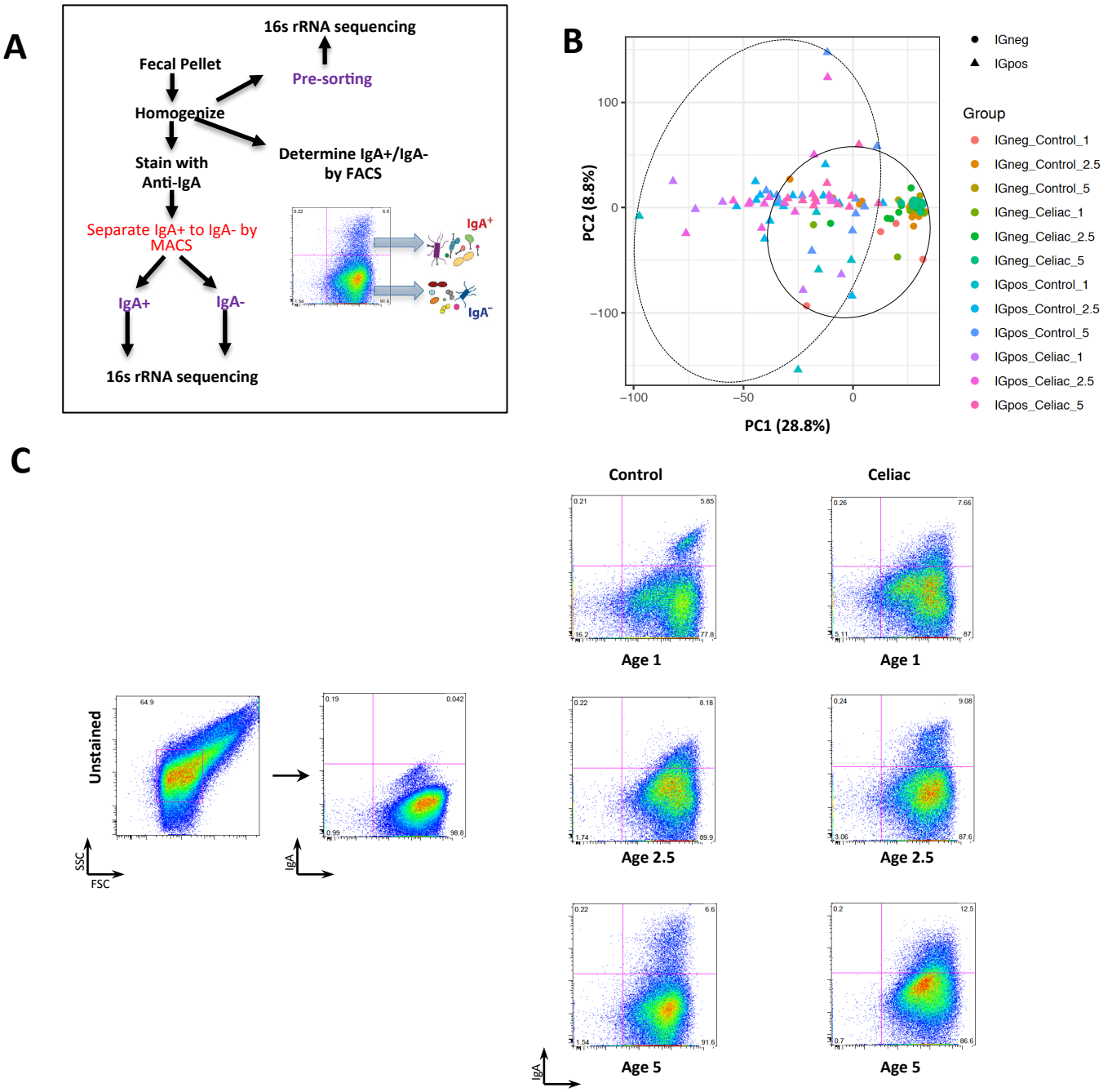
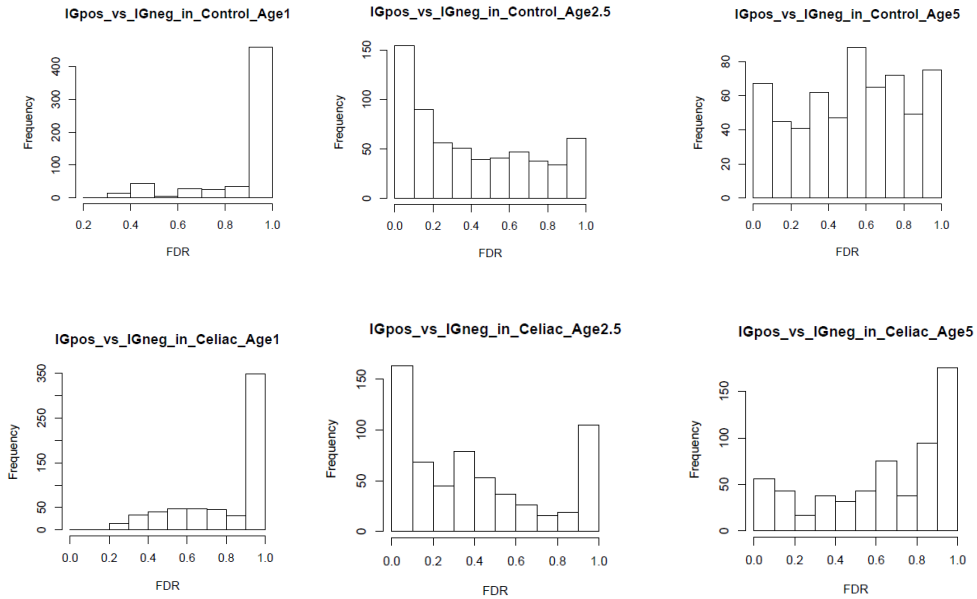
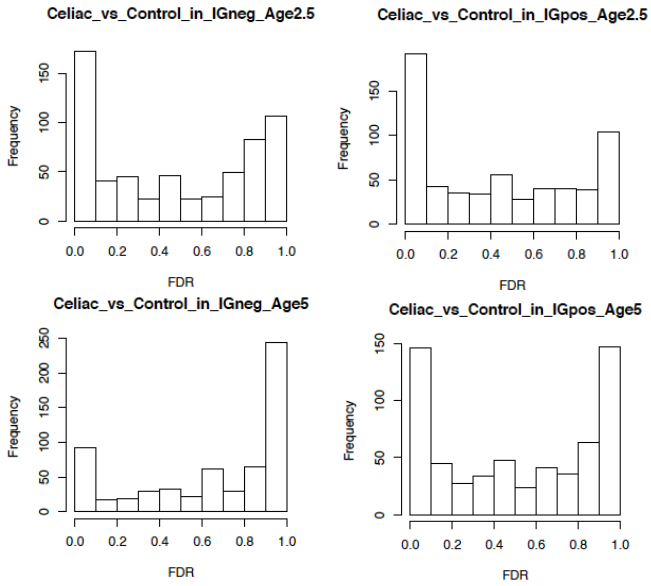


Figure S3. Post-sorting. Most Significant Differences for post-sorting

A



B



C

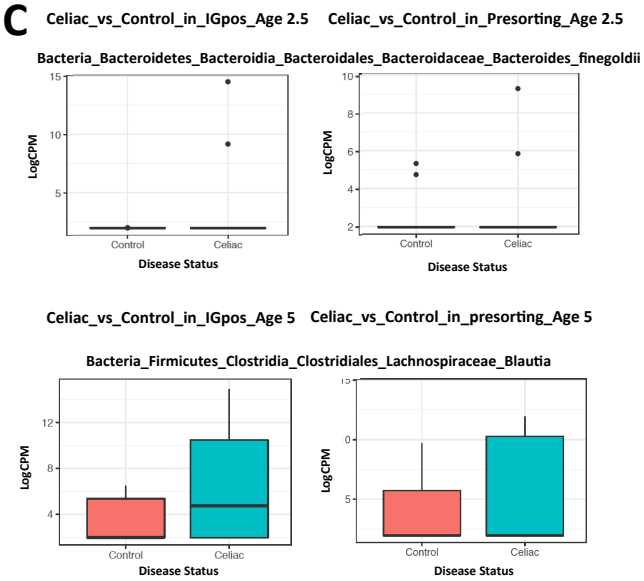


Figure S4. PICRUST genes heat map

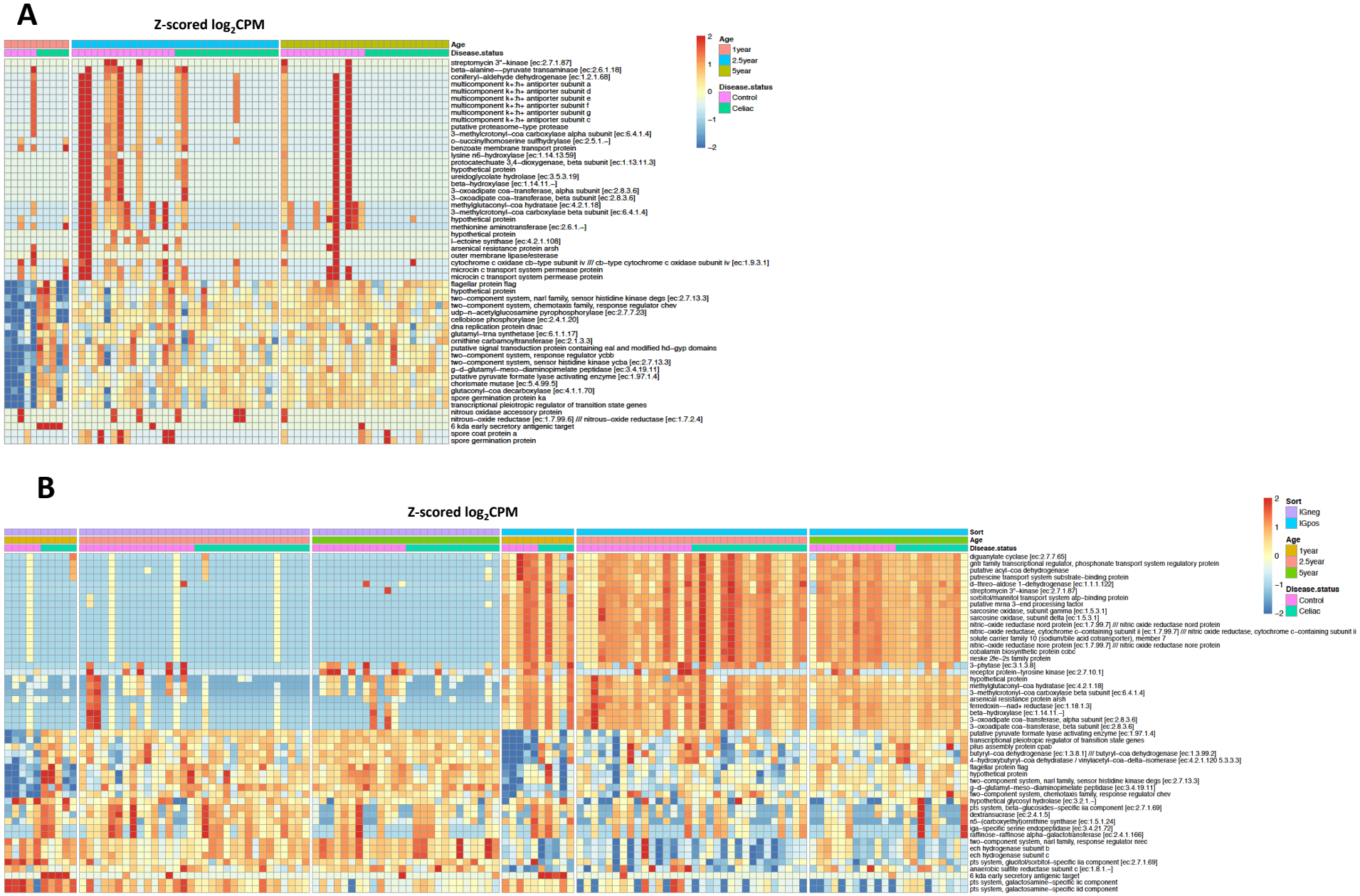


Figure S5. PICRUSt

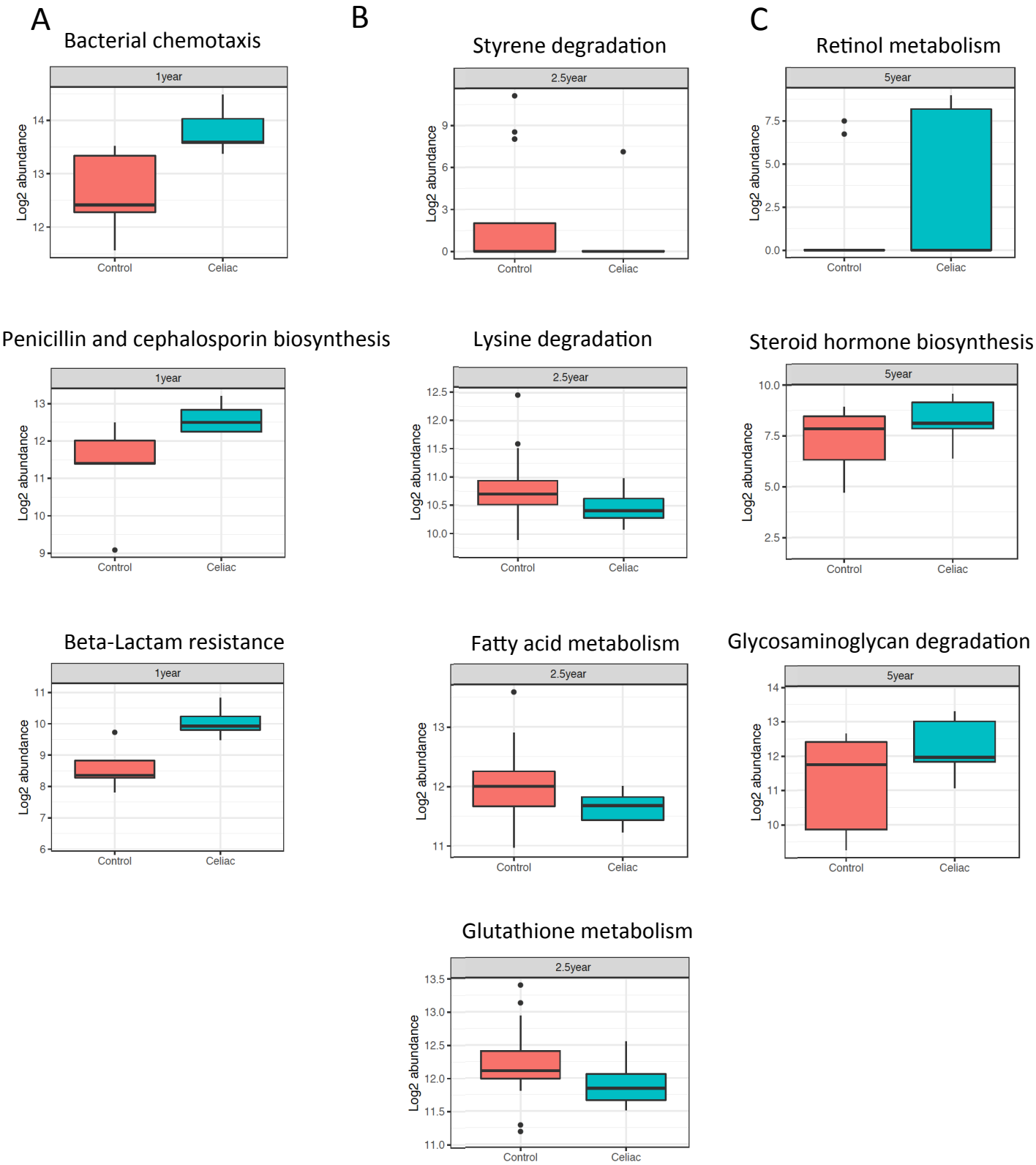
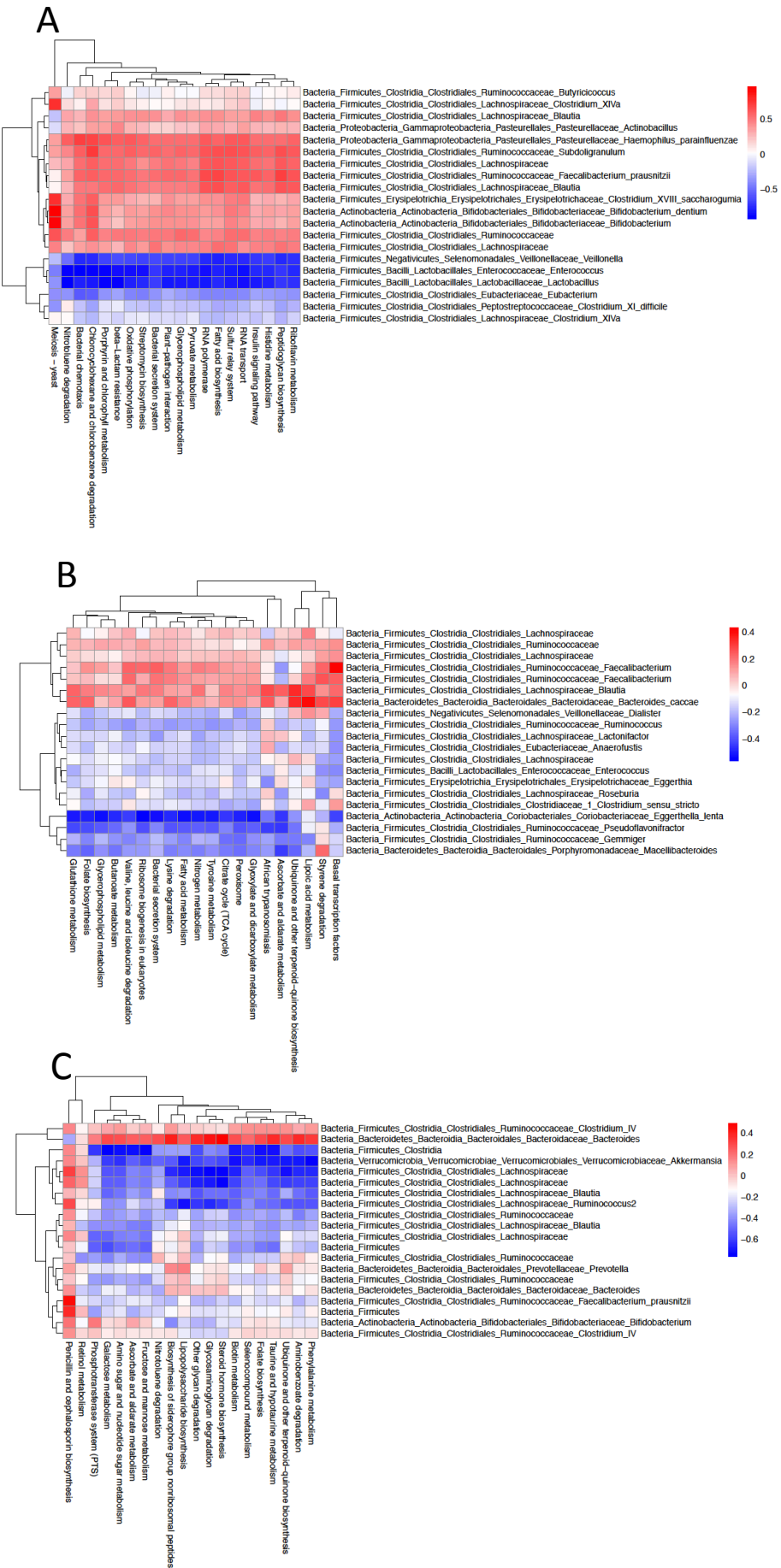



Figure S6. Pearson correlation analysis