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. CD is associated with HLA-DQ2 and HLA-DQ8 haplotypes, however, genetics alone cannot explain the increasing incidence rates. The main goal 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 focusing on three developmental phases of the gut microbiota (ages 1, 2.5 and 5 years). The fecal samples were obtained from 16 children who developed CD and 16 matched controls. We used 16S sequencing combined with functional analysis, flow cytometry, immunoglobulin A (IgA) 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 (CDP, children who developed CD during or after their gut microbiota were sampled) in each developmental phase. Pathogenesis and inflammation-related microbial pathways were enriched in CDP. Moreover, they had significantly more IgA coated bacteria and the IgA targets were significantly different compared to controls. Proinflammatory and pathogenesis-related metabolic pathways were enriched in CDP. Further, we identified inflammatory metabolites, particularly microbiota-derived taurodeoxycholic acid (TDCA) as increased in CDP.

Conclusion: Our study defines an inflammatory gut microbiota for the CDP including its composition, function, IgA response and related plasma metabolites. The inflammatory nature of CD gut microbiota during development is potentially related to the onset of the disease. Targeting inflammatory bacteria in this critical window could affect the pathogenesis and prognosis of CD.

Significance of this study

What is already known on this subject?

- Celiac Disease (CD) is a gluten induced immune-mediated disease in genetically predisposed individuals.
- CD incidence is increasing worldwide which genetics alone cannot explain. Previous studies have shown that the gut microbiota of CD patients differ from that of healthy populations. However, the role of the microbiome in CD pathogenesis and its role in chronic inflammation is yet be established.

What are the new findings?

- In a prospective longitudinal study in children using samples representing all three phases of gut microbiota development (ages 1, 2.5 and 5), we identified significant differences in the composition and function of gut microbiota at each phase. Pathogenesis and inflammation-related functions are enriched in the gut microbiome of CD progressors.
- We applied IgA-sequencing to identify inflammatory bacteria in both healthy subjects and CD progressors. Flow Cytometry analysis identified more IgA coated bacteria at ages 1 and 5 in CD progressors, indicating an early inflammatory response. CD bacterial IgA targets also differed significantly from healthy controls.
- We analyzed plasma metabolites obtained at age 5. The CD plasma metabolome was significantly different from healthy controls. Particularly, proinflammatory plasma metabolites, including microbiota-derived taurodeoxycholic acid (TDCA) and isobutyryl-L-carnitine, were increased two-fold in CD progressors.

How might it impact clinical practice in the foreseeable future?

- Our results establish a link between gut microbiota composition and chronic inflammation in CD during child development. The highly IgA-coated bacteria identified in IgA sequencing and inflammatory bacteria potentially contribute to CD pathogenesis. Targeting these bacteria in the early stages of CD development could be a preventative tool.
- TDCA is a microbiota-derived proinflammatory metabolite increased two-fold in CD progressors. Increased TDCA levels may be used as a predictive/diagnostic tool in genetically predisposed subjects. Moreover, targeting TDCA-producing bacteria (e.g., Clostridium XIVa species) could potentially help to control the intestinal inflammation in CD.
- Developing anti-inflammatory probiotics/prebiotics might be viable therapeutics for altering microbiota composition in children genetically predisposed for CD. These microbes/compounds may also complement a gluten-free diet in patients that continue to experience persistent CD symptoms.

INTRODUCTION

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

This evidence indicates that CD is a multifactorial disease and environmental factors play a role in CD onset. Various environmental factors are implicated in CD development[8], but the roles of these environmental factors in CD progression remain largely unknown. Gut microbiome studies observe an altered microbial[9] and metabolite composition[10 11] in both infant and adult CD patients but have not identified any causal link[12 13]. Immunoglobulin A (IgA) is the most abundant antibody isotype at mucosal surfaces and is a major mediator of intestinal immunity in humans[14]. IgA-sequencing (IgA-seq) combines bacterial flow cytometry with high-throughput sequencing to identify distinct subsets of highly IgA coated (IgA+) and non-coated microbiota (IgA-)[15-17]. It was previously shown in a mice model that IgA+ microbiota could induce more severe colitis than IgA- microbiota[18]. Similarly, IgA-seq identified *Escherichia coli*

as an inflammatory bacterium enriched in Crohn's disease-patients with spondyloarthritis[19]. Therefore, we hypothesized that the pathogenic bacteria and some immunoregulatory commensals involved in CD onset would be highly coated with IgA (IgA+) while most of the commensals would not be coated (IgA-). We also hypothesized that the immune response would have specific targets in CD progressors that differ from the control targets.

In this study, we assessed the composition and function of the gut microbiota in a prospective, longitudinal cohort of 32 children matched for human leukocyte antigen (HLA) genotype and breastfeeding duration (n=16/group). We focused on samples obtained at ages 1, 2.5 and 5 because these samples represent the three stages of gut microbiota development[20]. We then identified the functional pathways enriched in CD progressors, determined the targets of the IgA in the gut microbiota using IgA-seq. Lastly, we compared the plasma metabolome and identified significant differences. Our findings demonstrate that children who go on to develop CD have significant alterations in their gut microbiome years before diagnosis. CD-associated gut microbiota are enriched in inflammatory- and pathogenicity-related bacteria, as well as microbial functions and metabolites that potentially contribute to chronic inflammation in CD.

EXPERIMENTAL PROCEDURE:

Human Fecal Samples

The fecal samples were obtained from subjects in the All Babies in Southeast Sweden (ABIS) cohort. ABIS study was ethically approved by the Research Ethics Committees of the Faculty of Health Science at Linköping University, Sweden (Ref. 1997/96287 and 2003/03-092) and the Medical Faculty of Lund University, Sweden (Dnr 99227, Dnr 99321). All children born in southeast Sweden between 1st October 1997 and 1st October 1999 were recruited. Informed

consent from the parents was obtained. Fresh fecal samples were collected either at home or at the clinic. Samples collected at home were stored at -20 °C with freeze clamps, mailed to the WellBaby Clinic and stored dry at −80 °C. The questionnaire was completed by the parents to collect participants' health information including, but not limited to, breast feeding duration, antibiotic use, gluten exposure time, and more. In total 68 fecal samples were collected for the analysis (10 at 1 year old, 32 at 2.5 years old and 26 at 5 years old).

IgA+ and IgA- Bacteria Separation and Fecal IgA Flow Cytometry

IgA+ and IgA- bacteria separation was performed as previously describe[18]. Briefly, human fecal bacteria were stained with Anti-human IgA PE (clone IS11-8E10 Miltenyi Biotec) followed by Magnetic Activated Cell Separation (MACS) or flow cytometric analysis. Additional details are provided in the Supplemental Methods.

Flow Cytometric analysis.

Bacterial cells were isolated from fecal samples as described in the Supplemental Methods and analyzed by flow cytometry using a BD FACSAria™ IIIu cell sorter (Becton-Dickinson) as previously described[15].

16S rRNA Gene Sequencing and Statistical Analyses

16S rRNA sequencing of the V4 region sequencing for all bacteria samples were performed on the Miseq platform with barcoded primers. Microbial diversity and statistical analyses were performed with *data 2*, *phyloseq 2*, *vegan*, *edge,* and *PICRUSt2*. Additional details are provided in the Supplemental Methods.

165

166 **Plasma Metabolomics and Metabolite pathway Analyses**

167 Preparation of plasma samples for metabolomics analysis and subsequent plasma
168 metabolomics and metabolite pathway analyses were performed as previously described[21].
169 Additional details are provided in the Supplemental Methods.

170

171

172 **RESULTS**

173 **Study Cohort**

174 All Babies in Southeast Sweden (ABIS) is a prospective population-based study that
175 established a large biobank of biological specimens obtained longitudinally at birth and ages 1,
176 2.5, and 5. To determine the role of gut microbiota in CD pathogenesis, we used ABIS samples
177 selecting a sub cohort of 32 individuals born 1997-1999 in Sweden. We chose 16 subjects who
178 developed CD but were not diagnosed with any other autoimmune disease as of December 2017.
179 We matched this group with 16 healthy controls based on their HLA-risk class distribution and
180 breastfeeding duration (**Table S1**). The diagnosis of CD for 11 individuals occurred after the age
181 of 5, one subject was diagnosed at age 1.8 while the other four were diagnosed between the ages
182 of 2.5 and 5. Swedish National Patient Register[22] was used for verifying diagnosis of CD
183 according to international classification of disease (ICD) code-10 K90.0. In total, we used 68
184 longitudinal stool samples (**Table S1**). Although we did not match subjects for other parameters,
185 timing of gluten exposure, delivery method, breastfeeding duration, family history of CD,
186 infections and use of antibiotics were comparable between groups (**Table S1**).

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

By sequencing the V4 region of the 16S rRNA gene[18], we identified 661 operational taxonomic units (OTUs) (**Table S2**). Consistent with previous studies[23], gut microbiome alpha diversity increased until age 2.5 and remained stable up to age 5 in both groups. However, alpha diversity was significantly higher for the CD progressors at age 1, indicating a more diverse microbiota composition (**Figure 1A**). Beta diversity was comparable between CD and healthy controls in each phase (**Figure 1A**). Non-metric multidimensional scaling (NMDS) plots show a trend of separation of gut microbiome composition between CD and healthy control individuals at ages 1 and 2.5 (**Figure 1B**).

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

When we analyzed the differences at the level of OTUs, we identified 10 OTUs that are significantly (FDR≤0.05, $p<0.05$) different at age 1. Clostridium XVIII, *Ruminococcus bromii*, *Bifidobacterium dentium*, and *Clostridium XIVa sciendens* were highly enriched in CD progressors while Enterococcus was highly enriched in control samples (**Figure S1A, Figure 1D**). We identified more significant differences in OTU level at ages 2.5 and 5. 133 OTUs were different at age 2.5 and 112 OTUs were different at age 5 (**Figure 1E**). The most significantly

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

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

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

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

A Specific IgA Response to Bacteria Develops After Age 1

Because there are very few reports on the IgA response in early human gut microbiota development[17], we first focused on the results obtained from the healthy children. We did not observe any difference between IgA+ and IgA- samples in the control group at age 1 (**Figure S3A**). This result suggests that the IgA response does not target specific bacteria in this early stage of development. Consistent with the flow cytometry analysis, we identified 113 OTUs at age 2.5 and 43 OTUs at age 5 that were significantly different between IgA+ and IgA- samples in healthy controls (**Figure S3A, Table S2**). The top targets of IgA in healthy subjects were OTUs *Clostridium IV*, *Bifidobacterium* and *Bacteroides clarus* at age 2.5 and *Clostridium IV*, *Gemmiger* and *Elizabethkingia* at age 5.

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

IgA Response Targets Are Comprised of Different Bacteria in CD progressors

Consistent with the presorting data, the alpha diversity increased at age 2.5 and remained stable in both control and CD IgA- groups (**Figure 2C)**. Likewise, beta diversity was comparable 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[24] is designed to estimate the functional metagenome of gut bacteria using

16S rRNA data (**Figure S4A**). Combining PICRUSt with Kyoto Encyclopedia of Genes and Genomes (KEGG) metabolic pathway analysis, we identified 71 different metabolic pathways that differed between CD and control samples at age 1 (FDR<0.05, **Figure 3A**). Among these pathways, N-glycan biosynthesis, penicillin and cephalosporin biosynthesis, beta-Lactam resistance and bacterial chemotaxis pathways were the top pathways enriched in CD progressors (**Figure S5A**). Interestingly, most of these pathways are involved in bacterial pathogenesis[25 26] or shaping the composition of microbiota[27 28]. We also identified 9 pathways showing strong trends of difference at ages 2.5 and 5. For example, styrene degradation, lysine degradation, fatty acid metabolism and glutathione metabolism were decreased in CD progressors at age 2.5 ($P \leq 0.05$) (**Figure S5B**). Meanwhile, retinol metabolism, steroid hormone biosynthesis, and glycosaminoglycan degradation pathways were increased in CD progressors at age 5 ($P < 0.05$) (**Figure S5C**). We also identified interesting correlations between the most abundant 20 OTUs and most different 20 metabolic pathways (**Table S4).** Among these correlated OTUs, some were identified at the species level at each phase (**Figure S6A-C**)

We also used PICRUSt to examine the functional pathways comparing IgA+ to IgA- microbiota (**Figure 3B, S4B, Table S5)**. We identified styrene degradation pathway enriched in IgA+ population at age 1. Further, we identified 31 different functional pathways at age 2.5 and 23 functional pathways at age 5 (FDR<0.05) in the healthy subjects. The top enriched pathways in healthy IgA+ population were styrene degradation, chloroalkane and chloroalkene degradation, and toluene degradation at the ages of 2.5 and 5 years old. Analyzing CD samples, we identified 7 pathways at age 1, 28 pathways at age 2.5 and 13 pathways at age 5 enriched in IgA+ microbiota (FDR<0.05). The most significantly enriched pathways in CD IgA+ microbiota were beta- alanine metabolism, chloroalkane and chloroalkene degradation, and styrene degradation

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

Plasma Metabolomics Analysis Reveals an Inflammatory Metabolic Profile for CD Progressors

In order to determine the early markers of CD progression in the plasma metabolome and its link to gut microbiota, we applied a targeted plasma metabolomics analysis. We used 10 CD and 9 control plasma samples obtained at age 5. Three subjects in the CD group were diagnosed before age 5. In total, we identified 386 metabolites. Partial least squares-discriminant analysis (PLS-DA) showed a clear separation of the plasma metabolites between CD-progressors and healthy control groups (**Figure 4A).** Volcano plots show the most significantly altered metabolites (**Figure 4B, Table S6**). We identified a clear separation between these groups and 19 out of 387 metabolites were significantly different ($p < 0.05$, **Table S6**) between the two groups. The top three most altered metabolites ($p < 0.01$) were TDCA, Glucono-D-lactone and Isobutyryl-L-carnitine. All three top metabolites were increased in CD samples (**Figure 4C**). The most altered metabolite, TDCA, is a conjugated bile acid that was shown to be proinflammatory[29]. TDCA is mainly produced by gut microbes, particularly by Clostridium XIVa and Clostridium XI, with 7-α-dehydroxylation of taurocholic acid and cholic acid[30]. This result is consistent with our microbiota analysis since we identified several Clostridium XIVa OTUs that were significantly more abundant in CD samples, especially at age 5 (**Figure 4D**). The heat map shows 50 of the most altered metabolites and indicates a strong signature in the plasma metabolome in which 19 metabolites were significantly altered ($p<0.05$) (**Figure 4E**).

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

DISCUSSION

Recent studies have demonstrated strong associations between the gut microbiota and the pathogenesis of autoimmune diseases[31]. Studies of the gut microbiome in CD have demonstrated intestinal dysbiosis in CD patients[9 13 32 33]. However, the majority of these studies were performed using adult samples with diagnosed disease and none of them used a longitudinal approach as this study. Human gut microbiota development is divided into three phases; a developmental phase (months 3-14), a transitional phase (months 15-30), and a stable phase (months 31-46)[20]. Because recent data show that most childhood CD cases will develop by age 5years[34], we analyzed samples representing all of these critical phases. We report, for what we believe is the first time, that there are significant differences in microbiome composition and function at each developmental phase in CD progressors.

In this study, we first showed that alpha diversity was significantly higher in the CD progressors compared to healthy controls at age 1 (**Figure 1A**). This was an unexpected finding since alpha diversity is reported to be lower in children developing Type 1 Diabetes indicating a difference between two immune-mediated diseases[35]. Consistent with previous reports[33], we identified the proportion of phylum Firmicutes higher in CD progressors at age 1. (**Figure 1C**). Bacterial proteases of species mostly classified within the Firmicutes phylum are involved in gluten metabolism and this might be a link to CD[36]. Additionally, we observed that phylum Verrucomicrobia was only identified in control group at age 1. Akkermansia is the only genus of the Verrucomicrobia phylum identified in gut microbiota[37]. 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[38]. 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[39]. Consistent with our
353 findings, other studies have shown that the abundance of *B. dentium* was increased in the CD
354 patients[40]. 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. Interestingly, IgA-deficient patients are more
360 susceptible to variety of pathologies, including CD[41]. Planer et al described mucosal IgA
361 responses progression during two postnatal years in healthy US twins[17]. 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[42] and might be related to increased

inflammation and IgA responses. Flow cytometry analyses showed that the IgA response is highly selective and only a small fraction of the gut microbiota is highly coated with IgA in the first five years of life. More importantly, the percentage of IgA+ bacteria was higher in CD progressors compared with healthy controls at ages 1 and age 5. While a reduction of secretory IgA (sIgA) using infant (4-6 months) fecal samples in CD progressors[12] was reported previously, we did not observe such a defect in our cohort but identified a two fold increase in the number of IgA coated bacteria in CD progressors especially at age 5.

Our analysis revealed 144 OTUs at age 2.5 and 167 OTUs at age 5 years old that were highly IgA coated in the CD progressors. Among them *Coprococcus comes*, *Bacteroides finegoldii* at age 2.5 and *Faecalibacterium prausnitzii* and *Clostridium_XlVa* at age 5 were the main targets of the mucosal immune response in CD progressors. Among these bacteria, *Coprococcus comes* has been recently identified as the main IgA target in the human colon[43]. Notably, we identified 72 OTUs at age 2.5 and 45 OTUs at age 5 that are equally abundant in CD progressors and healthy controls, but selectively targeted by IgA in CD progressors. Some of these OTUs are at the species level. For example, *Bacterioides finegoldii*, and *Bacteroides vulgatus* at age 2.5 and *Peptostreptococcus stomatis* at age 5 were selectively targeted by IgA in CD progressors but not in controls. *B. vulgatus* was implicated in the development of gut inflammation and a previous report identified this bacterium as enriched in older children with CD[44 45]. In agreement with our results, a pathogenic role for Bacteriodes species is found to be related to the loss of integrity of the intestinal epithelial barrier[45].

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

example, glutathione metabolism was greatly decreased in CD progressors. Decreased glutathione redox cycle in CD patients is strongly associated with disease development[46]. At age 5, PICRUSt predicted that retinol metabolism, steroid hormone biosynthesis, and glycosaminoglycan degradation as over-represented pathways in CD progressors. Retinoic acid is one of the products of retinol metabolism and plays a key role in the intestinal immune response[47]. A previous study showed that retinoic acid mediated inflammatory responses to gluten in CD patients[48]. The increased retinol metabolism and glycosaminoglycan degradation pathways in CD progressors are potentially related to chronic inflammation. Indeed, glycosaminoglycan help to form a protective barrier for the intestinal mucin. The breakdown of glycosaminoglycan is reported to be associated with inflammatory response in intestinal disorders such as IBD[49]. These results suggest that during the developmental phase, the gut microbiota functions in CD progressors were related to shaping the gut microbiota composition. Entering the transition phase, the gut microbiota in CD progressors displayed more proinflammatory and oxidative stress related features. At stable phases, gut microbiota in CD progressors begin to become more involved in functions related to the clinical manifestation of the disease. This longitudinal observation provides insight into the proinflammatory and pathogenic function of gut microbiota in different stages of early CD pathogenesis.

Although the hallmark of the CD is intestinal inflammation, the disease affects different tissues. To determine the systemic effects of gut microbiota on different organs, we performed a comparative plasma metabolomics analysis at age 5. In agreement with the gut microbiota analysis, plasma metabolites were significantly altered prior to diagnosis in CD progressors. The top plasma metabolites altered were TDCA and Isobutyryl-L-carnitine (**Figure 4C**) in which both were increased two-fold in CD progressors. TDCA is a conjugated bile acid that is shown to

be proinflammatory[29] and is mainly produced by gut bacteria, particularly by Clostridium XIVa and Clostridium XI[30]. This observation suggests that the plasma TDCA detected in our study is secondary to the increased abundance of some Clostridium XIVa species in CD progressors at age 5 (**Fig 4D**). Likewise, Isobutyry-l-carnitine is a member of acylcarnitines. As the byproduct of incomplete beta oxidation, the increased isobutyry-l-carnitine is related to abnormalities in fatty acid metabolism[50] and activates proinflammatory signaling[51]. Pathway analysis for plasma metabolites identified several pathways including pentose phosphate pathway (PPP), lysine degradation, and glycerolipid metabolism. PPP was identified as the most significantly altered pathway and stimulates formation of NADPH for antioxidant, thereby controlling cell inflammation. Thus, plasma metabolites of CD progressors are a component of the inflammatory storm.

Currently, the only way to treat CD is strict adherence to a gluten-free (GF) diet, but 20% of patients do not respond to GF diet and continue to have persistent or recurrent symptoms[52]. CD permanently reshapes intestinal immunity and alterations in TCRγδ+ intraepithelial lymphocytes in particular may underlie non-responsiveness to the GF diet[53]. Our findings suggest that the inflammatory nature of the CD progressors' gut microbiota is a key component of intestinal inflammation in CD. The proinflammatory factors identified in this study potentially trigger local and systemic inflammation independent of the diet and may explain a failure to respond to GF diet in some patients. Taken together, our findings suggest that the gut microbiota of CD progressors in the first five years of life has profound effects on the inflammatory response and can potentially contribute to onset and progression of CD. These early markers, for example TDCA, have the potential to serve as useful biomarkers for CD diagnosis.

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

Acknowledgements:

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

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

461 assisted with research design and writing. All authors helped the analysis of the data that they
462 contributed to produce and approved the final version of the manuscript.
463

464 **REFERENCES**

465 1. Lebwohl B, Sanders DS, Green PHR. Coeliac disease. *Lancet* 2018;391(10115):70-81. doi:
466 10.1016/S0140-6736(17)31796-8 [published Online First: 2017/08/02]
467 2. Lindfors K, Ciacci C, Kurppa K, et al. Coeliac disease. *Nat Rev Dis Primers* 2019;5(1):3. doi:
468 10.1038/s41572-018-0054-z [published Online First: 2019/01/12]
469 3. Liu E, Lee HS, Aronsson CA, et al. Risk of pediatric celiac disease according to HLA haplotype and
470 country. *N Engl J Med* 2014;371(1):42-9. doi: 10.1056/NEJMoa1313977 [published Online First:
471 2014/07/06]
472 4. Wolters VM, Wijmenga C. Genetic background of celiac disease and its clinical implications. *Am J*
473 *Gastroenterol* 2008;103(1):190-5. doi: 10.1111/j.1572-0241.2007.01471.x [published Online
474 First: 2008/01/11]
475 5. King JA, Jeong J, Underwood FE, et al. Incidence of Celiac Disease Is Increasing Over Time: A
476 Systematic Review and Meta-analysis. *Am J Gastroenterol* 2020 doi:
477 10.14309/ajg.0000000000000523 [published Online First: 2020/02/06]
478 6. Greco L, Romino R, Coto I, et al. The first large population based twin study of coeliac disease. *Gut*
479 2002;50(5):624-8. doi: 10.1136/gut.50.5.624 [published Online First: 2002/04/16]
480 7. Kuja-Halkola R, Lebwohl B, Halfvarson J, et al. Heritability of non-HLA genetics in coeliac disease: a
481 population-based study in 107 000 twins. *Gut* 2016;65(11):1793-98. doi: 10.1136/gutjnl-2016-
482 311713 [published Online First: 2016/05/22]
483 8. Lionetti E, Catassi C. The Role of Environmental Factors in the Development of Celiac Disease: What Is
484 New? *Diseases* 2015;3(4):282-93. doi: 10.3390/diseases3040282 [published Online First:
485 2015/10/27]
486 9. Nistal E, Caminero A, Vivas S, et al. Differences in faecal bacteria populations and faecal bacteria
487 metabolism in healthy adults and celiac disease patients. *Biochimie* 2012;94(8):1724-9. doi:
488 10.1016/j.biochi.2012.03.025 [published Online First: 2012/05/01]
489 10. Serena G, Yan S, Camhi S, et al. Proinflammatory cytokine interferon-gamma and microbiome-
490 derived metabolites dictate epigenetic switch between forkhead box protein 3 isoforms in
491 coeliac disease. *Clin Exp Immunol* 2017;187(3):490-506. doi: 10.1111/cei.12911 [published
492 Online First: 2016/12/10]
493 11. Sellitto M, Bai G, Serena G, et al. Proof of concept of microbiome-metabolome analysis and delayed
494 gluten exposure on celiac disease autoimmunity in genetically at-risk infants. *PLoS One*
495 2012;7(3):e33387. doi: 10.1371/journal.pone.0033387 [published Online First: 2012/03/21]
496 12. Olivares M, Walker AW, Capilla A, et al. Gut microbiota trajectory in early life may predict
497 development of celiac disease. *Microbiome* 2018;6(1):36. doi: 10.1186/s40168-018-0415-6
498 [published Online First: 2018/02/21]
499 13. Cheng J, Kalliomaki M, Heilig HG, et al. Duodenal microbiota composition and mucosal homeostasis
500 in pediatric celiac disease. *BMC Gastroenterol* 2013;13:113. doi: 10.1186/1471-230X-13-113
501 [published Online First: 2013/07/13]
502 14. Bunker JJ, Bendelac A. IgA Responses to Microbiota. *Immunity* 2018;49(2):211-24. doi:
503 10.1016/j.immuni.2018.08.011 [published Online First: 2018/08/23]

15. Wilmore JR, Gaudette BT, Gomez Atria D, et al. Commensal Microbes Induce Serum IgA Responses that Protect against Polymicrobial Sepsis. *Cell Host Microbe* 2018;23(3):302-11 e3. doi: 10.1016/j.chom.2018.01.005 [published Online First: 2018/02/27]
16. Bunker JJ, Erickson SA, Flynn TM, et al. Natural polyreactive IgA antibodies coat the intestinal microbiota. *Science* 2017;358(6361) doi: 10.1126/science.aan6619 [published Online First: 2017/10/04]
17. Planer JD, Peng Y, Kau AL, et al. Development of the gut microbiota and mucosal IgA responses in twins and gnotobiotic mice. *Nature* 2016;534(7606):263-6. doi: 10.1038/nature17940 [published Online First: 2016/06/10]
18. Palm NW, de Zoete MR, Cullen TW, et al. Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease. *Cell* 2014;158(5):1000-10. doi: 10.1016/j.cell.2014.08.006 [published Online First: 2014/08/30]
19. Viladomiu M, Kivolowitz C, Abdulhamid A, et al. IgA-coated E. coli enriched in Crohn's disease spondyloarthritis promote TH17-dependent inflammation. *Sci Transl Med* 2017;9(376) doi: 10.1126/scitranslmed.aaf9655 [published Online First: 2017/02/10]
20. Stewart CJ, Ajami NJ, O'Brien JL, et al. Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature* 2018;562(7728):583-88. doi: 10.1038/s41586-018-0617-x [published Online First: 2018/10/26]
21. Drolet J, Tolstikov V, Williams BA, et al. Integrated Metabolomics Assessment of Human Dried Blood Spots and Urine Strips. *Metabolites* 2017;7(3) doi: 10.3390/metabo7030035 [published Online First: 2017/07/18]
22. Ludvigsson JF, Andersson E, Ekbom A, et al. External review and validation of the Swedish national inpatient register. *BMC Public Health* 2011;11:450. doi: 10.1186/1471-2458-11-450 [published Online First: 2011/06/11]
23. Koenig JE, Spor A, Scalfone N, et al. Succession of microbial consortia in the developing infant gut microbiome. *Proc Natl Acad Sci U S A* 2011;108 Suppl 1:4578-85. doi: 10.1073/pnas.1000081107 [published Online First: 2010/07/30]
24. Langille MG, Zaneveld J, Caporaso JG, et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol* 2013;31(9):814-21. doi: 10.1038/nbt.2676 [published Online First: 2013/08/27]
25. Tan FY, Tang CM, Exley RM. Sugar coating: bacterial protein glycosylation and host-microbe interactions. *Trends Biochem Sci* 2015;40(7):342-50. doi: 10.1016/j.tibs.2015.03.016 [published Online First: 2015/05/06]
26. Matilla MA, Krell T. The effect of bacterial chemotaxis on host infection and pathogenicity. *FEMS Microbiol Rev* 2018;42(1) doi: 10.1093/femsre/fux052 [published Online First: 2017/10/27]
27. Aharonowitz Y, Cohen G, Martin JF. Penicillin and cephalosporin biosynthetic genes: structure, organization, regulation, and evolution. *Annu Rev Microbiol* 1992;46:461-95. doi: 10.1146/annurev.mi.46.100192.002333 [published Online First: 1992/01/01]
28. Moore AM, Ahmadi S, Patel S, et al. Gut resistome development in healthy twin pairs in the first year of life. *Microbiome* 2015;3:27. doi: 10.1186/s40168-015-0090-9 [published Online First: 2015/06/27]
29. Fujisaka S, Ussar S, Clish C, et al. Antibiotic effects on gut microbiota and metabolism are host dependent. *J Clin Invest* 2016;126(12):4430-43. doi: 10.1172/JCI86674 [published Online First: 2016/10/25]
30. Ridlon JM, Kang DJ, Hylemon PB. Bile salt biotransformations by human intestinal bacteria. *J Lipid Res* 2006;47(2):241-59. doi: 10.1194/jlr.R500013-JLR200 [published Online First: 2005/11/22]

31. Marietta E, Horwath I, Balakrishnan B, et al. Role of the intestinal microbiome in autoimmune diseases and its use in treatments. *Cell Immunol* 2019;339:50-58. doi: 10.1016/j.cellimm.2018.10.005 [published Online First: 2018/10/28]
32. D'Argenio V, Casaburi G, Precone V, et al. Metagenomics Reveals Dysbiosis and a Potentially Pathogenic N. flavescens Strain in Duodenum of Adult Celiac Patients. *Am J Gastroenterol* 2016;111(6):879-90. doi: 10.1038/ajg.2016.95 [published Online First: 2016/04/06]
33. Wacklin P, Kaukinen K, Tuovinen E, et al. The duodenal microbiota composition of adult celiac disease patients is associated with the clinical manifestation of the disease. *Inflamm Bowel Dis* 2013;19(5):934-41. doi: 10.1097/MIB.0b013e31828029a9 [published Online First: 2013/03/13]
34. Lionetti E, Castellaneta S, Francavilla R, et al. Introduction of gluten, HLA status, and the risk of celiac disease in children. *N Engl J Med* 2014;371(14):1295-303. doi: 10.1056/NEJMoa1400697 [published Online First: 2014/10/02]
35. Kostic AD, Gevers D, Siljander H, et al. The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. *Cell Host Microbe* 2015;17(2):260-73. doi: 10.1016/j.chom.2015.01.001 [published Online First: 2015/02/11]
36. Caminero A, Herran AR, Nistal E, et al. Diversity of the cultivable human gut microbiome involved in gluten metabolism: isolation of microorganisms with potential interest for coeliac disease. *FEMS Microbiol Ecol* 2014;88(2):309-19. doi: 10.1111/1574-6941.12295 [published Online First: 2014/02/07]
37. Geerlings SY, Kostopoulos I, de Vos WM, et al. Akkermansia muciniphila in the Human Gastrointestinal Tract: When, Where, and How? *Microorganisms* 2018;6(3) doi: 10.3390/microorganisms6030075 [published Online First: 2018/07/26]
38. Cani PD, de Vos WM. Next-Generation Beneficial Microbes: The Case of Akkermansia muciniphila. *Front Microbiol* 2017;8:1765. doi: 10.3389/fmicb.2017.01765 [published Online First: 2017/10/12]
39. Bonder MJ, Tigchelaar EF, Cai X, et al. The influence of a short-term gluten-free diet on the human gut microbiome. *Genome Med* 2016;8(1):45. doi: 10.1186/s13073-016-0295-y [published Online First: 2016/04/23]
40. Collado MC, Donat E, Ribes-Koninckx C, et al. Imbalances in faecal and duodenal Bifidobacterium species composition in active and non-active coeliac disease. *BMC Microbiol* 2008;8:232. doi: 10.1186/1471-2180-8-232 [published Online First: 2008/12/24]
41. Chow MA, Lebwohl B, Reilly NR, et al. Immunoglobulin A deficiency in celiac disease. *J Clin Gastroenterol* 2012;46(10):850-4. doi: 10.1097/MCG.0b013e31824b2277 [published Online First: 2012/04/06]
42. Delpino MV, Fossati CA, Baldi PC. Occurrence and potential diagnostic applications of serological cross-reactivities between Brucella and other alpha-proteobacteria. *Clin Diagn Lab Immunol* 2004;11(5):868-73. doi: 10.1128/CDLI.11.5.868-873.2004 [published Online First: 2004/09/11]
43. Sterlin D, Fieschi C, Malphettes M, et al. Immune/microbial interface perturbation in human IgA deficiency. *Gut Microbes* 2019;10(3):429-33. doi: 10.1080/19490976.2018.1546520 [published Online First: 2018/11/20]
44. Schippa S, Iebba V, Barbato M, et al. A distinctive 'microbial signature' in celiac pediatric patients. *BMC Microbiol* 2010;10:175. doi: 10.1186/1471-2180-10-175 [published Online First: 2010/06/23]
45. Shiba T, Aiba Y, Ishikawa H, et al. The suppressive effect of bifidobacteria on Bacteroides vulgatus, a putative pathogenic microbe in inflammatory bowel disease. *Microbiol Immunol* 2003;47(6):371-8. doi: 10.1111/j.1348-0421.2003.tb03368.x [published Online First: 2003/08/09]

46. Stojiljkovic V, Pejic S, Kasapovic J, et al. Glutathione redox cycle in small intestinal mucosa and peripheral blood of pediatric celiac disease patients. *An Acad Bras Cienc* 2012;84(1):175-84. doi: 10.1590/s0001-37652012000100018 [published Online First: 2012/03/24]
47. Mucida D, Park Y, Kim G, et al. Reciprocal TH17 and regulatory T cell differentiation mediated by retinoic acid. *Science* 2007;317(5835):256-60. doi: 10.1126/science.1145697 [published Online First: 2007/06/16]
48. DePaolo RW, Abadie V, Tang F, et al. Co-adjuvant effects of retinoic acid and IL-15 induce inflammatory immunity to dietary antigens. *Nature* 2011;471(7337):220-4. doi: 10.1038/nature09849 [published Online First: 2011/02/11]
49. Winslet MC, Poxon V, Allan A, et al. Mucosal glucosamine synthetase activity in inflammatory bowel disease. *Dig Dis Sci* 1994;39(3):540-4. doi: 10.1007/bf02088339 [published Online First: 1994/03/01]
50. Vreken P, van Lint AE, Bootsma AH, et al. Quantitative plasma acylcarnitine analysis using electrospray tandem mass spectrometry for the diagnosis of organic acidaemias and fatty acid oxidation defects. *J Inherit Metab Dis* 1999;22(3):302-6. doi: 10.1023/a:1005587617745 [published Online First: 1999/06/29]
51. Rutkowsky JM, Knotts TA, Ono-Moore KD, et al. Acylcarnitines activate proinflammatory signaling pathways. *Am J Physiol Endocrinol Metab* 2014;306(12):E1378-87. doi: 10.1152/ajpendo.00656.2013 [published Online First: 2014/04/25]
52. Stasi E, Marafini I, Caruso R, et al. Frequency and Cause of Persistent Symptoms in Celiac Disease Patients on a Long-term Gluten-free Diet. *J Clin Gastroenterol* 2016;50(3):239-43. doi: 10.1097/MCG.0000000000000392 [published Online First: 2015/08/19]
53. Mayassi T, Ladell K, Gudjonson H, et al. Chronic Inflammation Permanently Reshapes Tissue-Resident Immunity in Celiac Disease. *Cell* 2019;176(5):967-81 e19. doi: 10.1016/j.cell.2018.12.039 [published Online First: 2019/02/12]

633 **Figure Legends**

634

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

637 **a.** Box plots showing the comparison between CD progressors and healthy controls : the alpha
638 diversity measured by observed (upper panel) and the beta diversity measured by Bray–Curtis
639 dissimilarity (lower panel).
640 **b.** Non-metric MultiDimentional Scaling (NMDS) analysis ordination of sample
641 similarity/dissimilarity between CD progressors and healthy controls at age 1, 2.5, and 5 years
642 old. Each circle represents an individual sample.
643 **c.** Average relative abundance of bacterial phylum (upper panel) or genera (lower panel) of
644 greater than 1% abundance (proportion) between the gut microbiota of CD progressors and
645 healthy controls at age 1, 2.5, and 5 years old (taxa average relative abundance>1%).
646 **d**. Heat map showing the relative abundance of the top OTUs significantly different between CD
647 progressors and healthy controls. Each column represents an individual participant and each raw
648 represents an OTU.
649 **e.** Empirical Bayes quasi-likelihood F-tests analysis for the comparisons of gut microbiota OTUs
650 between CD progressors and healthy controls age 1, 2.5, and 5 years old. Frequency: number of
651 OTUs. **a-e**: age 1: n=5/group; age 2.5: n=16/group; age 5: n=13/group.

652

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

655 **a.** Principle component analysis (PCA) of sample similarity/dissimilarity between IgA+ and IgA-

656 microbiota in healthy control (left) or CD progressors (right).

657 **b.** Flow cytometry results for IgA coating of fecal bacteria from CD progressors and healthy

658 controls at at age 1, 2.5 and 5. Indicated are mean±SEM. *P<0.05 (Two way ANOVA).

659 **c.** Box plots showing the comparison between CD progressors and healthy controls : the alpha

660 diversity measured by observed for IgA+/IgA- microbiota (upper panel), and the beta diversity

661 measured by Bray–Curtis dissimilarity for IgA+/IgA- microbiota (lower panel) at ages 1, 2.5 ,

662 and 5 years old. *P<0.05.

663 **d.** NMDS analysis ordination of IgA+/IgA- microbiota similarity/dissimilarity between CD

664 progressors and healthy controls at ages 1, 2.5, and 5 years old. Each circle represents an

665 individual sample.

666 **e.** Average relative abundance of IgA+/IgA- bacterial phylum (left) or genera (right) of greater

667 than 1% abundance (proportion) between the gut microbiota of CD progressors and healthy

668 controls at ages 1, 2.5, and 5 years old (taxa average relative abundance>1%).

669 **f** Heat map showing the relative abundance of the top OTUs significantly different between

670 IgA+/IgA- CD progressors and healthy controls (OTUs=51 based on p-value). Each column

671 represents an individual participant and each raw represents an OTU.

672 **a-f**: age 1: n=5/group; age 2.5: n=16/group; age 5: n=13/group.

673

674 **Figure 3. The Enriched Microbial Pathways Altered in CD Progressors' Gut Microbiota**

675 **are Significantly Different from Healthy Controls**

676 **a.** Heat map of PICRUSt predicted metabolic pathways of CD progressors and healthy controls.

677 Each column represents an individual participant and each raw represents a predicted microbial

678 functional pathway. Color code is shown on the figure.

679 **b.** Heat map of predicted metabolic pathways of CD progressors and healthy control obtained
680 from PICRUSt analysis. Each column represents an individual participant and each raw
681 represents a predicted microbial functional pathway. **a-b**: ages 1year old : n=5/group); age 2.5
682 years old: n=16/group; age 5 years old: n=13/group.
683
684 **Figure 4. CD Progressors Have a Distinct Plasma Metabolic Profile at Age 5**
685 **a.** Partial Least Square-Discrimination Analysis (PLS-DA) of plasma metabolites for CD
686 progressors (n=10) and controls (n=9).
687 **b.** Volcano plot of plasma metabolites with fold change threshold ($|\log_2 (FC)|>1.2$) and t-tests
688 threshold ($-\log_{10}(p)>0.1$). The red dots represent metabolites above the threshold. Fold changes
689 are $\log_2$ transformed and p values are $\log_{10}$ transformed.
690 **c.** Box plots showing three most significantly increased metabolites in CD plasma.
691 **d.** Box plots showing the representative Clostridium XIVa bacteria abundance between CD
692 progressors and healthy controls.
693 **e.** Heatmap showing 50 of the most altered metabolites
694 **f.** The Pathway Analysis (combined results from powerful pathway enrichment analysis with
695 pathway topology analysis) identify the most altered metabolic pathways between
696 CD progressors and healthy controls. Pathway impact value (x) is calculated from pathway
697 topology analysis. p is the original p value calculated from the enrichment analysis and depicted
698 on a logarithmic scale.
699 **Figure S1. Box plots of representative bacteria that are differentially distributed in CD**
700 **progressors and healthy controls at a. Age 1 year old. b. Age 2.5 years old**
701 **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.

b. Empirical Bayes quasi-likelihood F-tests analysis for the comparisons of IgA coated or non coated gut microbiota OTUs between CD progressors and healthy controls (upper row: age 2.5 years old; lower row: age 5 years old).

725 **c.** Box plots showing representative bacteria in which abundances were similar in the gut
726 microbiota (presorting samples) but differently targeted by IgA (upper row: age 2.5 years old;
727 lower row: age 5 years old).
728 **a-c**: Age 1 : n=5/group; Age 2.5: n=16/group; age 5:n=13/group.
729
730 **Figure S4. PICRUSt predicted genes enriched in pre-IgA seq and post-IgA seq samples.**
731 **a.** Heat map of PICRUSt predicted microbial genes enriched in CD progressors or healthy
732 controls Each column represents an individual participants and each raw represents a predicted
733 microbial gene. COLOR CODE?
734 **b.** Heat map of PICRUSt predicted microbial genes enriched in CD progressors or healthy
735 controls after IgA sequencing. Each column represents an individual participants and each raw
736 represents a predicted microbial gene.
737 **a-b**: age 1: n=5/group; age 2.5: n=16/group; age 5: n=13/group.
738
739 **Figure S5. Box plots of representative metabolic pathways that are affected in CD**
740 **progressors. A.** Age 1**, B.** Age 2.5 , **C.** Age 5
741
742 **Figure S6. Heat map showing the correlation of most significantly altered top 20 OTUs and**
743 **most significantly altered 20 metabolic pathways at a.** Age 1, **b.** Age 2.5 **, c.** Age 5**.** Color key
744 represents the Pearson correlation coeffient.
745

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

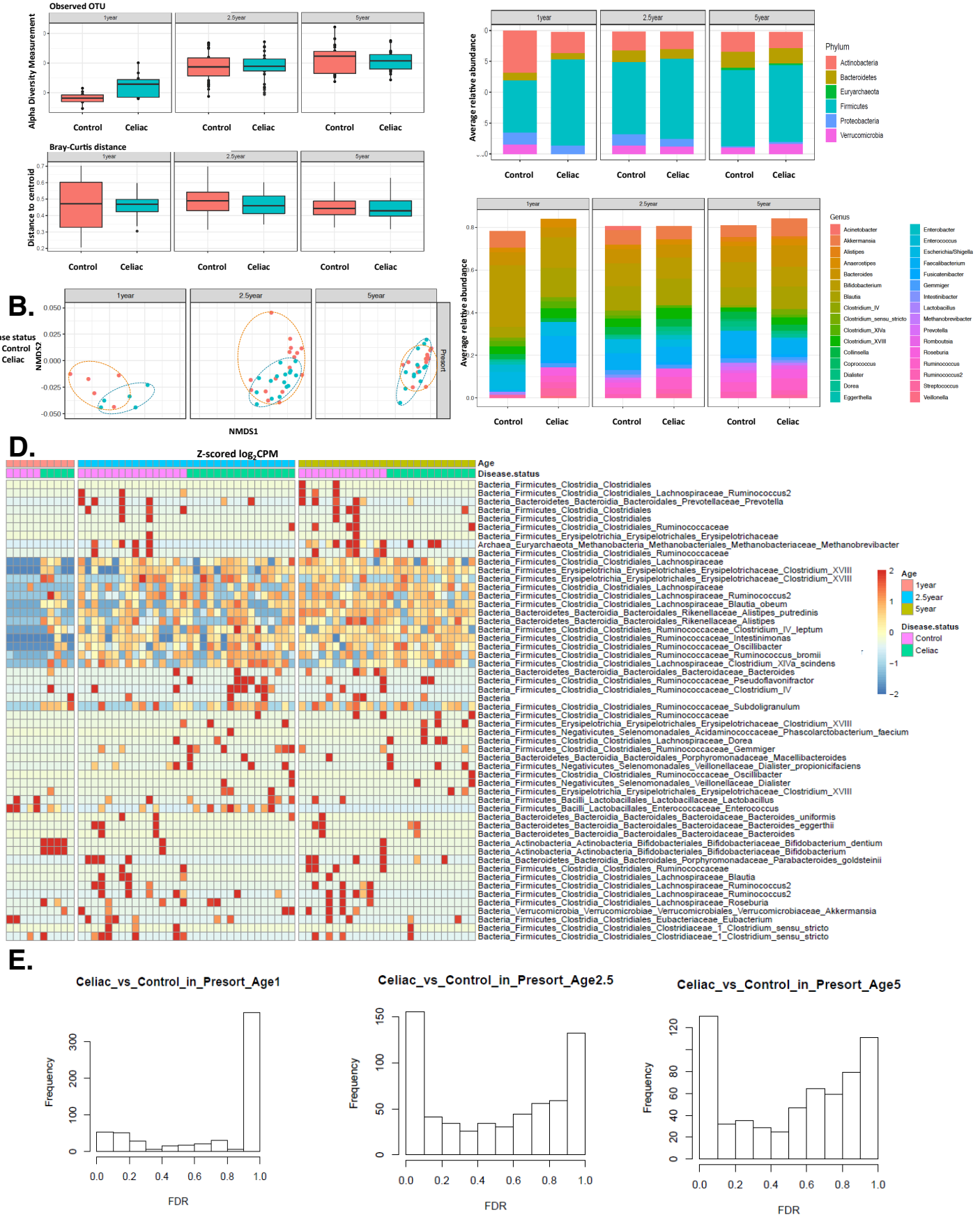


Figure 2 POSTSORTING FIGURES: IgA Seq analysis differences

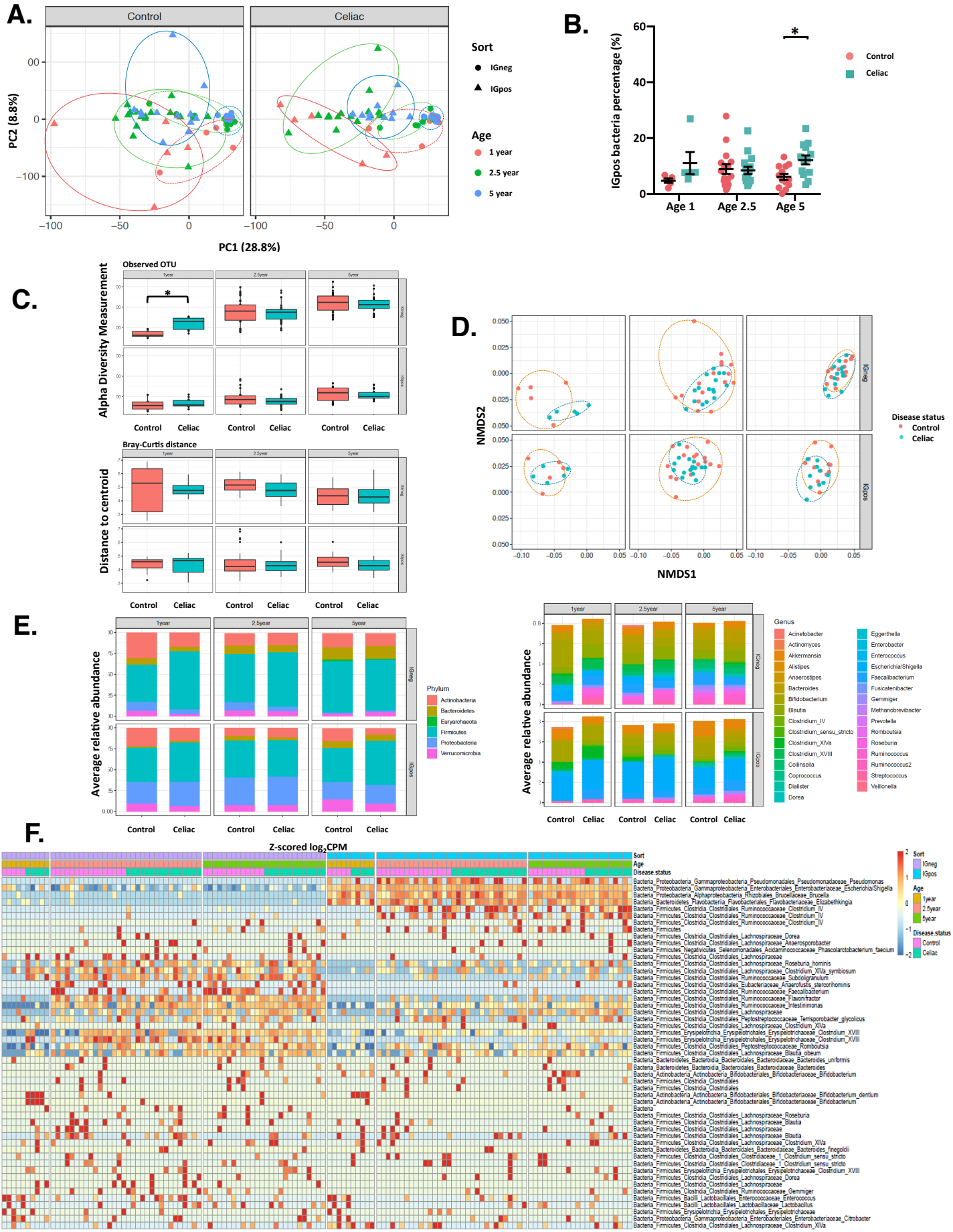


Figure 3 PICRUST

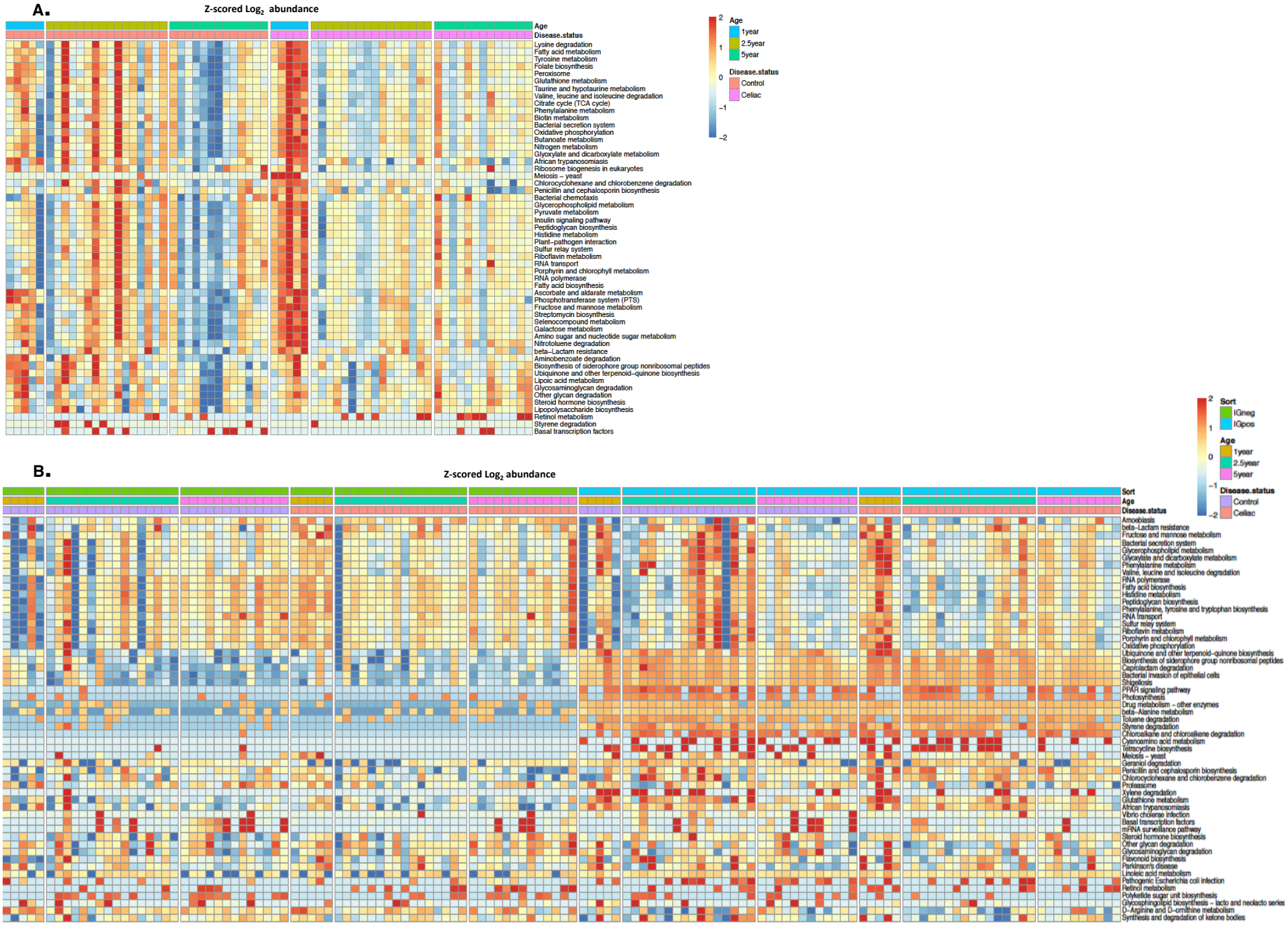


Figure 4 Plasma Metabolomics

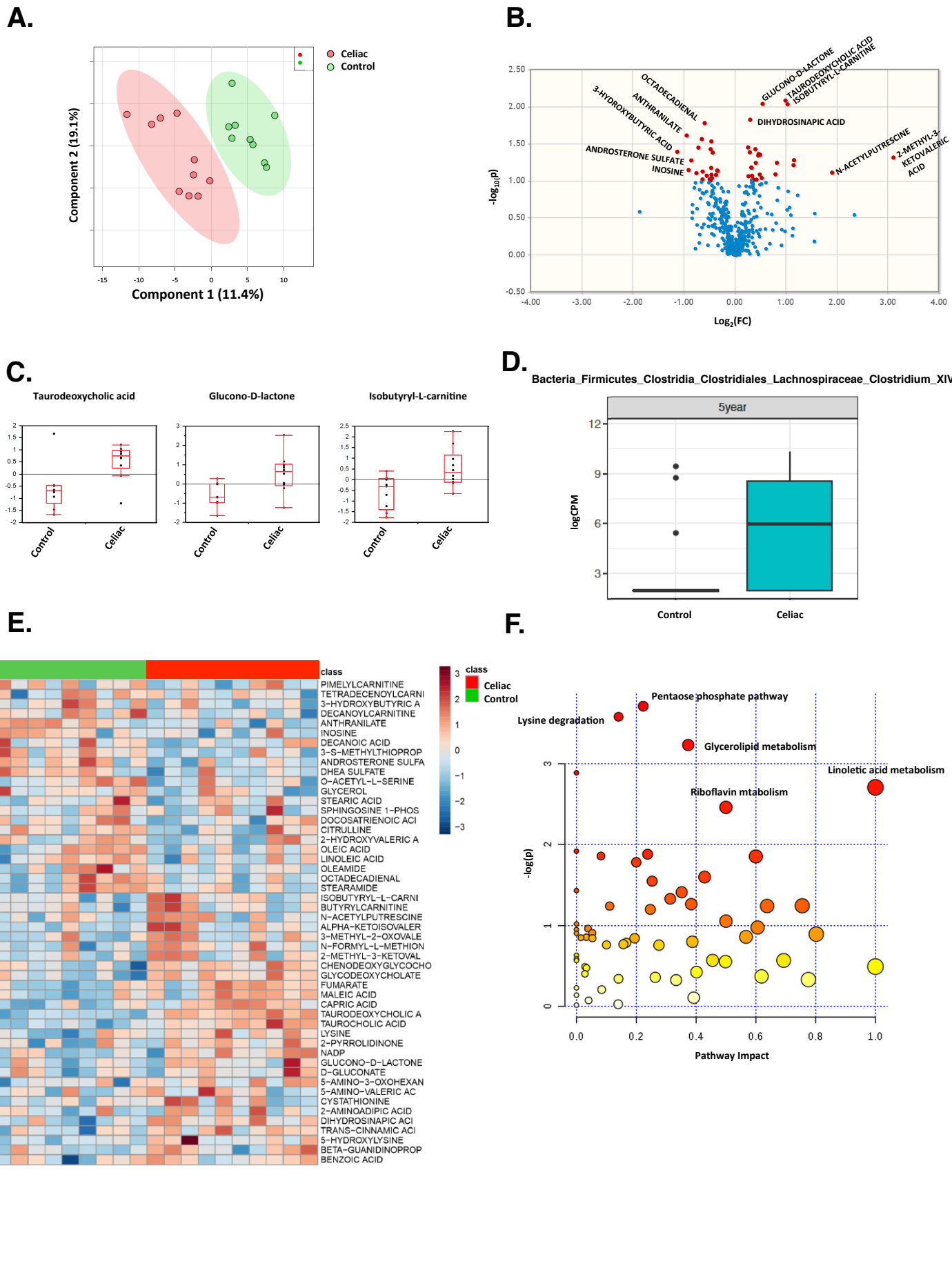
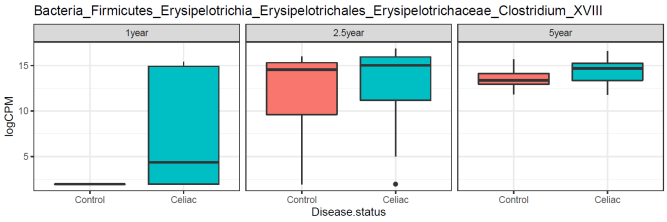
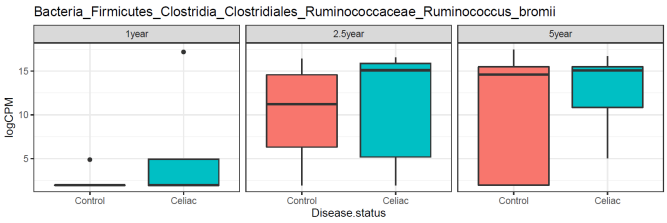
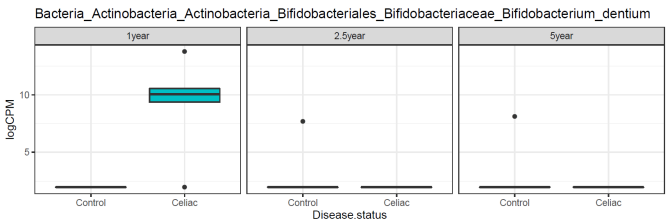


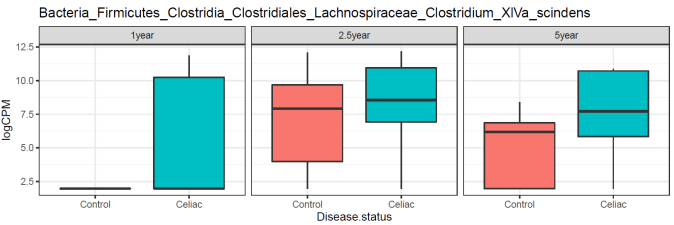
Figure S1. Pre-sorting. Most Significant Differences for PRESORTING

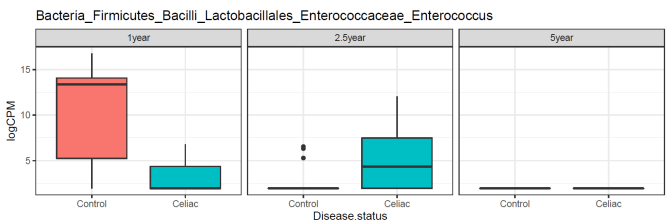
A. Age 1



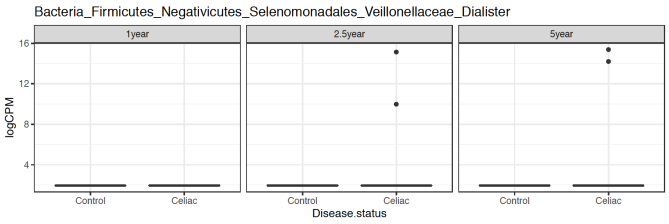


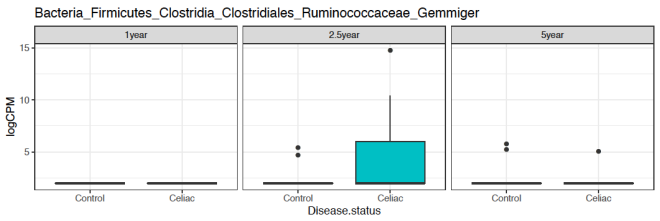


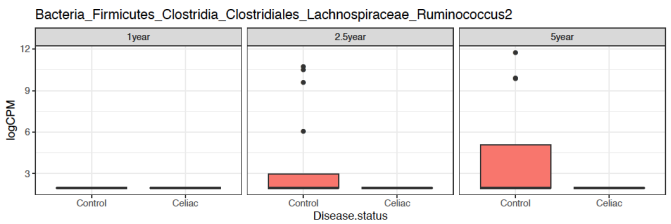


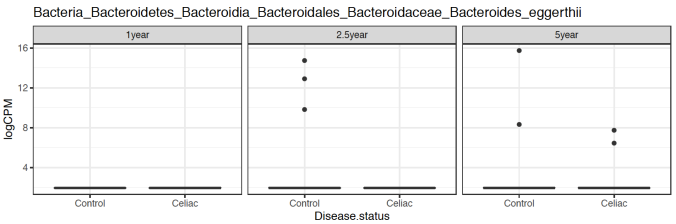


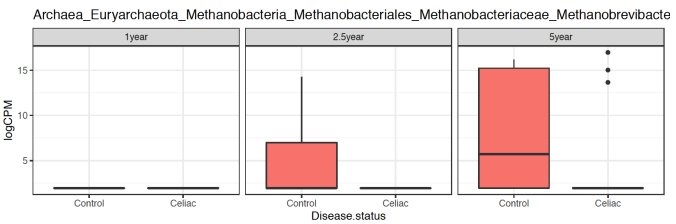
B. Age 2.5



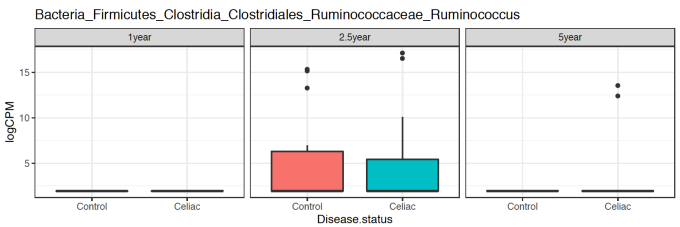


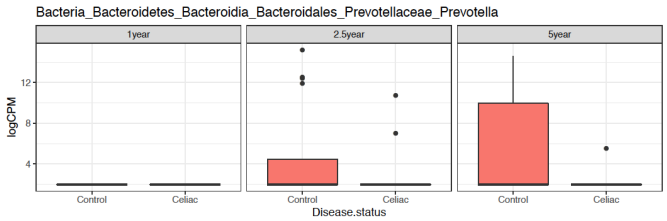


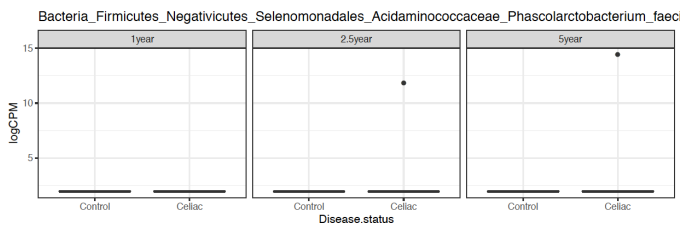




C. Age 5







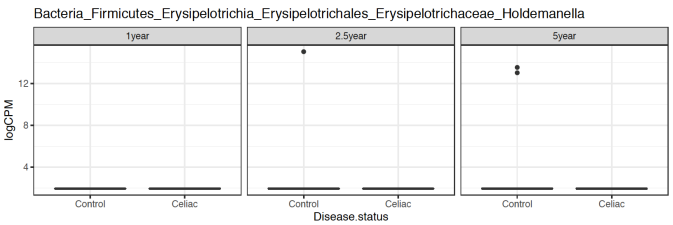
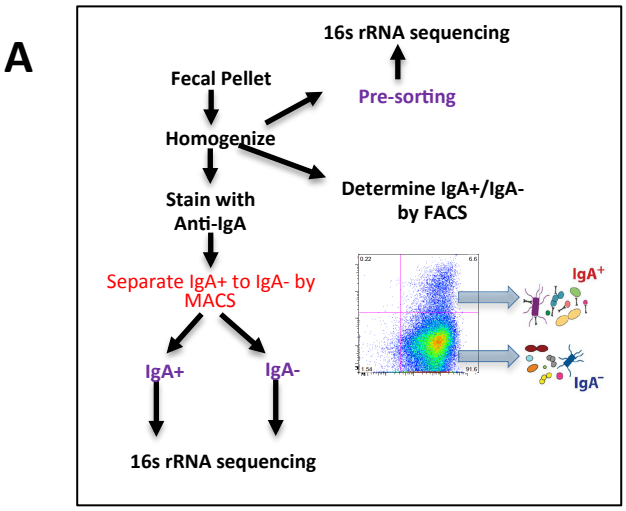
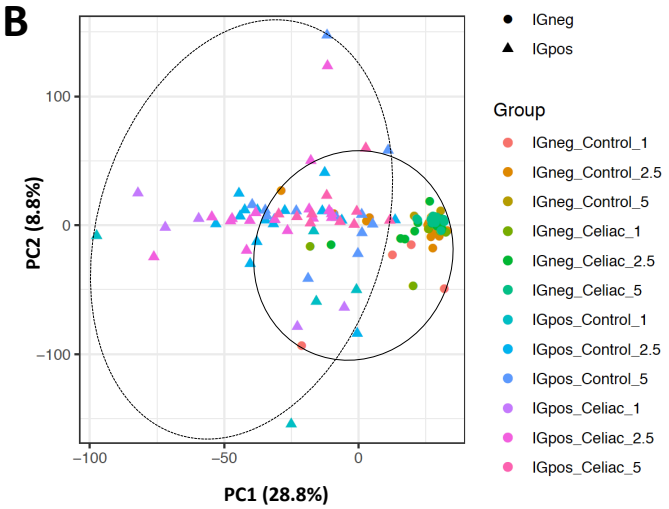
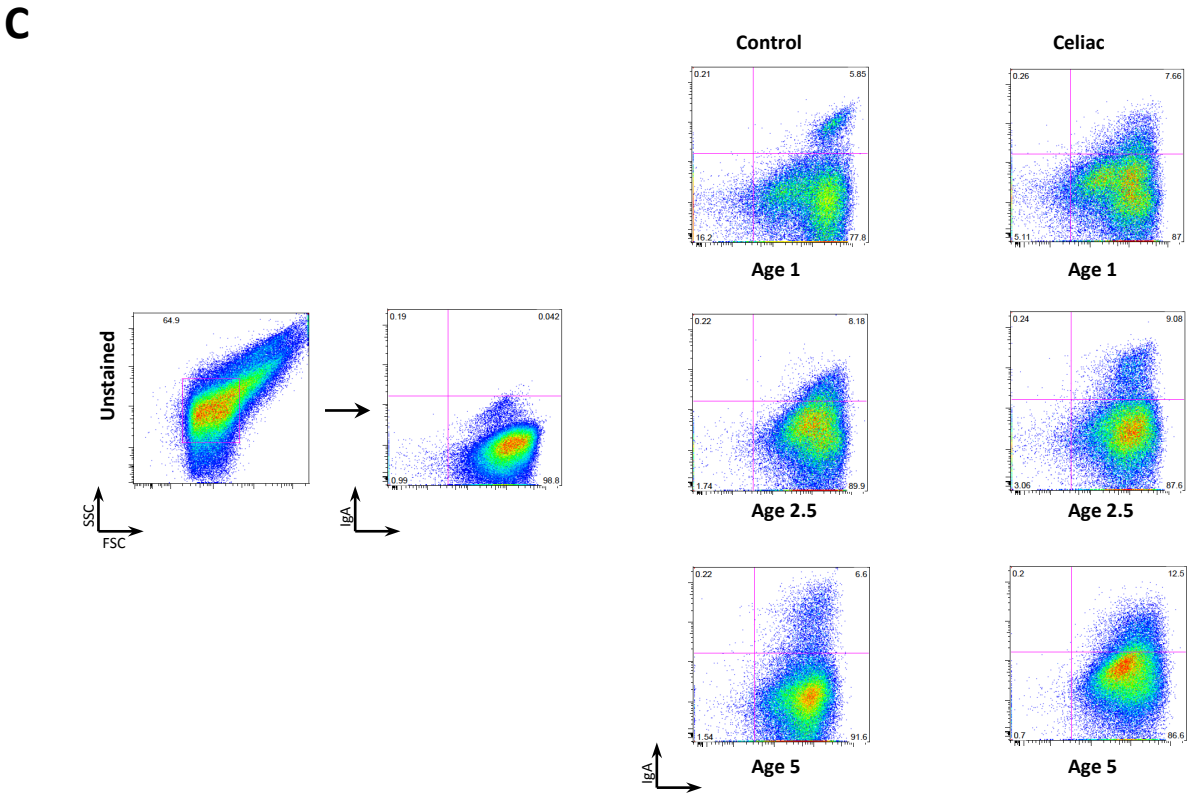


Figure S2. Post-sorting. IgA+ bacteria percentage







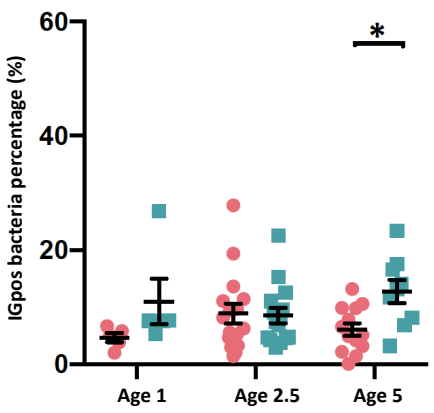
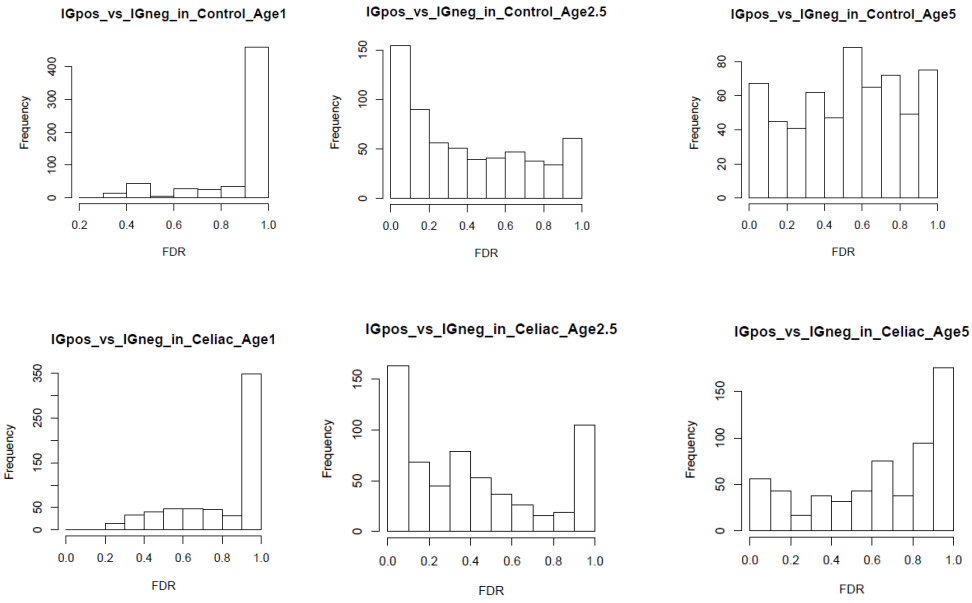
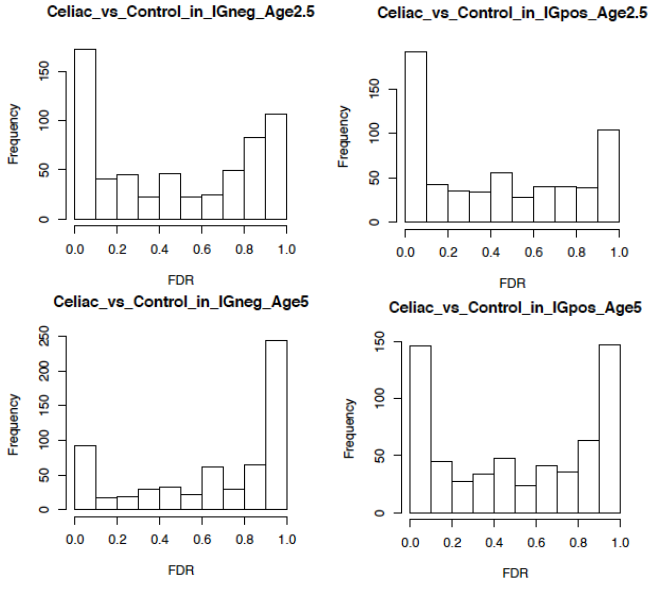


Figure S3. Post-sorting. Most Significant Differences for post-sorting

A



B



C

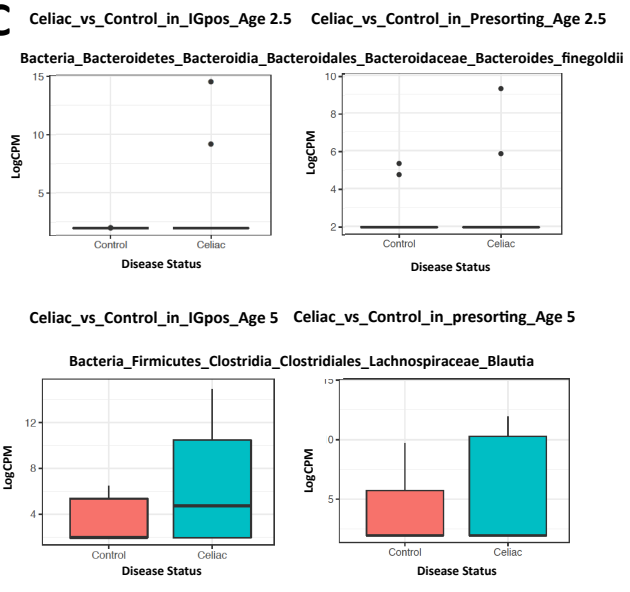


Figure S4. PICRUST genes heat map

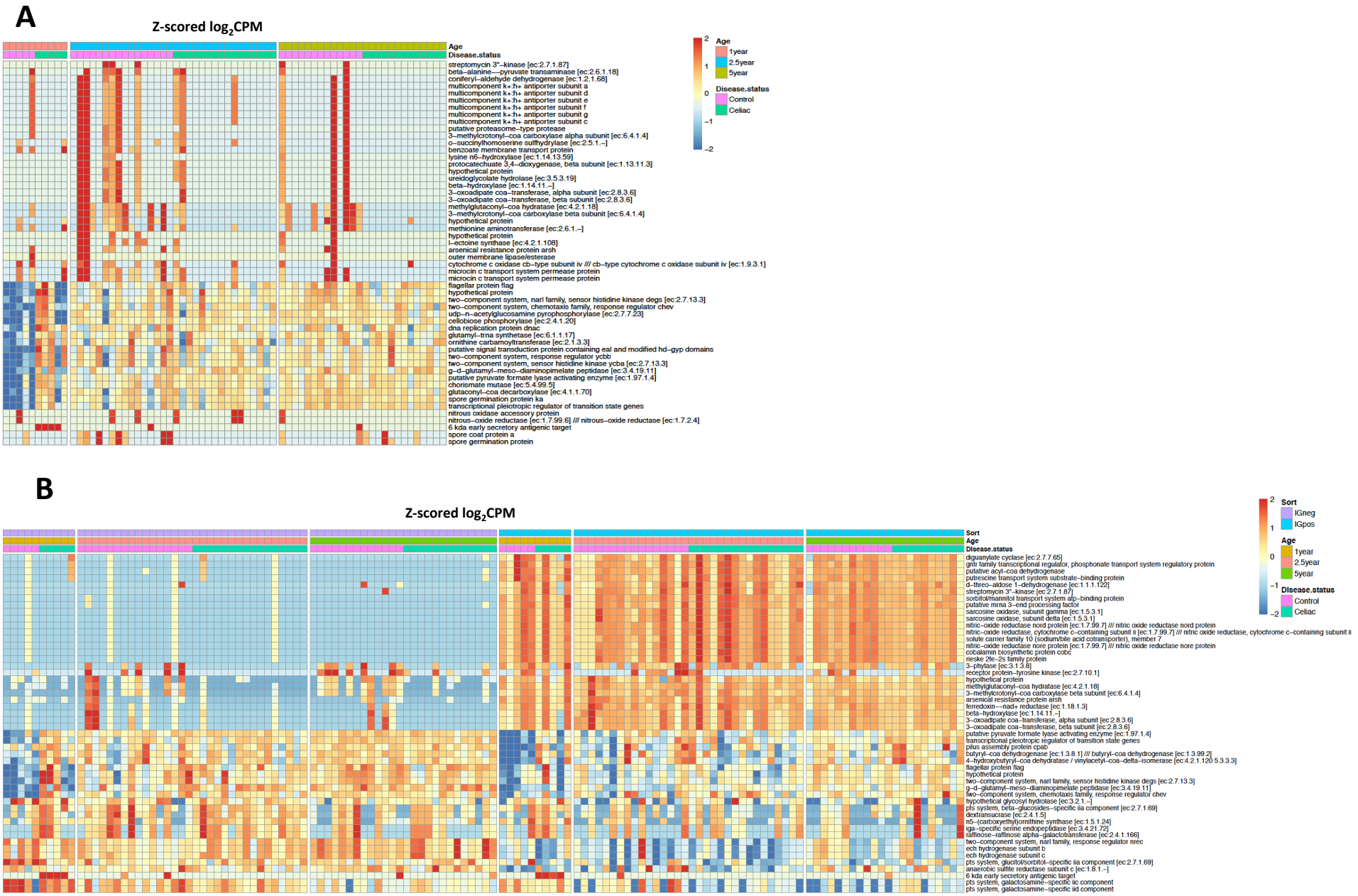


Figure S5. PICRUSt

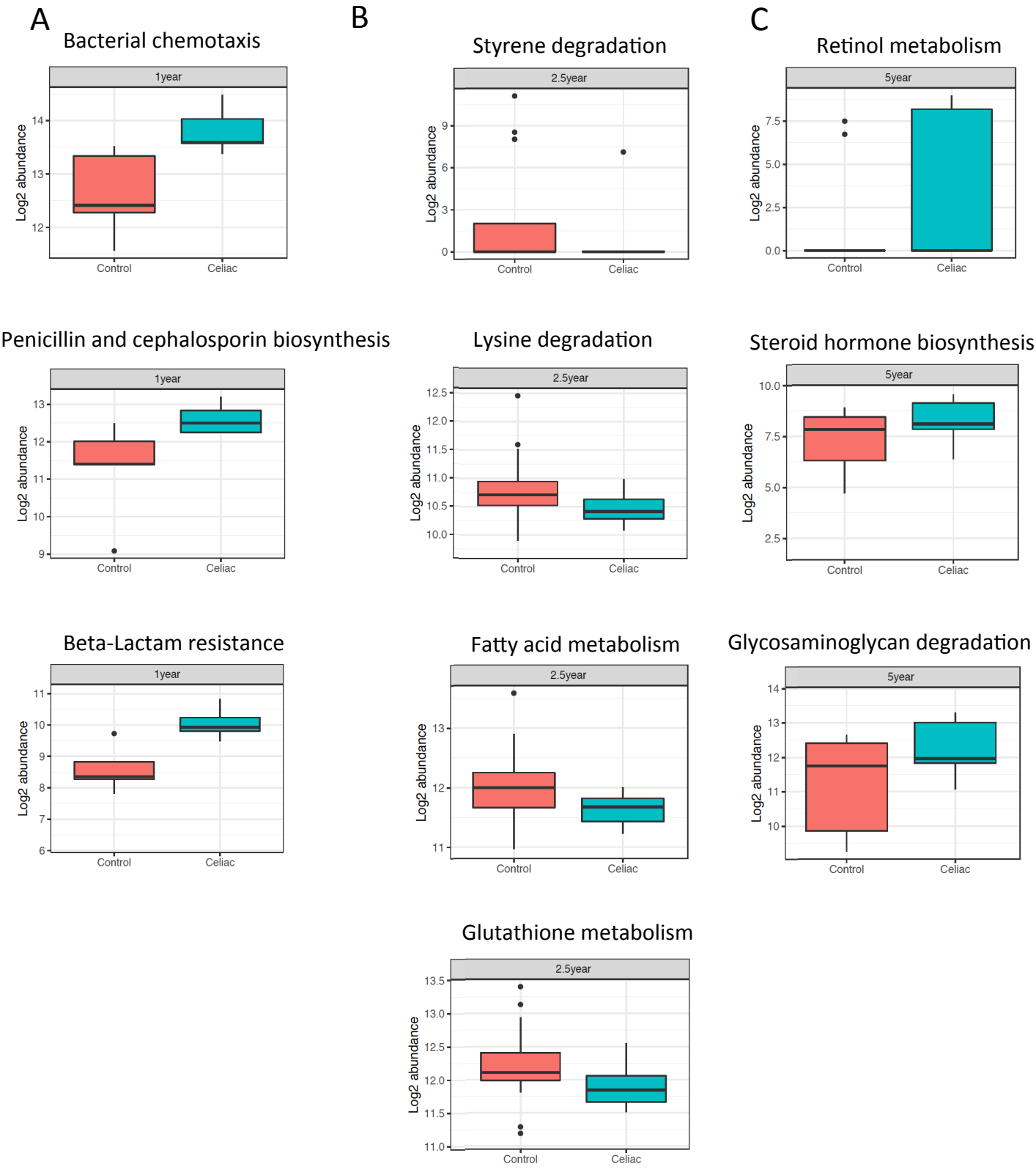


Figure S6. Pearson correlation analysis