Understanding microbial ecology of gut fecal samples of Crohn's disease patients when treated with Exclusive Enteral Nutrition

Student Name: YUCHENG ZHI
Student ID: 2540767z
Supervisor: Dr Umer Zeeshan Ijaz
Co-supervisor: Dr Nadine Elena Juarez Rendon

August 17th 2022

ACKNOWLEDGEMENT

I would want to thank my mentor and advisor for allowing me the chance to study this brand-new, highly fascinating subject. Additionally, I want to thank him for all of our meetings and for all of the advice he gave me.

Additionally, I want to thank my other Masters students for being there for me anytime I needed assistance.

ABSTRACT

In the process of EEN, the impact of specific bacteria on the intestinal tract also needs to be paid attention to, because studies have shown that the beneficial effects on the intestinal tract are often caused by some specific bacteria groups. Similarly, some specific bacteria groups can cause adverse reactions such as intestinal inflammation, indigestion, etc. Therefore, this study will pay additional attention to the dynamic changes of beneficial and harmful flora during the EEN process, and further analyze the significant flora.

The 117 fecal from 23 CD and 21 healthy kids were collected. Fecal samples from CD kids were taken before, during, and after patients resumed their regular diets. Shotgun metagenomics and 16S rRNA gene sequencing were used to characterize the composition of the microbiota and its functional capacity.

The mean Shannon diversity was higher in controls than CD children prior to EEN; differences were observed in 36 genera. NTI shows a deterministic environment in CD children and a stochastic environment in controls. It is more deterministic in healthy people. CD children show the percentage of dispersal limitation dropping from 69.52% to 59.14% when they are taking the treatment. Healthy children are undominated (41.97%). Also there is 27.53% of variable selection. C4 is significantly positive correlated with Shannon diversity (P= 7.653e-06). Furthermore, Atopobium and Fusobacterium have strong relationship with calprotectin which corresponds to Intestinal inflammation. Finally, Streptococcus is much more abundant in CD children. In contrast, we can find more Subdoligranulum and Collinsella in controls.

CD children before EEN treatment had lower levels of intestinal biodiversity than healthy children. Regression model showed that C4 in short chain fatty acid can effectively increase the diversity in children's gut (P=7.653e-06), and it plays an important role in resisting intestinal inflammation. It maintains the integrity of the intestinal wall, which is an important barrier within the gut. The bacteria that are more distributed in healthy children mostly belong to the Firmicutes genera, and such bacteria help to produce short chain fatty acid, which also explains the above findings. Therefore, in the subsequent treatment of Crohn's disease, means to promote such flora can be considered, thereby reducing inflammation in the intestine. Also, the experiment found that Atopobium (0.38), Fusobacterium (0.17) and calprotectin were positively correlated, and these two types of bacteria were directly related to causing Crohn's disease.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Abbreviation	Explanation
EEN	Exclusive Enteral Nutrition
CD	Crohn's Disease
CS	Corticosteroid
NG	Nasogastric
PCDAI	Pediatric Crohn's Disease Activity Index
IBD	Inflammatory Bowel Disease
NGS	Next-generation Sequencing
OTU	Operational Taxonomic Unit
NTI	Nearest Taxon Index
QPE	Quantitative Process Estimates
MNTD	Mean nearest phylogenetic taxon distance
NST	Normalized Stochasticity Ratio
HITChip	Human Intestinal Tract Chip
NMDS	Non-metric Multidimensional Scaling
SCFA	Short Chain Fatty Acid
ERS	Endoplasmic Reticulum Stress
CARD3	Caspase Activation and Recruitment Domain 3
H_2S	Hydrogen Sulfide
CDI	C. difficile Infection
C2	Acetic Acid
C3	Propionic Acid
C4	Butyric Acid
iC4	Isobutyric Acid
C5	Valerate Acid
iC5	Isovalerate Acid
C6	Caproic Acid
iC6	Iso-caproic Acid
C7	Heptanoate Acid
C8	Octanoate Acid

CHAPTER 1

INTRODUCTION

1.1 Background

Crohn's Disease is a type of inflammatory bowel disease (IBD), a chronic, non-specific inflammatory disease that affects the entire gastrointestinal tract and is characterized by repeated remission and recurrence. The symptoms of Crohn's disease are listed below, abdominal distension, abdominal pain, diarrhea (if the inflammation of gut is severe it will be blood), weight loss and so on. People with Crohn's disease have a higher risk of colon cancer and small bowel cancer. However, the cause of Crohn's disease is unclear, but research suggests it may be caused by a combination of genetics, individual environment, immune system and bacteria in the body. Crohn's disease may lead to chronic inflammatory diseases. The human body itself has a well-established immune system to protect the gut, but it may target some special microorganisms for antigens. While Crohn's disease is known to be an immune system, it can currently be inferred that it is not an autoimmune disease because autoimmune disease is not punished by the structure of the body itself, but many questions about immunity are currently It is unclear and further research is needed. Crohn's disease has an above-average risk associated with its own genes, and 70 genes have been found to be related to him. The study also found that smokers are twice as likely to develop Crohn's disease as non-smokers. There is also a trend that it often occurs after gastroenteritis. Methods used to detect Crohn's disease include biopsy, appearance of bowel wall, medical imaging and disease description.
bowel syndrome (IBS) and Behcet's disease also have symptoms similar to Crohn's disease.

About a quarter of people with Crohn's disease are diagnosed when they are in their childhood. Till now, to treat Crohn's disease, therapy is mainly conducted to induce and maintain remission and prevent relapse. During the treatment of Crohn's Disease, Hormone is used mainly to cure it. It is observed that the Crohn's disease symptoms in children are more extensive and more severe than adults (Turunen P et al., 2009) [1]. The disease itself and the long-term use of hormones often lead to complications such as severe malnutrition, growth retardation, and decreased bone density in children. Therefore, children with Crohn's disease are not recommended to take hormones as the first choice for induction of remission (Swaminath A et al., 2017) [2]. Nutritional therapy has received increasing attention to treat the childhood Crohn's disease because it cannot only induce disease remission but also improve children's nutritional status (Baert F et al., 2010) [3]. Currently, Corticosteroids can quickly improve symptoms in a short time, while another drug, such as methotrexate or azurine, can be used to prevent relapse. Multiple analyses have shown that exclusive enteral nutrition is as effective as steroids, another traditional treatment for CD, in treating children with Crohn's disease, with fewer side effects. Though, Hormone is an important drug to treat active Crohn's disease (Ruemmele FM et al., 2014) [4]. It is not the first choice for induction of remission in CD children (Marc A.Sidler et al.,

2008) [5]. So, EEN is getting more and more popular and is recommended as a first-line regimen for induction of remission in children with active Crohn's disease (Grover Z et al., 2016) [6].

About 99.9% of the microorganisms in the human digestive tract are bacteria, and the rest are fungi, archaea, viruses, etc. The relationship between intestinal flora and the development of intestinal inflammation is one of the hotspots of current research. *Firmicutes*, *proteobacteria*, *bacteroidetes*, and *actinobacteria* are the main parts of the gut microbial in the Crohn's disease patients according to the analysis of phylum level, whereas healthy group mainly consists of firmicutes, bacteroidetes, proteobacteria, and actinobacteria. In the study at the genus level, *Escherichia, Shigella, bacteroides, veillonella, lachnospiraceae, enterococcus* made up the majority of the gut microbial in the Crohn's disease group, while *bacteroides, faecalibacteria, prevotella, and roseburia* made up the majority in the control group. Samples are subjects who with Crohn's disease and healthy controls had significantly different intestine microecological compositions at the phylum and genus levels. When compared to a healthy population, Crohn's disease patients' gut microbiota is less diverse, and their gut's microecological makeup is more variable.

1.2 EEN (exclusive enteral nutrition)

The most recommended therapy for treating childhood Crohn's disease is Exclusive Enteral Nutrition (EEN) . The purpose of the formula-based (no solid meals) diet is to help patients go into remission. It is a brief program that could last six to twelve weeks. A nasogastric (NG) tube is put through your child's nasal and into their stomach to deliver exclusive enteral nourishment. Typically, therapy lasts six to twelve weeks and then usual diet is gradually reintroduced. In CD children, EEN has proven to be a practical and successful alternative to CS (corticosteroid). EEN offers extra advantages over and beyond those offered by CS, in addition to avoiding the negative effects of CS exposure. After treatment, EEN therapy is linked to improved quality of life, earlier rises in insulin-like growth factor 1, altered intestinal flora, higher rates of mucosal healing, increased weight gain, and better vitamin D status. There aren't many long-term studies after EEN, but the ones that have been done on kids suggest that EEN may shorten the period before recurrence. Once illness remission is attained, it has been demonstrated that the provision of supplemental enteral nourishment is preferable for maintaining remission in Japanese patients compared to a free diet.

EEN has proven to be a practical and successful alternative to CS. EEN offers extra advantages over and beyond those offered by CS, in addition to avoiding the negative effects of CS exposure. After treatment, EEN therapy is linked to improved quality of life, earlier rises in insulin-like growth factor 1, altered intestinal flora, higher rates of mucosal healing, increased weight gain, and better vitamin D status. There aren't many long-term studies after EEN, but the ones that have been done on kids suggest that EEN may shorten the period before recurrence. Once illness remission is attained, it has been demonstrated that the provision of supplemental enteral nourishment is preferable for maintaining remission in Japanese patients compared to a free diet.

1.3 Prior research

Prior research has mainly paid attention to small number of microbial species and bacterial metabolites. To examine the gut microbiota in CD patients, the prior research mainly uses primers and probes. Sequencing technology developed a lot during several years, so it enables research to evaluate the dynamics of the entire microbial population of patients without any hypothesis. We collected fecal samples from CD children and healthy people without any intestinal disease. We can then describe the microbiome's structural and genetic functional capabilities through large-scale parallel ultra-deep sequencing and cutting-edge computational methods. To achieve this, we targeted both the RNA genes of bacterial 16S and shotgun metagenomics (Quince C, Ijaz UZ et al., 2015) [7].

1.4 Aims and Objectives

At the same time, the analysis methods used by previous studies were also limited, and they focused mainly on whether there being relationship among the overall gut microbiota, EEN and CD. Looking at the overall impact of EEN on the gut and the patient's condition, some more detailed definitions of the microbiome could be added to the new study. At the same time, in the process of EEN, the impact of specific bacteria on the intestinal tract also needs to be paid attention to, because studies have shown that the beneficial effects on the intestinal tract are often caused by some specific bacteria groups. Similarly, some specific bacteria groups can cause adverse reactions such as intestinal inflammation, indigestion, etc. Therefore, this study will pay additional attention to the dynamic changes of beneficial and harmful flora during the EEN process, and further analyze the significant flora.

CHAPTER 2
METHODOLOGY

The report is based on the data of a previous research. The research collected fecal sample from 23 children who have Crohn's Disease. Among these 23 children, 15 of them are taking an 8-months EEN (Modulen, Nestle, Vevey, Switzerland) [8] with at least two consecutive samples. For each patient, the research collects at most five serial samples during different times. Sample A were collected before they receive EEN treatment or within 6 days of the treatment. When they are taking EEN sample B and sample C are collected (B: 1 ~ 16 & C: 16 ~ 32 days). The time of collecting sample D is closed to end of treatment (D: 32 ~ 54 days). Sample E was the last one and was collected when CD children finished EEN and returned to their habitual diet. (E: 54 ~ 63 days after EEN) Also, there are a group of controls which are 21 healthy children having no family history of inflammatory bowel disease. The experiment collects the two fecal samples from each control and make sure that the samples are at least 2 months apart. Simultaneously, the research contains no participant who had received antibiotics for at least 3 months before they attend the study. After the treatment, All these disease activity indices significantly decreased, and 62% of participants achieved clinical remission (PCDAI <10) at the end of EEN

2.1 Theory

It is essential to evaluate the genetic potential of gut microbes in order to comprehend how they affect human health and wellbeing. So, the research collected 576.7 Gb of material from fecal and demonstrated the sequencing, assemble, and characterization of 3.3 million non-redundant microbial genes by using an Illumina-based metagenomics approach (Junjie Qin et al., 2010) [8]. The gene set contains a vast majority of the most common microbial genes in the cohort. It also may contain quite a lot of the most common human intestinal microbial genes. Compared with the human gene complement, it is approximately 150 times larger. The cohort's members mostly share the same genes (Joyce K. Y. Tse, 2017) [9]. The results show that more than 99% of the genes are tend to be bacterial. This means that the whole cohort's number of common bacterial species ranging from 1000 to 1150. At the same time, each person has at least 160 of these species, many of which are also widely shared. To encode the gene set, the research defined and characterized the minimal gut metagenome and the minimal gut bacterial genome. Basing on the gene set, we can have a deep insight of the composition of human gut flora. The research provides great help for our study on several specific bacteria.

EEN has a long history of use in the treatment of pediatric CD, both as an induction therapy for those who have just received a diagnosis and as maintenance therapy for those with the condition. In meta-analyses, EEN has been demonstrated to be equally effective as corticosteroids in pediatrics with the added advantages of enhancing nutritional status and avoiding steroid-related side effects. After completing a first course of EEN, we have demonstrated a significant improvement in weight and BMI

that persisted through the end of the 2-year follow-up. We've shown that children who received a second round of EEN experienced a comparable rate of clinical response and a rise in weight, although this was noticeably smaller than it was after the first round of EEN. Additionally, the first four weeks of treatment were when weight and inflammatory indicators improved the most. Furthermore, genetic connections between host pathways and Crohn's disease (CD) suggest that abnormal immunological reactions to the gut flora have a subordinate role. However, it has been claimed that, even the comparison made between two IBD patients, there still have a lot of variation. We examined the microbiome in the biggest paediatric CD cohort to date using samples from several gastrointestinal sites taken before treatment in new-onset cases. The presence of a disease is significantly correlated with several variables. If there is an increase in the abundance of the bacteria *Enterobacteriaceae*, *Pasteurellaceae*, *Veillonellaceae*, and *Fusobacteriaceae*, also if there is the prevalence of *Erysipelotrichales*, *Bacteroidales*, and *Clostridiales* decreased. The disease will be affected significantly (Sushila R. Dalal et al., 2014) [10]. A comparison of the microbiomes of CD patients who had received antibiotics and those who had not shown that antibiotic usage exacerbates the microbial dysbiosis related to CD. According to a comparison of the microbial signatures between the ileum, rectum, and fecal samples, to detect the CD, especially in early stage of the disease, examining the rectal mucosa-associated microbiome can be potentially quick ways. Also, in children with active Crohn's disease (CD), EEN has a anti-inflammatory impact, which has been tested well, that is caused by modulating gut microbiota activity, which produces an anti-inflammatory SCFA pattern. In contrast, after receiving EEN medication, there have no kids with perianal illness displayed clinical or biochemical improvement (Bo Tjellström, 2012) [11].

Microbes living in the human gut confer additional metabolic capacity on their hosts and control the expression of genes involved in lipid and carbohydrate metabolism (Lola Rahib et al., 2007) [12], which affect nutrient availability, energy balance and weight. In addition, the gut microbiome plays a crucial role in the normal maturation of the immune system, helping to prevent infections and abnormal immunological reactions. As a result, the microbiome living in the digestive tract plays a variety of physiological roles in the human body, regulating immune and metabolic processes, especially when they are disrupted. Increases the intake of food into energy, by intestinal secretion of peptides, controls appetite, controls disorganized composition of fatty acids and absorption, storage and oxidation by secretion of glucagone-like noladin-2 regulates the intestinal barrier, activates congenital immunity, as well as through lipopolysaccharides toll-like receptor-4-axis liver fibrosis, these intestinal microbiotas connect with the host's energy exchange some mechanisms. Antibiotic, prebiotic, and probiotic modification of the gut microbiota has produced positive outcomes in the management of diabetes and obesity. Additionally, the results examining the development of IBD and CD indirectly suggest that endogenous or environmental components should be described when tackling multifactorial disorders instead of taking the bacterial profile into thoughts alone (Giovanni Musso et al., 2010) [13].

2.2 Primary analysis of genomes

2.2.1 16S rRNA gene sequencing

Commonly used amplicon ranking methods is used to identify and compare bacteria or fungi in a given sample, including 16S and ITS ribosomal RNA (rRNA) grading. Sequencing ITS and 16S rRNA genes using next-generation sequencing (NGS) is a well-established method to compare phylogeny and classification of samples from complex microbes or environments that are difficult or impossible to study.

The prokaryotic 16S rRNA gene is approximately 1500bps long and has nine variable regions between conserved regions. Variable regions of 16S rRNA gene are commonly used for the phylogenetic classification of genera or species in different microbial populations. The its1 area of rRNA ciscodon is a common DNA marker used in metagenomic samples to identify fungal species. (Identify and compare microbes from complex microbiomes or environments, 2022) [14]

2.2.2 Shotgun metagenome sequencing

Shotgun metagenomic ranking allows researchers to extract all the genes of all living things from a complex sample. Microbiologists can use this method to evaluate bacterial diversity and detect the abundance of microbes in different environments. Shotgun metagenomics can also be used to study microbes that cannot be grown that are otherwise difficult or impossible to do research.

Unlike capillary sequencing or PCR-based methods, next generation sequencing (NGS) allows researchers to rank hundreds of species at the same time. NGS-based metagenomic ranking may detect low abuntions of members of potential missing detection in microbiome, or cannot be identified because the use of other methods is too expensive, as it can combine several samples in one ranking and produce high sequence coverage for each sample. (Sequence thousands of organisms in parallel, 2022) [15]

2.3 Visualization of the results

Due to the huge amount of data, it is necessary to clearly display the data results through multiple visualization tools. The results are run by R vegan package and visualized by the plot tools in R language.

2.3.1 R Studio

All of the results were operated on the R Studio platform. R Studio is an integrated development environment (IDE) for R. It includes different functions for solving problems in various fields. It also supports direct code execution. At the same time, there are many built-in tools for plotting, history, debugging and workspace management (R Studio, Take control of your R code, 2022) [16]. R is a language specially developed for statistics and data analysis.

All the plots of the results are produced via R Studio using R packages like vegan, ggplot2 and so on. Vegan package provides tools for descriptive community ecology. It has most basic functions of diversity analysis, community ordination and dissimilarity

analysis. Vegan kits provide tools to describe community ecology. Its most basic functions are diversity analysis, community ordination and dissimilarity analysis (Jari Oksanen, 2022) [17]. Graphic syntax is the basis of the declarative graphic creation system called ggplot2. You give ggplot 2 data, instruct it how to map variables to aesthetics and what graphical primitives to employ, and then handle the rest. (ggplot2 3.35, 2022) [18]

2.3.2 Sample Analysis

Microbiota composition was described at multiple resolutions using "oligarchic" tasks of genus, operational taxonomic unit (OTU) and the 16S rRNA sequencing, the latter two of which were close to species level. Before EEN, we used the Adonis function in the R-package to determine the total amount and significance of differences between control and CD. For the 16S rRNA markers (genus, OTU and oligonucleotype), only markers with mean relative abundance greater than0.01% were tested. We compared control and CD children in a similar way to test for changes between 16S RNA and metagenome-labeled samples D (end of EEN) and E (habitual diet). For CD subjects with two or more samples who were still receiving or had just stopped EEN treatment in vitro ($\leqslant$3 days after treatment), the GLM function was used to recapitulate individual marker data relative to the number of ejaculation days in vitro for specific subjects. Therefore, any confounding effect of different sampling times is naturally taken into account in the analysis, especially after the activation and completion of the European nuclear material.

2.3.3 Specific analysis

There has a very powerful method through which we can do community ecology. We can apply it to abundance table whether it is abundance of microbes, if we have their phylogeny also available, then it will be able to figure out whether there are deterministic (homogeneous or variable selection) forces there, or stochastic (homogenizing dispersal, dispersal limitation, or undominated process)

1) Alpha diversity analysis is used to see the changes of microbe's abundance between different samples.

2) Some null model approaches are also applied. Affairs-based (Raup-Crick) β-diversity (βRC) tests stochasticity and certainty using indicators cited by (Chase et al. 2011) [19]. QPE (Quantitative Process Estimates) assembly processes involving phylogeny and β-diversity based on abundance (Raup-Crick) using the ways quoted by (Stegan et al, 2013) [20]. NTI (nearest-taxon-index) has been operated to figure out the problem of environmental filtering (phylogenetic overdispersion versus clustering). Nearest taxa index will calculate phylogenetic distances within each sample.

3) Differential analysis will take any two conditions at a type, and figure out what are the specific genera that are "differentially expressed".

4) By evaluating every conceivable combination of the explanatory factors, we did subset regression against various microbiome metrics, and then it can select the best model based on specific statistical parameters, with the recommendations given in

5) Any group of microbial taxa, or the genetic and functional traits linked to those taxa, that are exclusive to a host or environment of interest are referred to as the core microbiome. Most frequently, the microbial taxa shared by two or more samples from the same host or environment are used to quantify core microbiomes. (Alexander T. Neu et al., 2021) [21] To find the core microbiomes, R's microbiome package can be used. (Lahti, L., & Shetty, S., 2017) [22]

CHAPTER 3
RESULTS

The results will split into four different part, they are Diversity estimates, Null modeling approaches, Differential analysis, and the regression modeling separately. Using visuals like figures and charts, the results are explained for each unique genus in each section. Specifically, to analyze the microbial diversity, the research conducts the Alpha diversity and Beta diversity estimates. In next part, Nearest Taxon Index, Normalized Stochasticity Ratio and Quantitative Process Estimates make up the Null modeling approaches. Then, the core microbiome heat map and the DESeq2 can help find the difference between samples. Finally, the regression modeling consists of Subset regression and CODA GLMNET.

3.1 Diversity estimate

Alpha and beta diversity can be estimated under the guidance and help of R vegan packages. The mean species diversity of a site within a specific range is known as alpha diversity in ecology. Although Shannon entropy gives equal importance to every species, if we want to emphasis more on the “abundant species” then we use Simpson formula (The Use and Types of Alpha-Diversity Metrics in Microbial NGS, 2022) [23] Two factors, the average species diversity (alpha diversity) of local-scale sites and differentiations between these sites, together determine total species diversity in the landscape (beta diversity).

The mean Shannon diversity, a measure of summarizing a single sample in terms of diversity was higher in controls (Sample H) than CD children before they took EEN (Sample A). From Figure 1, we can see a clear trend that, when children take EEN, their microbiota richness will drop. Then they will get a bit higher compared with Group A.

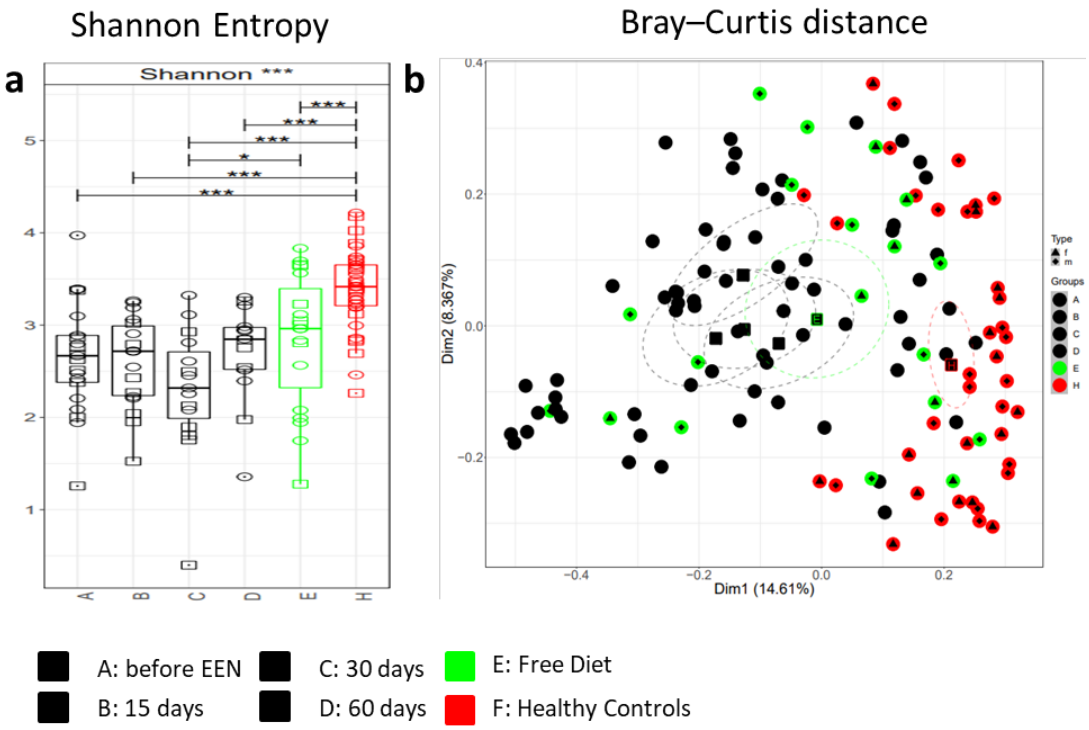

Figure 1: Genus Shannon diversity in richness equivalents among Groups A, B, C, D, E and F (a), and beta diversity of operational taxonomic unit (OTU) community structures for CD children during their treatment and healthy controls (b). (a) Diversity fluctuates and shows a decreased trend during EEN treatment. Furthermore diversity was higher in healthy controls versus CD children. (If P value < 0.01, the different groups will be linked above and marked with two stars. If P value < 0.05, the different groups will be linked above and marked with one star. If P value < 0.001, the results will be linked and marked with three stars.) (b) CD children's community structure was much more different from Healthy controls.

Unweighted Unifrac, a phylogenetic distance metric, is a β-diversity measure that uses phylogenetic information to compare targeted samples. It determines the distance between samples without taking into account their abundances by dividing the sum of unshared branch. It has also been applied in more than 150 research publications that seek to understand the relationship between microbial communities in systems ranging from human disease to ecology in general. To calculate the Unifrac distances, R's phyloseq package can be implemented. (McMurdie and Holmes, 2013) [24].

Furthermore, from Figure 1 we can see that, beta diversity, showing whether the OTU community structure between samples is similar, was closer for CD children than healthy controls. But it worth paying attention to that when CD children are taking treatment their diversity was getting much more different from healthy controls. However, it is clearly shown that CD children's OTU community structure, compared with before taking EEN, were getting much closer to healthy controls.

Figure 2 shows the top genera exist in both Crohn's disease (CD) children in regard to the relative abundance. Mostly, genera belonging to the *Actinobacteriota* and *Firmicutes* were more abundant in controls for example, there are *Bifidobacterium*, *Veillonellaceae* and *Streptococcus*. In contrast, *[Ruminococcus]_gnavus_group* which belongs to *Firmicutes* are more abundant in CD children. Also, genera like *Blautia* and *Dorea* are both prevalent in CD children and healthy controls. Compared with CD children, healthy children's genera are much more abundant. EEN shows a positive impact on CD children's genera diversity. The color of the taxa bars was getting richer and richer.

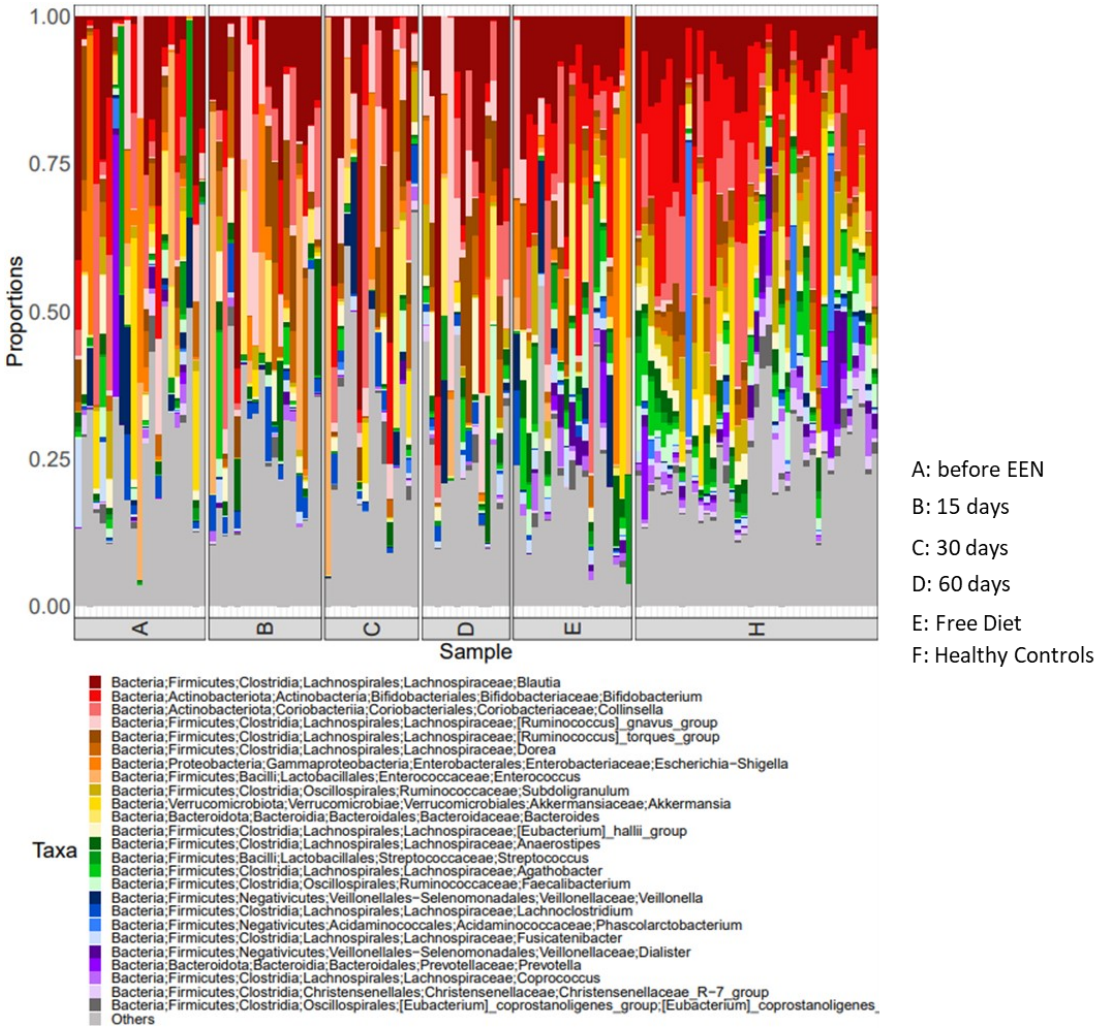

Figure 2: Top 26 bacterial genera which is more abundant in relative abundance for CD children in their EEN treatment versus healthy controls.

3.2 Null model approaches

3.2.1 NTI (nearest-taxon-index)

This approach can effectively show whether the status of OTU community structure is stochastic (ecological drift driven by random birth-death events and random colonisation) or deterministic (clustering and driven by strong environmental varialbes like pH, temperature, calprotectin and so on). NTI can be calculated by using mntd() and ses.mntd() functions which can be acquired from picante package (Kembel et al.,

2010) [25]. First, get the output from ses.mntd(). And then, NTI is the negatives of the output above. The standard deviation between the observed MNTD and the zero distribution mean is quantified by NTI (999 randomizations). An NTI greater than +2 in a single community indicates a closer relationship between coexisting taxa than predicted by chance (phylogenetic clustering). Coexisting taxa are more distantly related than would be expected by chance if NTI was less than -2. (Phylogenetic overdispersion). In general, clustering (NT>0) or over-dispersion (NT<0) is the average NTI for all communities, which is significantly different from zero.

NTI shows a deterministic environment in CD children and a stochastic environment in controls. (Figure 3) It is more deterministic in healthy people. The P value shows that there are several significant changes happened during CD children's EEN treatment. Especially between 30 days and 60 days, the NTI fluctuates a lot (P value<0.01). But, there is almost no difference between CD children and healthy controls. This means EEN may influence the CD children's community structure significantly. It should be also paid attention that the results of nearest taxon index opposite to the Quantitative Process Estimates (QPE) procedure.

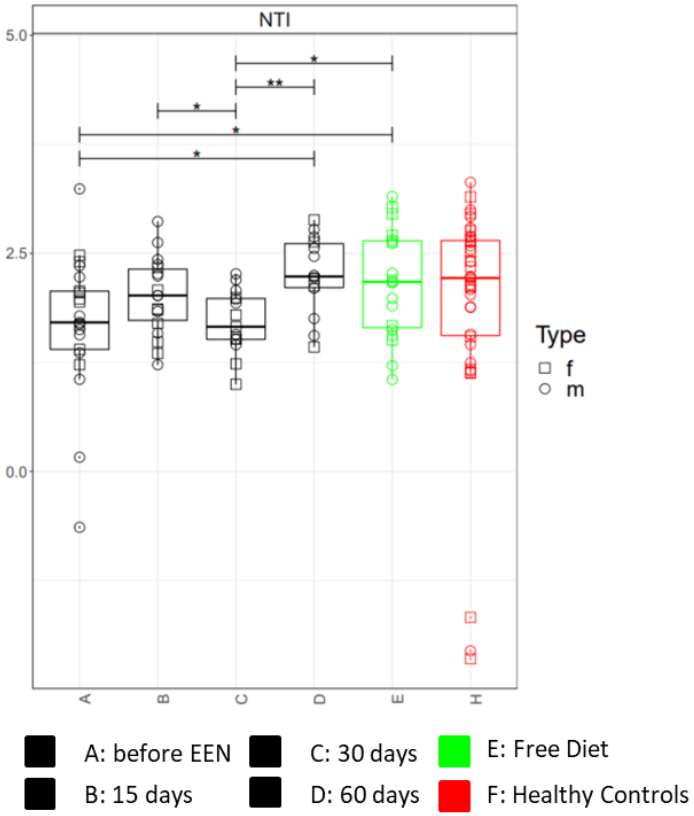

Figure 3: Weighted nearest taxon index (NTI) for CD children and healthy controls. CD children changed significantly during their EEN treatment. If P value < 0.01, the correlated two results will be linked together and marked by two stars. If P value < 0.05, these results will be linked and marked by one star. If P value < 0.001, the results will be linked and marked by three stars.

3.2.2 NST (Normalized Stochasticity Ratio)

The 50 percent line that separates more deterministic (50 percent) assembly from an index called normalized stochasticity ratio (NST) was devised. By taking into account abiotic filtration, competition, ambient noise, and spatial scales, NST was put to the test with fictitious communities. At large spatial scales or when subjected to extremely loud environmental noise, all studied techniques had limited performance. In contrast, NST demonstrated higher accuracy (0.90 to 1.00) and accuracy (0.91 to 0.99), with average accuracy (0.1 to 0.7) and accuracy (0.33 to 1.8) being 0.37 higher than previous methods (Daliang Ning et al., 2019) [26].

NST elucidates healthy people have a stochastic assembly process (59%). In contrast, CD children have a deterministic assembly process. We can observe a trend that, during EEN treatment and after returning to their habitual diet, the community structure was becoming more and more stochastic. Especially, they will be more stochastic when CD children finish the EEN treatment.

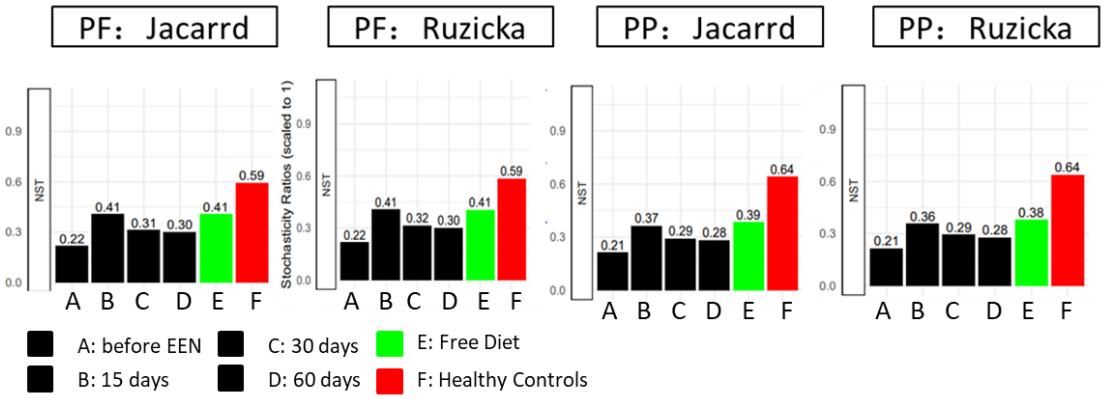

Figure 4: Normalized Stochasticity Ratio (NST) for CD children and healthy controls. When “abundance.weighted” = TRUE, “Jacarrd” can be called as “Ruzicka”. PF and PP are two variables, the first variable is the way to control taxa occurrence frequency and the second variable is the way to control richness in each sample. So “PP” means “proportional & proportional”. "Proportionality" refers to the probability of a taxon occurring in direct proportion to the frequency of occurrence observed. The CD children show a trend of getting more stochastic during EEN.

3.2.3 QPE Procedure (Quantitative Process Estimates)

The integration of phylogenetic information into null model techniques is based on the idea that phylogenetic relatedness is a sign of shared environmental response qualities. Empirical modelling techniques based on phylogenetic and abundance (Raup-Crick) measurements of β-diversity (RCbray) were combined to quantify the relative effects of processes such as selection, drift, dispersion limitation, and mass effects (Stegen et al., 2013) [27]. Additionally, this quantitative process estimate (QPE) method shows the results whether the environment is homogeneous and heterogeneous selection processes, which both have an adverse effect on beta diversity. The results of QPE can be divided into several parts. Selection pressure can show whether the microbial samples have an underlying signature of a single environment or multiple environments in the absence of any recorded environmental

parameter like PH, temperature and Calprotectin. The microbial community structure is influenced by environment, so the results can illustrate what is the % of homogeneous selection and what is the % of variable selection. Then Homogeneous dispersal microbes from one sample are continuously appearing in other samples. Furthermore, if the results show dispersal limitation, it means there is no dispersal and "Undominated" means we cannot explain it (Máté Vass et al., 2020) [28].
From Figure 5, CD children shows the percentage of dispersal limitation dropping from 69.52% to 59.14%, compared with healthy children (29.96% of dispersal limitation) when they are taking the treatment. Healthy children are undominated and the percentage is 41.97%. Also there is 27.53% of variable selection. This means that, for CD children, the dispersal of the microbiome in their gut are limited by some specific factors.

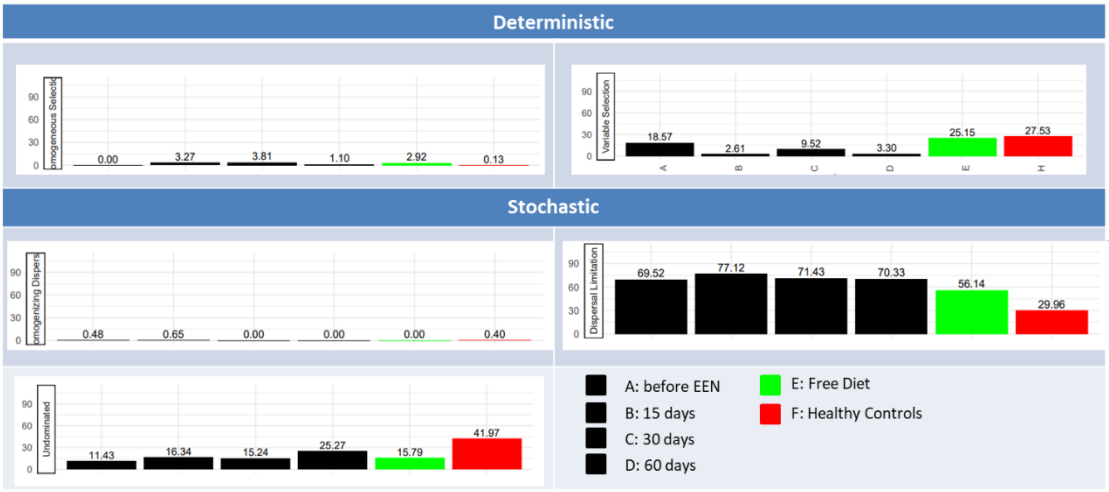

Figure 5: Quantitative Process Estimates (QPE) Procedure for CD children and healthy controls. CD children were getting less stochastic during EEN.

3.3 Differential Analysis

It can be find that the community structure differed significantly between CD children and Healthy controls. It is necessary to check if there have any different genera between different categories or samples. The DESeqDataSetFromMatrix() method from the DESeq2 (Love et al., 2014) [29] package was used to settle the above problem, and the log2 fold change cut-off was set to 2 . Also, the adjusted p-value significance threshold are set to 0.05. In order to estimate maximum probability of OTU log fold change between two conditions, the function applies a negative binomial GLM.
Following the application of Bayesian shrinkage to generate shrunken log fold changes, the Wald test is used to determine significance. Changes were detected locally by DESeq2 (in combination with diversity analysis) to identify genera driving microbial community shifts, while MINT provided discriminating information shared in stratified models globally.
Table 1 shows the most abundant microbiome in both CD children (before taking EEN) and healthy controls separately. It can be found that *Dialister* (P value adjusted = 0.008276), *Parabacteroides* (P value adjusted = 0.006915), *Slackia* (P value adjusted =

0.00301) and *Parasutterella* (P value adjusted = 0.000756) are much more abundant in healthy controls compared with CD children who didn't take EEN treatment. In contrast, *Eisenbergiella* (P value adjusted = 4.05E-05), *Lactonifactor* (P value adjusted = 1.86E-05) are more prevalent in CD's body. Then the sample of CD children who had returned to their habitual diet was compared with healthy children. It can be found that *Parabacteroides, Holdemanella, Slackia and Enterorhabdus* are more abundant and *Tyzzerella, Erysipelatoclostridium* and *Sellimonas* are more abundant in CD children. This means EEN help CD children decrease the abundance of *Eisenbergiella*, *Lactonifactor* and other microbiomes and increase the abundance of *Dialister*, *Parasutterella* and other microbiomes.

Table 1 : Community structure of bacterial genera which are significantly different in the healthy control group and CD children before taking EEN.

Genera	**P value adjusted**	**Sample**
Bacteria; Firmicutes; Negativicutes; Veillonellales-Selenomonadales; Veillonellaceae; Dialister	0.008276	H
Bacteria; Bacteroidota; Bacteroidia; Bacteroidales; Tannerellaceae; Parabacteroides	0.006915	H
Bacteria; Actinobacteriota; Coriobacteriia; Coriobacteriales; Eggerthellaceae; Slackia	0.00301	H
Bacteria; Proteobacteria; Gammaproteobacteria; Burkholderiales; Sutterellaceae; Parasutterella	0.000756	H
Bacteria; Firmicutes; Clostridia; Oscillospirales; Ruminococcaceae; CAG-352	0.000579	H
Bacteria; Actinobacteriota; Actinobacteria; Actinomycetales; Actinomycetaceae; Varibaculum	0.000438	H
Bacteria; Firmicutes; Clostridia; Oscillospirales; Ruminococcaceae; Subdoligranulum	0.000411	H
Bacteria; Firmicutes; Bacilli; Erysipelotrichales; Erysipelotrichaceae; Turicibacter	0.000327	H
Bacteria; Firmicutes; Clostridia; Monoglobales; Monoglobaceae; Monoglobus	0.000314	H
Bacteria; Bacteroidota; Bacteroidia; Bacteroidales; Prevotellaceae; Prevotella	0.000314	H
Bacteria; Firmicutes; Clostridia; Lachnospirales; Lachnospiraceae; Lachnospiraceae_UCG-010	0.000269	H
Bacteria; Desulfobacterota;Desulfovibrionia;Desulfovibrionales;Desulfovibrionaceae;Bilophila	0.000269	H
Bacteria; Actinobacteriota;Coriobacteriia;Coriobacteriales;Eggerthellaceae;Enterorhabdus	0.000258	H
Bacteria; Actinobacteriota;Coriobacteriia;Coriobacteriales;Coriobacteriaceae;Enorma	0.000206	H
Bacteria; Firmicutes;Bacilli;Lactobacillales;Lactobacillaceae;Lactobacillus	0.000187	H
Bacteria; Firmicutes;Clostridia;Oscillospirales;Ruminococcaceae;Faecalibacterium	7.73E-05	H
Bacteria; Firmicutes; Negativicutes; Veillonellales-Selenomonadales; Selenomonadaceae; Megamonas	7.69E-05	H
Bacteria; Firmicutes; Clostridia; Oscillospirales; Oscillospiraceae; NK4A214_group	6.60E-05	H

Bacteria; Actinobacteriota; Coriobacteriia; Coriobacteriales; Eggerthellaceae; Adlercreutzia	5.23E-05	H
Bacteria; Firmicutes; Clostridia; Oscillospirales; UCG-010;UCG-010	5.16E-05	H
Bacteria; Firmicutes; Clostridia; Lachnospirales; Lachnospiraceae; Eisenbergiella	4.05E-05	A
Bacteria; Firmicutes; Clostridia; Lachnospirales; Lachnospiraceae; [Eubacterium]_ruminantium_group	2.80E-05	H
Bacteria; Firmicutes; Clostridia; Lachnospirales; Lachnospiraceae; Lactonifactor	1.86E-05	A
Bacteria; Firmicutes; Clostridia; Lachnospirales; Lachnospiraceae; Lachnospiraceae_UCG-001	1.67E-05	H
Bacteria; Firmicutes; Clostridia; Peptostreptococcales-Tissierellales; Peptostreptococcaceae; Terrisporobacter	1.41E-05	H
Bacteria; Firmicutes; Clostridia; Lachnospirales; Lachnospiraceae; Frisingicoccus	1.18E-05	H
Bacteria; Actinobacteriota; Coriobacteriia; Coriobacteriales; Eggerthellaceae; Eggerthella	9.30E-06	A
Bacteria; Firmicutes; Clostridia; Lachnospirales; Lachnospiraceae; Moryella	5.97E-06	H
Bacteria; Firmicutes; Clostridia; Oscillospirales; Ruminococcaceae; Incertae_Sedis	5.97E-06	H

Table 1: Children with Crohn's disease before taking EEN contribute to Group A. Healthy controls contribute to Group H.

Table 2: Community structure of bacterial genera which are significantly different in the children with Crohn's after EEN and CD children before EEN treatment.

Genera	**P value adjusted**	**Sample**
Bacteria; Bacteroidota; Bacteroidia; Bacteroidales; Tannerellaceae; Parabacteroides	0.011884	H
Bacteria; Firmicutes; Bacilli; Erysipelotrichales; Erysipelotrichaceae; Holdemanella	0.011088	H
Bacteria; Actinobacteriota; Coriobacteriia; Coriobacteriales; Eggerthellaceae; Slackia	0.00827	H
Bacteria; Actinobacteriota; Coriobacteriia; Coriobacteriales; Eggerthellaceae; Enterorhabdus	0.005293	H
Bacteria; Firmicutes; Clostridia; Lachnospirales; Lachnospiraceae; Lachnospira	0.003829	H
Bacteria; Firmicutes; Negativicutes; Acidaminococcales; Acidaminococcaceae; Acidaminococcus	0.00237	H
Bacteria;Firmicutes; Clostridia; Oscillospirales; Ruminococcaceae; CAG-352	0.001316	H
Bacteria; Firmicutes; Clostridia; Lachnospirales; Lachnospiraceae; Tyzzerella	0.001279	E
Bacteria; Firmicutes; Clostridia; Peptostreptococcales-Tissierellales; Anaerovoracaceae; Family_XIII_UCG-001	0.001192	H
Bacteria; Firmicutes; Clostridia; Lachnospirales; Lachnospiraceae; [Eubacterium]_ventriosum_group	0.001192	H
Bacteria; Actinobacteriota; Coriobacteriia; Coriobacteriales; Coriobacteriaceae; Enorma	0.000835	H
Bacteria; Bacteroidota; Bacteroidia; Bacteroidales; Prevotellaceae; Paraprevotella	0.000835	H
Bacteria; Bacteroidota; Bacteroidia; Bacteroidales; Marinifilaceae; Odoribacter	0.000708	H

Bacteria; Proteobacteria; Gammaproteobacteria; Burkholderiales; Sutterellaceae; Sutterella	0.000649	H
Bacteria; Firmicutes; Bacilli; Erysipelotrichales; Erysipelatoclostridiaceae; Erysipelatoclostridium	0.000615	E
Bacteria; Firmicutes; Clostridia; Lachnospirales; Lachnospiraceae; Sellimonas	0.000419	E
Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales; Methanobacteriaceae; Methanobrevibacter	0.000408	H
Bacteria;Firmicutes; Clostridia; Lachnospirales; Lachnospiraceae; Lachnospiraceae_UCG-003	0.000392	H
Bacteria; Bacteroidota; Bacteroidia; Bacteroidales; Barnesiellaceae; Barnesiella	0.000287	H
Bacteria; Firmicutes; Clostridia; Oscillospirales; Ruminococcaceae; Fournierella	0.000287	H
Bacteria; Firmicutes;Clostridia; Peptostreptococcales-Tissierellales; Anaerovoracaceae; Family_XIII_AD3011_group	0.000241	H
Bacteria; Firmicutes; Negativicutes; Veillonellales-Selenomonadales; Selenomonadaceae; Megamonas	0.000191	H
Bacteria; Proteobacteria; Gammaproteobacteria; Burkholderiales; Sutterellaceae; Parasutterella	0.000127	H
Bacteria; Firmicutes; Clostridia; Oscillospirales; Ruminococcaceae; DTU089	0.000127	E
Bacteria; Firmicutes; Clostridia; Lachnospirales; Lachnospiraceae; [Eubacterium]_ruminantium_group	5.53E-05	H
Bacteria; Firmicutes; Clostridia; Lachnospirales; Lachnospiraceae; Hungatella	3.54E-05	E
Bacteria; Firmicutes; Clostridia; Lachnospirales; Lachnospiraceae; Moryella	2.27E-05	H
Bacteria; Actinobacteriota; Actinobacteria; Actinomycetales; Actinomycetaceae; Actinomyces	2.05E-05	E
Bacteria; Firmicutes; Clostridia; Christensenellales; Christensenellaceae; Christensenellaceae_R-7_group	1.89E-05	H

Table 2: Crohn's disease (CD) Children before EEN treatment contributes to Group A. Healthy controls contribute to Group H.

3.4 Subset regression

By evaluating every conceivable combination of the explanatory factors, we did subset regression against several microbiome metrics, and then we chose the best model based on specific statistical criteria, as recommended in (Kassambara et al., 2018) [30]. By specifying the option nvmax (set as the maximum number of predictions to be included in the model), the best models of different sizes were selected using the set of rules for the R function of the jump package (Mattias Nyström et al., 2012) [31].

Cross-validation requires that the data be decomposed into k subsets after obtaining the best available subsets. The remaining subsets (90%) were used for training data, and each subset (10%) was used in turn for test data sets. Then the model prediction error calculated as the average cross validation error. All of this was accomplished using a

custom function that uses the train method for inserting packages in R. (Kuhn, 2008) [32]. To collect statistics for each model, use the Tab Model() function in the sjPlot package. We used recommendations from the R's microbiome package (Lahti et al.., 2017) [33] to determine the core microbiome.

This model investigated the relationship between various parameters such as pH, Calprotectin dry, and various short fatty acid chains and Shannon diversity. Our research explores which of bacteria or acid are correlated with Shannon diversity. The research got the data of calprotectin, pH, Acetic acid (C2), Propionic acid (C3), Butyric acid (C4), Isobutyric acid (iC4), Valerate acid (C5), Isovalerate acid (iC5), Caproic acid (C6), Iso-caproic acid (iC6), Heptanoate acid (C7) and Octanoate acid (C8). Then the subset regression modeling shows 13 different groups. Model 4 was chosen to make deep research which cross-validation errors equal to 0.53277 (Figure 9). As can be seen from the Figure 5, C4 marked with three stars has a fairly significant statistical significance, and there is a positive correlation between it and Shannon diversity, which means that C4 contributes to the intestinal biodiversity of CD children. Increased sexuality deserves further study and discussion. However, it can be seen from the figure that there is no statistical significance for other parameters except C4, which indicates that changes in intestinal biodiversity may not be directly related to these types of substances.

	Shannon								
Predictors	*Estimates*	*std. Error*	*std. Beta*	*standardized std. Error*	*CI*	*standardized CI*	*Statistic*	*p*	*df*
(Intercept)	2.33064 ***	0.14574	-0.00000	0.08228	2.04138 – 2.61989	-0.16331 – 0.16331	15.99157	**6.068e-29**	97.00000
Calprotectin dry	-0.00002	0.00001	-0.14933	0.09364	-0.00003 – 0.00000	-0.33519 – 0.03653	-1.59465	1.140e-01	97.00000
C4	0.00941 ***	0.00199	0.44328	0.09373	0.00546 – 0.01336	0.25724 – 0.62931	4.72907	**7.653e-06**	97.00000
iC5	0.01855	0.00985	0.17645	0.09368	-0.00100 – 0.03809	-0.00947 – 0.36237	1.88363	6.261e-02	97.00000
iC6	-0.03531	0.01920	-0.17864	0.09715	-0.07342 – 0.00280	-0.37146 – 0.01418	-1.83880	6.900e-02	97.00000
Observations	102								
R^2 / R^2 adjusted	0.337 / 0.309								

** p<0.05 ** p<0.01 *** p<0.001*

Figure 6: Correlation between Shannon diversity and variables collected from CD children. Estimates of Calprotectin, C4, iC5 and iC6 represents their β coefficients separately. If P value < 0.01, the estimates of the predictors will be marked by two stars. If P value < 0.05, the estimates of the predictors will be marked by one star. If P value < 0.001, the results will be marked by three stars. If the estimates are minus, it means that the variables do negatively to the Shannon diversity. If the estimates are positive, contrary to above

3.5 CODA GLMNET

In ecological surveys, multivariate abundance data --numerous interacting species (or other taxonomic groups) observed from a set of samples --are often collected to describe a biological community or combination.

The term "abundance" here refers to counting, existence-absence of records, biomass data or any other measure of the likely presence of a species in a given location (Jenni Niku et al.., 2019) [34]. Common ecological problems addressed by these data include determining whether a group of sites has a similar species composition (Bjork, Hui, O'Hara, and Montoya, 2018) [35], discovering interactions between species and

visualizing correlation patterns between species, testing environmental impact assumptions (Lammel et al., 2018) [36], and predicting abundance (Buisson & Grenouillet, 2008) [37]. Consider the logarithmic-contrast model as a backup regression strategy for CoDA.

In order to ensure the scale invariance principle, the regression coefficients (except intercept) have a zero-sum constraint, which is a linear regression model that combines logarithmic transformation variables with results in the form of a logarithmic-contrast function. This model can modify the selection of CoDA variable selection by penalty regression. In the context of microbiome studies, a punitive linear logarithmic regression model optimization method was proposed and extended to generalized linear regression models (Antoni Susin et al.., 2020) [38]. For code, in a hypothesis space, you simply select the rows in source data you are interested in, and then assign a column to the meta-table $Groups, which you want to use to capture all environments to regress according to the environmental covariance (source data column) listed in the environment _ covaries.

From the Figure 7 we can observe that *Atopobium*, belonging to Actinobacteria, is associated with halitosis, and is also have potential relationship with disease activity in pediatric Crohn's disease. *Fusobacterium*, belonging to *Fusobacteriota*, has been known that are correlated with Crohn's disease. They have strong relationship with calprotectin which corresponds to Intestinal inflammation. By contrast, *NK4A214_group* belong to Firmicutes may have some negative effect on calprotectin. So, it means that finding any therapy can promote the *NK4A214_group* can help the patients to decrease the calprotectin. Furthermore, Genera like *Actinomyces* which belongs to *Actinobacteriota* are strongly related with pH in samples' gut. In contrast, other genera belonging to Firmicutes are negative with pH (e.g., *Streptococcus* and *Lachnospiraceae_UCG−004*).

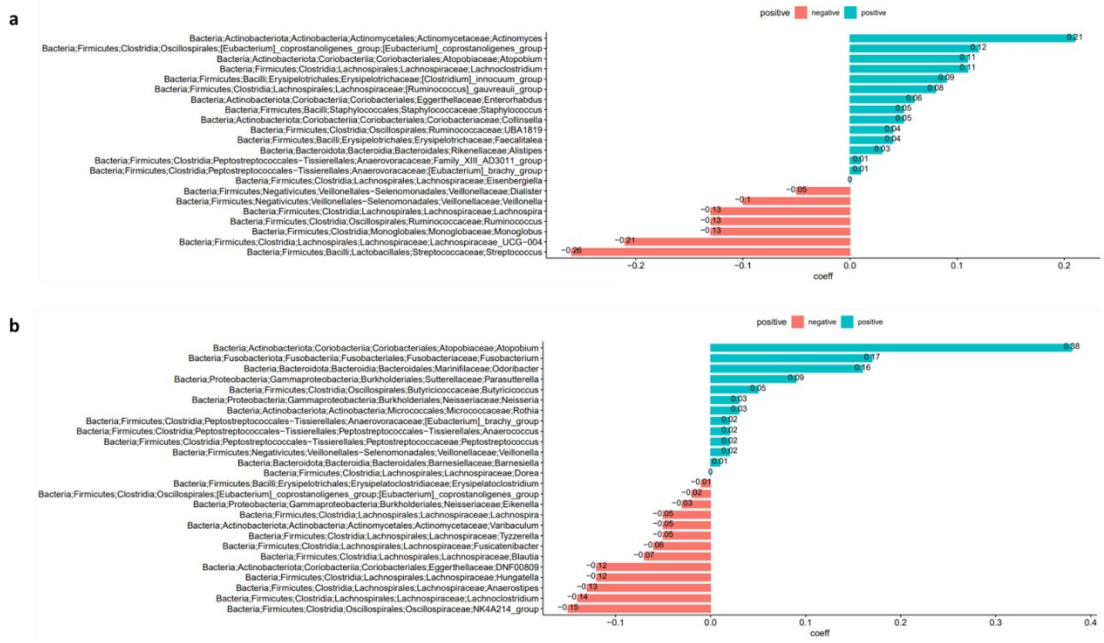

Figure 7: CODA GLMNET results for CD children before taking EEN treatment. Perform variable

selection through penalized regression on the set of all pairwise log-ratios. (a) The top bacteria that have impact on calprotectin, both positive and negative. (b) The top bacteria that have impact on the patients guts' pH, both positive and negative.

3.6 Core Microbiome Analysis

The main purpose of studying and analyzing human core microbiome is to identify and describe various microbial structures, bacterial taxa and their respective functions. From a taxonomic standpoint, the bacterial community exhibits an increasing degree of complexity from the phylum level to the species level. Additionally, gut bacteria engage in interactions with the host and one another at different cellular, population, and community levels. Due to the complexity of this multi-scale interaction, it is difficult to understand community function and its impact on host health (Greenblum, Turnbaugh and Borenstein, 2012; Greenblum et al, 2013) [39] [40].

The HITChip (Human Intestinal Tract Chip) a phylogenetic microarray targeting the 16S rRNA gene sequences of bacterial taxa recorded in the human gut and has been used to address the analytical definition of core microbiota (Salonen et al., 2012) [41]. This phylogenetic microarray is a measurement platform that integrates common and uncommon species into microbiome profiling, with excellent repeatability, robustness, and dynamic range. While these and other phylogenetic microarrays capture the vast majority of gut microbiota, novel taxa can be identified that can be evaluated by next-generation sequencing (Hazen et al., 2010; Roh et al., 2010; Tottey et al., 2013) [41] [42] [43]. Previously, a common core of 290 phylotypes was found in 115 adults who had >1000 rare and abundant bacterial phylotypes from 130 genus-level groups (Jalanka-Tuovinen et al., 2011) [44].

In Figure 8, compared with CD children, the core microbiome types of healthy children were significantly more abundant. From the comparison of group A and group H, it can be seen that the three types of bacteria *Dorea, Blautia and Bifidobacterium* are equally distributed in CD children and healthy children. But in CD children, compared with healthy children, Streptococcus distribution is more common, and its prevalence remains at 1 when Detection Threshold is around 16. *Subdoligranulum, Collinsella* and *Agathobacter* were more common in healthy children. When Detection Threshold is around 40, these bacteria's prevalence remain keeps at 1.

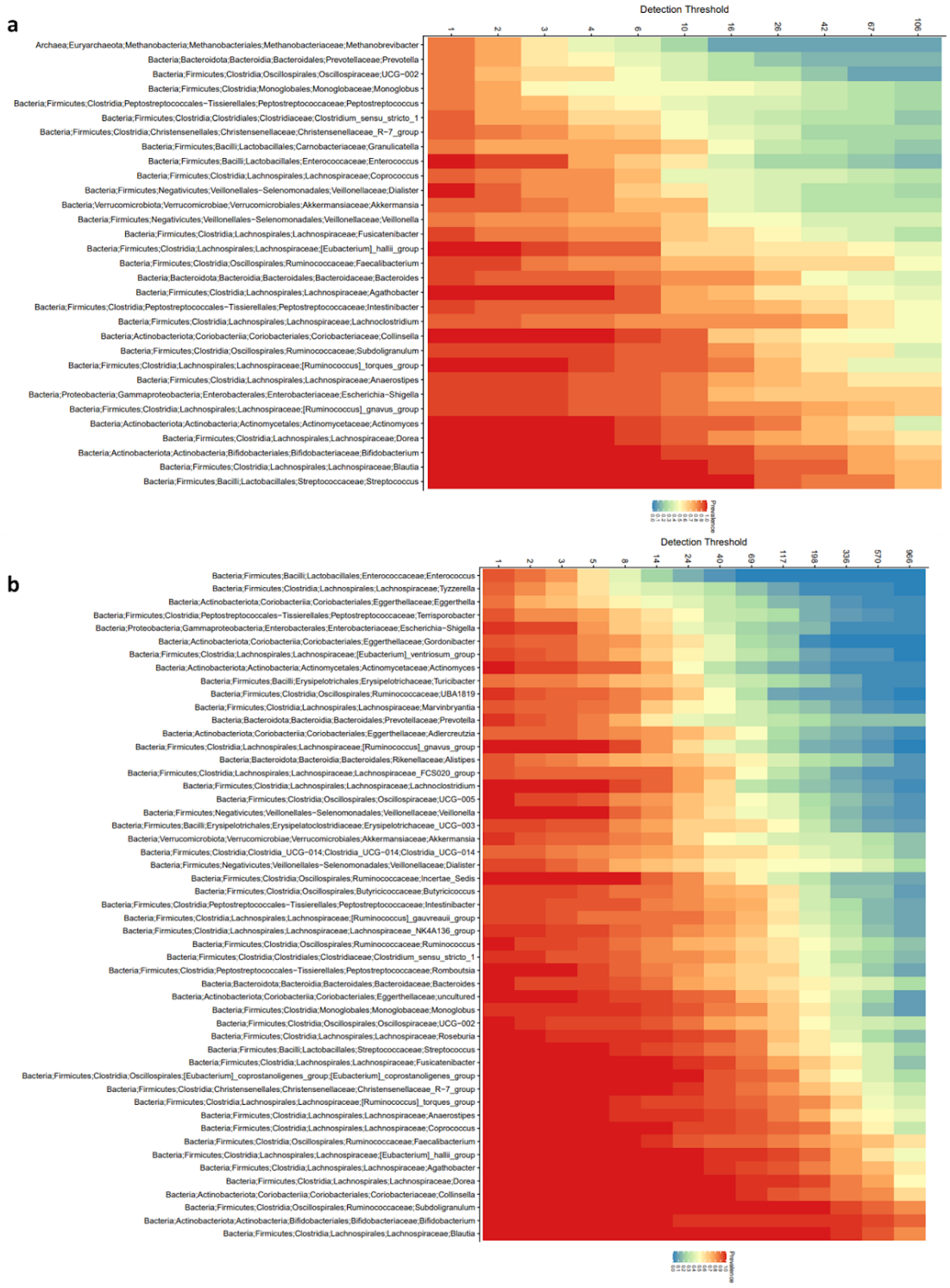

Figure 8: Core microbiome analyses for children with Crohn's disease before EEN and healthy controls. In the heat maps, the OTUs are ranked by their abundances. Those OTUs abundance which is low will be put on the top, whereas those at the bottom are highly abundant.

CHAPTER 4

DISCUSSION

In this study, we detail the effects on the entire fecal microbiota during EEN in children with CD, using the gut microbiota of healthy children as a reference. This study identified various bacterial taxa associated with children with CD, as well as other bacterial taxa that changed significantly during EEN treatment. In exploring the effects of EEN on the gut microbiota, several new biome analysis methods were used for the first time, filling the gaps in previous research methods.

4.1 Variations of diversity between CD children and controls

In this experiment, from the results we found that there were great differences in intestinal biodiversity and tissue structure between children with Crohn's disease and healthy children in the control group. The intestinal biodiversity of children with Crohn's disease is significantly lower than that of healthy children, and the diversity among these children with Crohn's disease is also significantly different, which may be attributed to differences in specific disease manifestations, disease location, or inflammation severity controls. At the same time, the living environment of intestinal flora in children with Crohn's disease may be affected by more factors, such as temperature, pH and so on. In contrast, the intestinal environment of healthy children is only affected by some single factors. Therefore, another focus of this study is to explore the environmental factors of impression flora and the relationship between intestinal flora and environmental factors.

When children received total gut nutrients, the relative abundance of gut microbiota of children with CD decreased significantly, and many species were already less abundant than the control group. However, this is a paradox because we expect EEN treatment to normalize and make healthier the "unhealthy" microbiome known to have a negative impact on the gut. In contrast, in the healthy group, EEN treatment made the microbiome more dysfunctional and more distant from NMDS than before EEN.

4.2 Bacteria correlated with Crohn’s Disease

The regression model and Core microbiome analysis allowed us to clearly observe the differences in gut microbiota between healthy children and CD children. In the regression model, it was found that C4 type short chain fatty acids could effectively improve Shannon diversity in the intestinal tract of patients. And C4 is important for our digestive health and the prevention of diseases, including neurological diseases. The production of short chain fatty acids is less than that of others, but studies show that they are vital to human health (Ygor Parladore Silva et al., 2020) [46]. It is very useful in fighting inflammation, which is also one of the symptoms of Crohn's disease.

At the same time, one of its main functions is to act as a major source of energy for intestinal lining cells, known as "colon cells." In fact, butyric acid accounts for 90 percent of its total energy requirements. These cells need this SCFA so they can perform

their vital functions, particularly maintaining the integrity of the intestinal wall. The intestinal wall is important because it acts as a barrier between the gut environment and the rest of the body (Gijs den Besten et al.., 2013) [47].
When the lining works effectively, it allows beneficial substances such as vitamins and minerals to enter the bloodstream and reach all parts of the body that need them. At the same time, it prevents opportunistic pathogens, toxins and food ingredients from getting into your bloodstream and getting sick. The barrier is made up of tightly linked proteins that control the opening and closing of the lining. However, if these connections are not closed, this can lead to a phenomenon called intestinal leakage. However, by increasing the production of butyric acid, you will be able to increase the production of this SCFA, which in turn will protect your bowel from leakage. Another advantage of this fiber breakdown product is its antioxidant and anticancer properties. It also has an important role: it leads to rogue cells committing suicide and prevents cancer from developing.

In the Core microbiome analysis, we found that Streptococcus was more prevalent in CD children and relatively rare in healthy children. It has been known as a Gram-positive, non-sporulating, facultative anaerobe. Some of the bacterial belong to these genera are strongly correlated to the inflammation. Like other streptococci, cell division S. haemoglobin occurs along a single axis, leading to the formation of streptococcus or pairs of *streptococcus*. (Kasper DL, 2007) [48] *Streptococcus* haemoglobin is a symbiotic bacterium widely distributed in the mouth, mainly on the surface of teeth, oral mucosa, and human saliva. Schistosoma's hair and adhesives facilitate its initial attachment, and glucan and eDNA production contribute to the maturation of the *S. sanguinis* biofilm.
Epidemiological studies have shown that *streptococcus* can inhibit dental caries. In vitro studies have shown competition between erythrosporidium and amoeba, the most common cariogenic species (Jens Kreth et al.., 2005) [49]. 16S rRNA sequencing results showed that *Streptococcus* sanguinis may be associated with periodontal health.
Compared with the diseased subgingival microbiome, the abundance of *streptococcus* was significantly increased in healthy individuals. However, in vitro studies suggest that schistosomiasis may also facilitate subsequent attachment of pathogens associated with periodontitis. Interaction between Streptococcus sanguinis and pathogens associated with periodontitis (Bin Zhu et al.., 2018) [50]. This reminds us that ”disease comes from the mouth, good teeth live" is not empty words. Most people with chronic conditions have oral or periodontal problems.
At the same time, research experiments found that in healthy children, such as *Subdoligranulum*, *Agathobacter* and CD children, the distribution is more common, and they all belong to the *Firmicutes* genera, and members of the *Firmicutes* family are known for making short chain fatty acids. The main producers of butyrate are anaerobic bacteria such as *Faecalibacterium prausnitzii*, *Eubacterium rectale* and *Roseburia spp* (DP Venegas et al., 2019) [51]. This also points to the importance of short-chain fatty acids for gut health. Therefore, supplementation of short-chain fatty acids is important for the treatment of Crohn's disease in the follow-up and detection of childhood CD.

At the same time, in the Generalized Linear Latent Variable Model, we found a variety of bacterial groups that were positively and negatively correlated with calprotectin. Among them, *Atopobium*, *Fusobacterium* and other bacterial groups were positively correlated with calprotectin, that is, with the increase of such microorganisms, it may lead to the increase of calprotectin. And calprotectin has a direct relationship with intestinal inflammation. In *Fusobacterium nucleatum (F. nucleatum)*, Crohn's disease (CD) is associated, but the mechanism by which F. nucleatum promotes CD development is unclear.

During CD progression, *F. nucleatum* infection plays a role in triggering the endoplasmic reticulum stress (ERS) pathway, promoting the destruction of the intestinal mucosal barrier. Nucleus activate the ERS pathway by targeting caspase the caspase activation recruitment recruitment domain 3 (CARD3, disrupting the mucosal barrier in vitro and in vivo. To regulate the CD process, *F. nucleatum* coordinates a molecular network involving CARD3 and ERS. Monitoring and focusing on *F. nucleatum* and its associated pathways will greatly contribute to CD prevention and treatment (Pan Cao et al.., 2020) [52].

Atopobium, which is also positively correlated to Calprotectin, it has been known that Chronic gastrointestinal tract inflammation is a hallmark of Crohn's disease (CD). The severity of CD is positively correlated with the presence of microbes that create hydrogen sulphide (H_2S), which is associated with a dysbiotic microbiome and a compromised immune system. *Atopobium* parvulum is a significant H_2S generator in CD patients' microbiomes. Two *Atopobium* parvulum cysteine desulfurases, *ApSufS* and *ApCsdB*, were biochemically characterized to determine their allosteric regulation. According to structural studies, *ApSufS* forms a dimer with the same properties as type II cysteine desulfurases. Cysteine desulfurylation requires four residues close to the active site to be catalyzed, and a section of short-chain residues provides access for substrate binding. Future CD therapeutic options will benefit from a deeper understanding of *ApSufS*. (Gapisha Karunakaran, 2022) [53]

At the same time, the experiment also found that *NK4A214_group* was negatively correlated with calprotectin, which indicated that *NK4A214_group* could effectively suppress intestinal inflammation. However, Acetate is a short-chain fatty acid (SCFA) produced by gut bacteria. Microbiome alpha-diversity was positively correlated with circulating acetate levels. And *NK4A214_group* has been found that are positively associated with acetate. But negative correlations were observed with *Lachnoclostridium* (Ana Nogal et al., 2021) [54]. So, the reason why *Lachnoclostridium* can help decrease the calprotectin still need to be deeply researched.

4.3 Mechanism of action of EEN

The mechanism of action of EEN on the intestinal tract of children is not clear, so this experiment also focused on the specific impact of EEN on the intestinal flora of children with CD. Through differential analysis, it was found that EEN effectively reduced a number of bacterial species. The abundance of some bacterial species was also increased. Among them, *Clostridium difficile* is a Gram-positive anaerobic bacterium, also known as *Lactonifactor*, which is highly transmitted by the fecal-oral route and is

one of the colonizing bacteria in the human intestine. Two protein exotoxins, toxin A (tcdA gene) and toxin B (tcdB gene), can be produced, which mediate tissue damage and inflammation and cause CDI-related symptoms (W K Smits at al., 2016) [55]. *Clostridium difficile* can colonize the intestine asymptomatically (no exotoxin production), or only present with mild diarrhea, and life-threatening pseudomembranous colitis and toxic megacolon can also occur. Healthy gut microbes have protective effects against *Clostridium difficile*, but patients with IBD suffer from dysbiosis of gut microbiome, decreased bacterial diversity, and loss of gut resistance to *C. difficile* colonization, leading to susceptibility to CDI and relapse.

From the existing research, we found that the increased abundance of microflora after EEN does not seem to be closely related to intestinal inflammation. For example, *Dialister* was found to be a marker for detecting Crohn's disease, and *Parasutterella* was only found in Crohn's disease. One of a type of bacteria was found in CD patients. Therefore, the method of this experiment is still limited (SR Dalal et al., 2014) [56]. It can only judge the changes of the flora before and after the experiment, but it cannot control and study whether these changes are caused by a single variable of EEN. At the same time, there is a lack of further research on these bacteria that produce changes, such as their mechanism of action on the gut, which is also a direction worthy of further research and discussion.

This research has several restrictions. The analysis in fecals allowed for the non-invasive collection of repeated samples; however, the outcomes might not be the same as for mucosal adherent bacteria. Despite accounting for multiple testing, the sample number of our participants was small, and this study may not have the power to investigate secondary outcomes including links between disease location and behavior and microbiome traits. In order to complete and advance this work, metatranscriptomic and metabolomic data are also needed because a community only expresses a variable subset of its genome at any one time. A small age gap existed between controls and CD children, and some patients were receiving concurrent treatment. The alterations seen were primarily driven by EEN or differences in community structure between groups (CD vs. Healthy), therefore any impact was modest.

4.4 Suggestions for further work

Current research has only identified some of the bacteria that potentially contribute to Crohn's disease or intestinal inflammation, but there is no next step in whether these bacteria can be effective in treating Crohn's disease and how to use these bacteria in actual treatment. At the same time, there is a lack of relevant research on the pathogenic mechanism in biology. In the same way, for those bacteria that are interested in the gut, whether they can effectively improve the intestinal environment of patients and how to put them into actual treatment still requires further experiments. For example, studies have found that short chain fatty acids such as C4 can effectively reduce intestinal inflammation, so how to effectively increase the secretion of such substances in patients is worthy of further exploration. As for the specific mechanism of EEN, it is necessary

to further observe the changes of the flora before and after EEN under the premise of controlling other objective factors, and analyze the role of each community.

CHAPTER 5

CONCLUSION

Excluding potential interference factors such as the interference of intestinal diseases on the body's microbiota, CD children before EEN treatment had lower levels of intestinal biodiversity than healthy children. Regression model showed that C4 in short chain fatty acid can effectively increase the diversity in children's gut (P=7.653e-06), and it plays an important role in resisting intestinal inflammation. It maintains the integrity of the intestinal wall, which is an important barrier within the gut. The experiment compared the difference between the specific flora in the intestines of healthy children and CD children, and found that the distribution of Streptococcus in CD children is much higher than that in healthy children, and such bacteria often cause periodontal health problems, which may also cause chronic diseases such as Crohn's disease. The bacteria that are more distributed in healthy children mostly belong to the Firmicutes genera, and such bacteria help to produce short chain fatty acid, which also explains the above findings. Therefore, in the subsequent treatment of Crohn's disease, means to promote such flora can be considered, thereby reducing inflammation in the intestine. There are also some bacteria that are also worth noting. For example, the experiment found that *Atopobium* (0.38), *Fusobacterium* (0.17) and calprotectin were positively correlated, and these two types of bacteria were directly related to causing Crohn's disease. And *NK4A214_group* (0.15), which is negatively correlated with calprotectin, can effectively alleviate intestinal inflammation. When exploring the effect of EEN on the intestinal flora of children, we found some microbial changes. EEN reduced the abundance of *Lactonifactor*, and its presence may lead to mild diarrhea or even pseudomembranous colitis in patients. EEN also increases the abundance of *Dialister*, *Parasutterella*, but there is currently no research and support to show the relationship between this type of bacteria and gut health.

REFERENCES

[1] Turunen P, Ashorn M, Auvinen A, et al. Long-term health outcomes in pediatric inflammatory bowel disease: a population-based study [J]. Inflamm Bowel Dis, 2009, 15(1):56-62.

[2] Swaminath A, Feathers A, Ananthakrishnan AN, et al. Systematic review with Meta-analysis: enteral nutrition therapy for the induction of remission in paediatric
Crohn's disease[J]. Aliment Pharmacol Ther, 2017, 46(7): 645-656

[3] Baert F, Moortgat L, van Assche G, et al. Mucosal healing predicts sustained clinical remission in patients with early-stage Crohn′ s disease[J]. Gastroenterology, 2010, 138(2):463-468.

[4] Ruemmele FM, Veres G, Kolho KL, et al. Consensus guidelines of ECCO/ESPGHAN on the medical management of pediatric Crohn′ s disease[J]. J Crohns Colitis, 2014, 8(10): 1179-1207.

[5] Marc A. Sidler, Daniel A. Lemberg, Andrew S. Day, Combination of nutritional therapy and medical therapy for the management of enterovesical fistula in paediatric Crohn's disease: A case report, e-SPEN, the European e-Journal of Clinical Nutrition and Metabolism, Volume 3, Issue 2, 2008, Pages e89-e91, ISSN 1751-4991,

[6] Grover Z, Burgess C, Muir R, et al. Early mucosal healing with exclusive enteral nutrition is associated with improved outcomes in newly diagnosed children with luminal Crohn's disease[J]. J Crohns Colitis, 2016, 10(10): 1159-1164.

[7] Quince C, Ijaz UZ, Loman N, Eren AM, Saulnier D, Russell J, Haig SJ, Calus ST, Quick J, Barclay A, Bertz M, Blaut M, Hansen R, McGrogan P, Russell RK, Edwards CA, Gerasimidis K. Extensive Modulation of the Fecal Metagenome in Children With Crohn's Disease During Exclusive Enteral Nutrition. Am J Gastroenterol. 2015 Dec;110(12):1718-29; quiz 1730. doi: 10.1038/ajg.2015.357. Epub 2015 Nov 3. PMID: 26526081; PMCID: PMC4697132.

[8] Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, Mende DR, Li J, Xu J, Li S, Li D, Cao J, Wang B, Liang H, Zheng H, Xie Y, Tap J, Lepage P, Bertalan M, Batto JM, Hansen T, Le Paslier D, Linneberg A, Nielsen HB, Pelletier E, Renault P, Sicheritz-Ponten T, Turner K, Zhu H, Yu C, Li S, Jian M, Zhou Y, Li Y, Zhang X, Li S, Qin N, Yang H, Wang J, Brunak S, Doré J, Guarner F, Kristiansen K, Pedersen O, Parkhill J, Weissenbach J; MetaHIT Consortium, Bork P, Ehrlich SD, Wang J. A human gut microbial gene catalogue established by metagenomic sequencing. Nature. 2010 Mar 4;464(7285):59-65. doi: 10.1038/nature08821. PMID: 20203603; PMCID: PMC3779803.

[9] Tse JKY. Gut Microbiota, Nitric Oxide, and Microglia as Prerequisites for Neurodegenerative Disorders. ACS Chem Neurosci. 2017 Jul 19;8(7):1438-1447. doi: 10.1021/acschemneuro.7b00176. Epub 2017 Jul 7. PMID: 28640632.

[10] Dalal SR, Chang EB. The microbial basis of inflammatory bowel diseases. J Clin Invest. 2014 Oct;124(10):4190-6. doi: 10.1172/JCI72330. Epub 2014 Aug 1. PMID: 25083986; PMCID: PMC4191005.

[11] Tjellström B, Högberg L, Stenhammar L, Magnusson KE, Midtvedt T, Norin E, Sundqvist T. Effect of exclusive enteral nutrition on gut microflora function in children with Crohn's disease. Scand J Gastroenterol. 2012 Dec;47(12):1454-9. doi: 10.3109/00365521.2012.703234. Epub 2012 Sep 28. PMID: 23016828.

[12] Rahib, L., MacLennan, N., Horvath, S. et al. Glycerol kinase deficiency alters expression of genes involved in lipid metabolism, carbohydrate metabolism, and insulin signaling. Eur J Hum Genet 15, 646–657 (2007). https://doi.org/10.1038/sj.ejhg.5201801

[13] Musso G, Gambino R, Cassader M. Gut microbiota as a regulator of energy homeostasis and ectopic fat deposition: mechanisms and implications for metabolic disorders. Curr Opin Lipidol. 2010 Feb;21(1):76-83. doi: 10.1097/MOL.0b013e3283347ebb. PMID: 19915460.

[14] Identify and compare microbes from complex microbiomes or environments, available from

https://www.illumina.com/areas-of-interest/microbiology/microbial-sequencing-methods/16s-rrna-sequencing.html [accessed on 28th July 2022]

[15] Sequence thousands of organisms in parallel, available from https://www.illumina.com/areas-of-interest/microbiology/microbial-sequencing-methods/shotgun-metagenomic-sequencing.html [accessed on 28th July 2022]

[16] R Studio, Take control of your R code, available from https://www.rstudio.com/products/rstudio/ [accessed on 9th June 2022]

[17] Jari Oksanen, Gavin L. Simpson, F. Guillaume Blanchet et al., Package 'vegan' , April 17, 2022

[18] ggplot2 3.35, available from https://ggplot2.tidyverse.org/ [accessed from 9th June 2022]

[19] Chase JM, Myers JA. Disentangling the importance of ecological niches from stochastic processes across scales. Philos Trans R Soc Lond B Biol Sci. 2011 Aug 27;366(1576):2351-63. doi: 10.1098/rstb.2011.0063. PMID: 21768151; PMCID: PMC3130433.

[20] Stegen, J., Lin, X., Fredrickson, J. et al. Quantifying community assembly processes and identifying features that impose them. ISME J 7, 2069–2079 (2013). https://doi.org/10.1038/ismej.2013.93

[21] Neu, Alexander & Allen, Eric & Roy, Kaustuv. (2021). Defining and quantifying the core microbiome: Challenges and prospects. Proceedings of the National Academy of Sciences. 118. e2104429118. 10.1073/pnas.2104429118.

[22] Lahti, L., & Shetty, S. (2017). microbiome R package. Bioconductor. https://doi.org/10.18129/B9.bioc.microbiome

[23] The Use and Types of Alpha-Diversity Metrics in Microbial NGS, available from https://www.cd-genomics.com/microbioseq/the-use-and-types-of-alpha-diversity-metrics-in-microbial-ngs.html [accessed from 7th July 2022]

[24] McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One. 2013 Apr 22;8(4):e61217. doi: 10.1371/journal.pone.0061217. PMID: 23630581; PMCID: PMC3632530.

[25] Steven W. Kembel, Peter D. Cowan, Matthew R. Helmus, William K. Cornwell, Helene Morlon, David D. Ackerly, Simon P. Blomberg, Campbell O. Webb, Picante: R tools for integrating phylogenies and ecology, Bioinformatics, Volume 26, Issue 11, 1 June 2010, Pages 1463–1464

[26] Ning D, Deng Y, Tiedje JM, Zhou J. A general framework for quantitatively assessing ecological stochasticity. Proc Natl Acad Sci U S A. 2019 Aug 20;116(34):16892-16898. doi: 10.1073/pnas.1904623116. Epub 2019 Aug 7. PMID: 31391302; PMCID: PMC6708315.

[27] Stegen, J., Lin, X., Fredrickson, J. et al. Quantifying community assembly processes and identifying features that impose them. ISME J 7, 2069–2079 (2013). https://doi.org/10.1038/ismej.2013.93

[28] Vass, M., Székely, A.J., Lindström, E.S. et al. Using null models to compare bacterial and microeukaryotic metacommunity assembly under shifting environmental conditions. Sci Rep 10, 2455 (2020). https://doi.org/10.1038/s41598-020-59182-1

[29] Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550 (2014). https://doi.org/10.1186/s13059-014-0550-8

[30] Barry Kassambara, Homayoon Ganji, Takamitsu Kajisa, Impact of agricultural water allocation on the ecosystems in the Inner Niger River Delta, January 2018International Journal of GEOMATE 14(42) DOI:10.21660/2018.42.7251

[31] Mattias Nyström, Johan Holmgren, Håkan Olsson, Prediction of tree biomass in the forest–tundra ecotone using airborne laser scanning, Remote Sensing of Environment, Volume 123, 2012, Pages 271-279, ISSN 0034-4257, https://doi.org/10.1016/j.rse.2012.03.008.

[32] Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. Journal of Statistical Software, 28(5), 1–26. https://doi.org/10.18637/jss.v028.i05

[33] Lahti M, Savolainen K, Tuovinen S, Pesonen AK, Lahti J, Heinonen K, Hämäläinen E, Laivuori H, Villa PM, Reynolds RM, Kajantie E, Räikkönen K. Maternal Depressive Symptoms During and After Pregnancy and Psychiatric Problems in Children. J Am Acad Child Adolesc Psychiatry. 2017 Jan;56(1):30-39.e7. doi: 10.1016/j.jaac.2016.10.007. Epub 2016 Oct 24. PMID: 27993226.

[34] Bjork, J. R., Hui, F. K. C., O'Hara, R. B., & Montoya, J. M. (2018). Uncovering the drivers of host-associated microbiota with joint species distribution modelling. Molecular Ecology, 27, 2714–2724. https://doi.org/10.1111/mec.14718

[35] Björk JR, Hui FKC, O'Hara RB, Montoya JM. Uncovering the drivers of host-associated microbiota with joint species distribution modelling. Mol Ecol. 2018 Jun;27(12):2714-2724. doi: 10.1111/mec.14718. Epub 2018 Jun 4. PMID: 29761593; PMCID: PMC6025780.

[36] Lammel, D.R., Barth, G., Ovaskainen, O. et al. Direct and indirect effects of a pH gradient bring insights into the mechanisms driving prokaryotic community structures. Microbiome 6, 106 (2018). https://doi.org/10.1186/s40168-018-0482-8

[37] BUISSON, L., THUILLER, W., LEK, S., LIM, P., & GRENOUILLET, G. (2008). Climate change hastens the turnover of stream fish assemblages. Global Change Biology, 14(10), 2232–2248. https://doi.org/10.1111/j.1365-2486.2008.01657.x

[38] Susin A, Wang Y, Lê Cao KA, Calle ML. Variable selection in microbiome compositional data analysis. NAR Genom Bioinform. 2020 May 13;2(2):lqaa029. doi: 10.1093/nargab/lqaa029. PMID: 33575585; PMCID: PMC7671404.

[39] Greenblum S, Turnbaugh PJ, Borenstein E. Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. Proc Natl Acad Sci U S A. 2012 Jan 10;109(2):594-9. doi: 10.1073/pnas.1116053109. Epub 2011 Dec 19. PMID: 22184244; PMCID: PMC3258644.

[40] Greenblum, Catherine A. PhD, FNP-BC, ARNP1; Rowe, Meredeth A. PhD, RN, FAAN1; Neff, Donna Felber PhD, RN2; Greenblum, Jesse S. MD, MS, FACOG3. Midlife women: symptoms associated with menopausal transition and early postmenopause and quality of life. Menopause: The Journal of The North American Menopause Society: January 2013 - Volume 20 - Issue 1 - p 22-27 doi: 10.1097/gme.0b013e31825a2a91

[41] Salonen, A., Salojärvi, J., Lahti, L., & de Vos, W. M. (2012). The adult intestinal core microbiota is determined by analysis depth and health status. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases, 18 Suppl 4, 16–20. https://doi.org/10.1111/j.1469-0691.2012.03855.x

[42] Hazen, T. C., Dubinsky, E. A., DeSantis, T. Z., Andersen, G. L., Piceno, Y. M., Singh, N., Jansson, J. K., Probst, A., Borglin, S. E., Fortney, J. L., Stringfellow, W. T., Bill, M., Conrad, M. E., Tom, L. M., Chavarria, K. L., Alusi, T. R., Lamendella, R., Joyner, D. C., Spier, C., Baelum, J., … Mason, O. U. (2010). Deep-sea oil plume enriches indigenous oil-degrading bacteria. Science (New York, N.Y.), 330(6001), 204–208. https://doi.org/10.1126/science.1195979

[43] Roh, J. D., Sawh-Martinez, R., Brennan, M. P., Jay, S. M., Devine, L., Rao, D. A., Yi, T., Mirensky, T. L., Nalbandian, A., Udelsman, B., Hibino, N., Shinoka, T., Saltzman, W. M., Snyder, E., Kyriakides, T. R., Pober, J. S., & Breuer, C. K. (2010). Tissue-engineered vascular grafts transform into mature blood vessels via an inflammation-mediated process of vascular remodeling. Proceedings of the National Academy of Sciences of the United States of America, 107(10), 4669–4674. https://doi.org/10.1073/pnas.0911465107

[44] Tottey, W., Denonfoux, J., Jaziri, F., Parisot, N., Missaoui, M., Hill, D., Borrel, G., Peyretaillade, E., Alric, M., Harris, H. M., Jeffery, I. B., Claesson, M. J., O'Toole, P. W., Peyret, P., & Brugère, J. F. (2013). The human gut chip

"HuGChip", an explorative phylogenetic microarray for determining gut microbiome diversity at family level. PloS one, 8(5), e62544. https://doi.org/10.1371/journal.pone.0062544

[45] Jalanka-Tuovinen, J., Salonen, A., Nikkilä, J., Immonen, O., Kekkonen, R., Lahti, L., Palva, A., & de Vos, W. M. (2011). Intestinal microbiota in healthy adults: temporal analysis reveals individual and common core and relation to intestinal symptoms. PloS one, 6(7), e23035. https://doi.org/10.1371/journal.pone.0023035

[46] Silva, Y. P., Bernardi, A., & Frozza, R. L. (2020). The Role of Short-Chain Fatty Acids From Gut Microbiota in Gut-Brain Communication. Frontiers in endocrinology, 11, 25. https://doi.org/10.3389/fendo.2020.00025

[47] den Besten, G., van Eunen, K., Groen, A. K., Venema, K., Reijngoud, D. J., & Bakker, B. M. (2013). The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. Journal of lipid research, 54(9), 2325–2340. https://doi.org/10.1194/jlr.R036012

[48] Sebastian, S., Dillon, S. T., Lynch, J. G., Blalock, L. T., Balon, E., Lee, K. T., Comstock, L. E., Conlan, J. W., Rubin, E. J., Tzianabos, A. O., & Kasper, D. L. (2007). A defined O-antigen polysaccharide mutant of Francisella tularensis live vaccine strain has attenuated virulence while retaining its protective capacity. Infection and immunity, 75(5), 2591–2602. https://doi.org/10.1128/IAI.01789-06

[49] Kreth, J., Merritt, J., Shi, W., & Qi, F. (2005). Competition and coexistence between Streptococcus mutans and Streptococcus sanguinis in the dental biofilm. Journal of bacteriology, 187(21), 7193–7203. https://doi.org/10.1128/JB.187.21.7193-7203.2005

[50] Zhu, B., Macleod, L. C., Kitten, T., & Xu, P. (2018). Streptococcus sanguinis biofilm formation & interaction with oral pathogens. Future microbiology, 13(8), 915–932. https://doi.org/10.2217/fmb-2018-0043

[51] Parada Venegas, D., De la Fuente, M. K., Landskron, G., González, M. J., Quera, R., Dijkstra, G., Harmsen, H., Faber, K. N., & Hermoso, M. A. (2019). Short Chain Fatty Acids (SCFAs)-Mediated Gut Epithelial and Immune Regulation and Its Relevance for Inflammatory Bowel Diseases. Frontiers in immunology, 10, 277. https://doi.org/10.3389/fimmu.2019.00277

[52] Cao, P., Chen, Y., Guo, X., Chen, Y., Su, W., Zhan, N., & Dong, W. (2020). Fusobacterium nucleatum Activates Endoplasmic Reticulum Stress to Promote Crohn's Disease Development via the Upregulation of CARD3 Expression. Frontiers in pharmacology, 11, 106. https://doi.org/10.3389/fphar.2020.00106

[53] Karunakaran, G., Yang, Y., Tremblay, V., Ning, Z., Martin, J., Belaouad, A., Figeys, D., Brunzelle, J. S., Giguere, P. M., Stintzi, A., & Couture, J. F. (2022). Structural analysis of Atopobium parvulum SufS cysteine desulfurase linked to Crohn's disease. FEBS letters, 596(7), 898–909. https://doi.org/10.1002/1873-3468.14295

[54] Nogal, A., Valdes, A. M., & Menni, C. (2021). The role of short-chain fatty acids in the interplay between gut microbiota and diet in cardio-metabolic health. Gut microbes, 13(1), 1–24. https://doi.org/10.1080/19490976.2021.1897212

[55] Smits, W. K., Lyras, D., Lacy, D. B., Wilcox, M. H., & Kuijper, E. J. (2016). Clostridium difficile infection. Nature reviews. Disease primers, 2, 16020. https://doi.org/10.1038/nrdp.2016.20

[56] Dalal, S. R., & Chang, E. B. (2014). The microbial basis of inflammatory bowel diseases. The Journal of clinical investigation, 124(10), 4190–4196. https://doi.org/10.1172/JCI72330

APPENDIX

7.1 Appendix A

	Model	Cross-validation Errors
4	Shannon ~ Calprotectin_dry + C4 + iC5 + iC6	0.53277
5	Shannon ~ Calprotectin_dry + C2 + C4 + iC5 + iC6	0.53339
7	Shannon ~ Calprotectin_dry + C2 + C4 + iC5 + iC6 + C6 + C7	0.54025
3	Shannon ~ Calprotectin_dry + C4 + iC5	0.54129
6	Shannon ~ Calprotectin_dry + C2 + C4 + iC5 + iC6 + C7	0.54296
8	Shannon ~ Calprotectin_dry + pH + C2 + C4 + iC5 + iC6 + C6 + C7	0.54400
2	Shannon ~ Calprotectin_dry + C4	0.54444
9	Shannon ~ Age + Calprotectin_dry + pH + C2 + C4 + iC5 + iC6 + C6 + C7	0.54875
1	Shannon ~ C4	0.55326
11	Shannon ~ Age + Calprotectin_dry + pH + C2 + C3 + C4 + iC5 + iC6 + C6 + C7 + C8	0.55378
12	Shannon ~ Age + Calprotectin_dry + pH + C2 + C3 + C4 + iC5 + C5 + iC6 + C6 + C7 + C8	0.55404
10	Shannon ~ Age + Calprotectin_dry + pH + C2 + C4 + iC5 + iC6 + C6 + C7 + C8	0.55449
13	Shannon ~ Age + Calprotectin_dry + pH + C2 + C3 + iC4 + C4 + iC5 + C5 + iC6 + C6 + C7 + C8	0.56392

Figure 9: The different models generated by subset regression modeling. Each model shows their cross-validation errors. The less the value of cross-validation errors, the model will be more significant.