

Gut microbial and SCFA profiling identifies a distinct Dorea-Megasphaera-Valeric acid axis associated with Pakistani beta thalassemia cohorts

Muhammad Bilal Jamshaid

COMSATS University Islamabad

Hajra Khurshid

COMSATS University Islamabad

Iqra Anwar

COMSATS University Islamabad

Jannat Ejaz

COMSATS University Islamabad

Aqsa Ameer

University of Management and Technology

Konstantinos Gerasimidis

University of Glasgow

Zobia Noreen

COMSATS University Islamabad

Sundus Javed

COMSATS University Islamabad

Umer Zeeshan Ijaz

Umer.Ijaz@glasgow.ac.uk

University of Glasgow

Aamira Tariq

COMSATS University Islamabad

Article

Keywords: Beta-thalassemia, Short chain fatty acids (SCFA), Microbiome, Iron -overload

Posted Date: October 22nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10695064/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Gut microbial and SCFA profiling identifies a distinct *Dorea-Megasphaera*-Valeric acid axis associated with Pakistani beta thalassemia cohorts

Muhammad Bilal Jamshaid^a, Hajra Khurshid^a, Iqra Anwar^a, Jannat Ejaz^a, Aqsa Ameer^{a,b}, Konstantinos Gerasimidis^c, Zobia Noreen^a, Sundus Javed^a, Umer Zeeshan Ijaz^{d,e,f,*}, and Aamira Tariq^{a,*}

^a*Department of Biosciences, COMSATS University Islamabad, 45550, Pakistan*

^b*School of Food and Agricultural Sciences, University of Management and Technology, Pakistan*

^c*School of Medicine, Dentistry & Nursing, Glasgow Royal Infirmary, Glasgow, G31 2ER, UK*

^d*Water & Environment Research Group, University of Glasgow, Mazumdar-Shaw Advanced Research Centre, Glasgow G11 6EW, UK*

^e*Department of Molecular and Clinical Cancer Medicine, University of Liverpool, Liverpool L69 7BE, UK*

^f*National University of Ireland, Galway, University Road, Galway, H91 TK33, Ireland*

***Joint Corresponding Authors**

email: aamira_tariq@comsats.edu.pk

email: Umer.Ijaz@glasgow.ac.uk (<http://userweb.eng.gla.ac.uk/umer.ijaz>)

Abstract

Background: β -thalassemia is the most prevalent inherited blood disorder in Pakistan, driven by a high rate of consanguineous marriages. It has led to high rate of childhood morbidity that require lifelong blood transfusions causing toxic iron overload. Exploring the associated gut microbiome and metabolite shifts in this population can reveal how iron toxicity disrupts crucial bacterial pathways, offering a foundation for targeted nutritional therapies. This study investigates the gut microbial structural and functional alteration in three cohorts, namely healthy, Thalassemia Minor, and transfusion-dependent Thalassemia Major.

A comprehensive multi-omic approach integrating 16S rRNA sequencing, metabolic pathway analysis via MetaCyc, and short-chain fatty acid (SCFA) profiling using the Multiblock Sparse Partial Least Squares Discriminant Analysis, DIABLO framework, was employed to identify disease-specific microbial signatures.

Results: A strong decrease in abundance of *Alistipes* and *Bacteroides*, was observed with the disease progression. The Thalassemia Minor group exhibited an abundance of *Prevotella* and *Eubacterium*, whereas Thalassemia Major showed an increase in abundance of iron-tolerant species, including *Megasphaera*, *Sarcina*, and *Akkermansia*. On the functional side, a significant upregulation of pro-inflammatory lipopolysaccharide biosynthesis pathways was observed in the thalassemia cohorts as opposed to the healthy cohort. The healthy cohort maintained essential Mevalonate and Flavin biosynthetic pathways. Furthermore, a distinct metabolic signature of the *Dorea-Megasphaera*-Valeric acid axis was identified in the Thalassemia Major cohort.

Conclusion: Beta-thalassemia leads to shift from a balanced butyrate-rich environment towards an endotoxin-loaded pro-inflammatory state mainly due to iron overload-induced selective pressure. These findings support incorporation of iron-stable probiotics (e.g., *Bifidobacterium*) and postbiotic valeric acid along with standard chelation therapy, to rejuvenate the gut integrity and reduce systemic inflammation.

Keywords: Beta-thalassemia, Short chain fatty acids (SCFA), Microbiome, Iron -overload

Introduction

β -thalassemia is a heterogeneous autosomal recessive haematological disorder caused by an excess of unpaired globin chains due to a mutation in the β -globin gene at chromosome 11 that reduces or eliminates β -globin chain production (1) (2). This imbalance between the alpha and beta chains results in defective erythropoiesis followed by haemolysis (3), a hallmark of β -thalassemia. Annually, 25,500 Transfusion-Dependent- β -thalassemia (TDT β -thalassemia) affected infants are born globally in addition to 56,000 with non-transfusion dependent forms (4). The prevalence is high in developing countries like Pakistan with a carrier frequency of 5-8% and 5000 new-borns with TDT β -thalassemia (5) mainly attributed to consanguineous marriages. Ineffective erythropoiesis leads to an increase in iron absorption by enhancing erythroferrone production and suppressing hepcidin (6). Low hepcidin upregulates ferroportin on the enterocytes leading to 3-4 times higher iron absorption than the normal rate. Meanwhile, accelerated hemolysis yields a surplus of heme, subsequently processed into unconjugated bilirubin. Chronic iron overload in the liver causes inflammation and scarring over time, which impairs the liver's ability to process and clear bilirubin from the blood causing overflow of the bilirubin into the intestinal lumen creating a metabolic pressure on the gut. High bilirubin in the gut potentially favour the microbial taxa that can thrive in the high bilirubin environment (7). At the same time iron overload can also cause microbial dysbiosis favouring the growth of pathogenic bacteria (8,9). It has been observed that β -thalassemia mice exhibit increased intestinal mucosal injury due to iron overload leading to enhanced gut permeability and inflammatory responses triggered by bacterial endotoxins (10). Gut inflammation and disruption in the gut barrier integrity is not solely influenced by microbial dysbiosis but also by the altered metabolites produced by the altered microbial taxa(11).

Short chain fatty acids (SCFAs) are microbial metabolites that play a pivotal role in maintaining intestinal integrity by inhibiting colonization of pathogenic microbes(12) such as *Klebsiella spp.* in Asian regions and *Yersinia enterocolitica* in Western nations (13). SCFA also serve as potent anti-inflammatory molecules that modulate immunological responses and suppress pro-inflammatory cytokine synthesis(14). At the same time recurrent blood transfusion in β -thalassemia may lead to iron overload causing epithelial cell degeneration and gut inflammation which may also influence bacterial physiology and virulence (15). For instance, *Pseudomonas aeruginosa*, *Legionella pneumophila*, *Streptococcus pneumoniae*, *Escherichia coli*, *Klebsiella species*, and *Listeria monocytogenes* have been observed to exhibit increased virulence *in vitro* when exposed to excessive levels of iron (16).

Patients with β -thalassemia have a high rate of malnutrition, and weakened immune system. Investigating the gut bacteria of these patients helps doctors manage iron overload, improving gut health through probiotic, prebiotic and synbiotic interventions(17). Since efficacy of these treatments vary in different racial populations, there is an urgent need to explore gut microbiome alterations in β -thalassemia in Pakistani patients. In this study, we analysed alteration in gut microbial communities at taxonomic and functional level using fecal samples from thalassemia patients and healthy controls. The altered microbial profile was further correlated with the SCFA levels. This discovery of gut microbial dysbiosis and associated SCFAs in thalassemia patients can facilitate development of tailored nutritional interventions for disease management of β -thalassemia patients exhibiting iron overload.

Methods

Aim and design of the study

The main aim of the study was to associate the gut microbiome and metabolic profiling with beta-thalassemia. The study design is depicted in Figure 1.

Study area and participants recruitment

A study was carried out in Islamabad, Pakistan, involving thalassaemia patients who were undergoing medical treatment at Sundus Foundation, Islamabad. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Review Board at COMSATS University Islamabad (CUI/Bio/ERB/2021/60). Written informed consent of the participants was obtained before data collection.

Clinical histories and blood transfusion records were obtained for all participants, who were aged between 4 and 17 years old. Patients pre-screened for beta-thalassemia major on a regular blood transfusion regimen with no other diagnosed complications, pre-screened beta-thalassemia minor patients, and healthy control subjects who tested negative for thalassemia were recruited. Patients with chronic infections or antibiotic use in the past 3 months were excluded.

Haematological profiling

All the CBC parameters were regressed in an ANOVA model by using `param ~ Groups` to obtain a P-value separately for each parameter. Afterwards, we adjusted the P-values for multiple comparisons (all CBC parameters) using Benjamini & Hochberg (1995) and selected only those where $\text{Adj. } p < 0.05$. For each CBC parameter, we have retained only covariates

Adj. $P < 0.05$. The complete table is shown in Table I, whilst individual plots against Groups are shown in Figure 1.

Sample collection for faecal microbiome analysis

A total of 25 fecal samples were obtained from 14 patients with beta-thalassemia major, 7 individuals carrying the beta-thalassemia trait, and 4 healthy control subjects. Routine complete blood counts (CBCs) were performed, and thalassemia was diagnosed by standard hematologic tests such as hemoglobin electrophoresis and genetic analysis (18). Fresh fecal samples were collected in sterile containers and directly submitted to the Microbiology and Public Health Laboratory, COMSATS University Islamabad, after freezing them in a freezer at -20°C storage at -80°C .

DNA extraction, total microbial load, and sequencing

DNA was extracted using extraction kits (QIAGEN DNeasy PowerSoil Pro Kits), according to manufacturer's specifications. Illumina MiSeq platform (Illumina, UK) utilizing Schloss 250 bp paired end method for 16S rRNA Sequencing of V4 region was performed (19).

Bioinformatics

A total of 1074419 paired-end reads were obtained from 25 samples from which Amplicon Sequencing Variants (ASVs) were identified. Briefly, the Dada2 algorithm (20) was implemented within the QIIME2 platform (version 2019.7.0) (21) with the parameters `--p-trim-left-f 0 --p-trim-left-r 0 --p-trunc-len-f 225 --p-trunc-len-r 215` in `qiime dada2 denoise-paired` plugin. The analysis generated an abundance table comprising of 1,723 ASVs. The summary statistics of sample-wise reads mapped to ASVs were as follows [Min:43,981; 1st Quartile:52,678; Median:54,810; Mean:57,559; 3rd Quartile:61,896; Max:82,227]. Furthermore, within the QIIME2 platform, a rooted phylogenetic tree for the ASVs was constructed using `qiime phylogeny align-to-tree-mafft-fasttree` plugin. Taxonomic classification was performed against the recent SILVA SSU Ref NR database release v.138 (22) using the `qiime feature-classifier classify-sklearn` plugin. Predicted functional and metabolic profiles were obtained using `qiime picrust2 full-pipeline` (<https://github.com/picrust/picrust2/wiki/q2-picrust2-Tutorial/>) with the parameters `--p-hsp-method pic --p-max-ntsi 2`. Out of the total 1,723 identified ASVs, 24 were discarded because they showed inadequate alignment with the available reference genomes, while another 13

were filtered out for surpassing the NSTI cut-off value of 2.0. After these quality control steps, 1,686 ASVs, representing 97.9% of the original dataset were retained and successfully mapped to the PICRUSt2 reference database (23). This high proportion of matched ASVs supports the reliability of the inferred functional predictions. Subsequent PICRUSt2-based functional profiling across the 25 samples resulted in the identification of 10,543 KEGG Orthologs (KOs) and 489 MetaCyc metabolic pathways.

Statistics

In accordance with the QIIME2 filtering guidelines <https://docs.qiime2.org/2022.8/tutorials/filtering/>, preliminary data processing was performed to enhance dataset quality. Sequences identified as common contaminants, including mitochondrial and chloroplast reads, were removed, along with ASVs that remained unassigned at all taxonomic levels. Additionally, samples with low sequencing depth (<5,000 reads) and those not relevant to the study were excluded from subsequent analyses. This processing resulted in a final abundance matrix comprising 25 samples and 1,099 ASVs.

Alpha and beta diversity analyses were conducted using the vegan package in R (24). For alpha diversity assessment, samples were rarefied to the minimum sequencing depth prior to analysis. Community diversity and richness were evaluated using the Shannon diversity index, H' , and the Chao1 estimator. Subsequently, group-wise comparisons were performed using analysis of variance (ANOVA) implemented through R's `av()` function, where statistically significant differences were detected. The corresponding p-values were annotated on the alpha diversity plots above the lines connecting the respective groups.

To visualize inter-sample variation, Principal Coordinate Analysis (PCoA) was performed using the `cmdscale()` function from R's Vegan package, applying multiple distance metrics. Specifically, (i) Bray–Curtis dissimilarity was calculated from the ASV abundance table to assess compositional differences among samples; (ii) unweighted UniFrac distances were computed using the `phyloseq` package (25) based on the rooted phylogenetic tree to evaluate phylogeny-driven differences; and (iii) *Hierarchical Meta-Storms* (HMS) (26) analysis was conducted on the KEGG Ortholog (KO) abundance table to explore functional dissimilarities.

For PCoA visualization, 95% confidence ellipses representing the standard error of group centroids were generated using `ordiellipse()` function from R's Vegan package. The three distance measures were further subjected to the permutational ANOVA (PERMANOVA) analysis using `adonis2()` function in R's Vegan package. For statistically significant models, the returned R^2 value was interpreted as the proportion of variance in microbial composition

or predicted function explained by the respective covariate i.e. scaled to 1. These results were primarily illustrated alongside the PCoA ordinations.

Additionally, PERMANOVA was used along with Redundancy Analysis (RDA) on reduced set of covariates. To identify the covariates most strongly associated with community variation, distance-based RDA was applied using Vegan's `capscale()`, followed by stepwise model selection with `ordistep()` functions in the following set of commands: `cap.env = capscale(abund_table.dist~., meta_table); mod0.env = capscale(abund_table.dist~1, meta_table); step.env=ordistep(mod0.env, scope=formula(cap.env), direction="both", Pin=0.1, perm.max=9999, R2scope=TRUE)`. Variables retained in the final model `step.env$anova` with $p < 0.05$ were considered significantly associated with community variation and subsequently included in the PERMANOVA analysis.

To find differentially abundant genera (ASVs collated at genus level using SILVA taxonomy), we have used `DESeqDataSetFromMatrix()` function from DESeq2 package with the adjusted p -value significance cut-off of 0.05 and log2 fold change cut-off of 2 (27). Afterwards, Bayesian shrinkage is applied to obtain shrunken log fold changes subsequently employing the Wald test for obtaining significances. To visualise the differentially abundant taxa or pathways, we have used *Total Sum Scaling* followed by a *Centralized Log Ratio* (TSS+CLR) transformation on the raw abundance tables (genera/MetaCyc pathways).

We have used DIABLO algorithm (28) to integrate $M = 3$ datasets denoted by $X^{(1)}(N \times P_1)$, $X^{(2)}(N \times P_2)$, $X^{(3)}(N \times P_3)$, where $X^{(1)}$ represents the TSS+CLR (Total Sum Scaling + Centralised Log Ratio) normalised ASVs abundance table collated at genus level, $X^{(2)}$ represents the autoscaled dry SCFA table, and $X^{(3)}$ denotes the dummified (i.e., binary representation of the three cohorts: *Normal*, *Minor* Thallasemia, and *Major* Thallasemia).

To predict the number of latent components (associated loading vectors) and the number of discriminants in DIABLO, `block.splsda()` and `tune.block.splsda()` functions were used. We have used `ncomp=3` components and `dist="mahalanobis.dist"` in `tune.block.splsda()` function, respectively for optimization.

Results

Haematological parameters in β -beta-thalassemic patients

In the first step, parametric regression analysis on different haematological parameters like haemoglobin (Hb), platelet count (PLT), mean corpuscular haemoglobin concentration

(MCHC), haematocrit (HCT), mean corpuscular volume (MCV), mean red blood cell (RBC) count and red blood cell distribution width (RDW) was performed against Thalassemia major (N=14), thalassemia minor (N=7) and normal samples (N=4) groups. A significant decrease was observed in β -thalassemia major patients was observed for Hb (adjusted $p < 0.05$), PLT (adjusted $p < 0.001$), HCT (adjusted $p < 0.05$) and RDW (adjusted $p < 0.001$) parameters as opposed to normal and thalassemia minor group. Significant difference in RBC counts among the β -thalassemia major and β -thalassemia minor were observed (adjusted $p = 0.3090147785791502$). MCV (adjusted $p < 0.05$) and MCHC (adjusted $p < 0.001$) were higher in thalassemia major compared to normal and thalassemia minor group (Figure 2). Although the MCV and MCHC are higher in normal than in thalassaemia patients, but the post transfusion values in thalassaemia major patients exhibit elevated MCH and MCHC because of the presence of the normal people blood or elevated MCH values.

Reduced alpha diversity accompanies β -thalassemia major trait

To assess how β -thalassemia affects microbial alpha diversity (the diversity within samples), we calculated the number of microbial species found in each fecal sample (richness) along with the Shannon index, which evaluates both richness and evenness at the species level. Samples from individuals with thalassemia minor exhibited greater mean richness compared to both control and thalassemia major samples, suggesting a potential difference in microbial diversity between these groups. Shannon index, was significantly different ($p < 0.05$) between thalassemia minor and major groups. The Shannon index was predominantly elevated in thalassemia minor compared to thalassemia major, suggesting a greater alpha diversity in the minor group. These findings demonstrate that beta-thalassemia has a substantial impact on modulating the gut microbiome, with thalassemia minor exhibiting a distinct microbial profile compared to thalassemia major.

To analyse sample dissimilarity, Bray-Curtis distance was used to assess microbial compositional changes (Figure 3D). Weighted Unifrac estimated phylogenetic changes while considering ASV abundances (Figure 3E), and Hierarchical Meta-Storms indicated metabolic changes (Figure 3F). Control samples showed a clear distinction in beta diversity from those in the thalassemia samples, as determined by Bray Curtis analysis ($R^2 = 0.123$, $p > 0.001$) and Hierarchical Meta-Storms ($R^2 = 0.15045$, $p > 0.05$). However, no significant variability was detected in the phylogenetic analysis through Permutational Multivariate Analysis of Variance (PERMANOVA).

Since measurements for serum ferritin was missing for a few samples, it was not included in the RDA analysis and instead on a reduced sample (taking samples from all three cohorts: Normal, Minor Thalassemia and Major Thalassemia), and considering different beta diversity distances, PERMANOVA was performed to see how much variability was explained. PERMANOVA returned significant results for only Weighted Unifrac with $R^2 = 0.10618$ ($p > 0.05^*$) and not any other distance measures.

Key taxa associated with β -thalassemia

As illustrated in Figure 3G, the ten most abundant taxa were presented. Notably, there was a significant increase in the abundance of *Prevotella* in the thalassemia minor samples compared to the thalassemia major samples. Conversely, *Subdoligranulum* and *Dorea* exhibited substantial enrichment in the thalassemia major samples relative to the thalassemia minor samples.

Iron overload in β -thalassemia favours anaerobic bacterial taxa

The differential analysis based on ASVs was next performed to assess the differentially regulated taxa between the control, thalassemia minor and thalassemia major group using the DEseq2 tool. Microbial taxa found to be significantly upregulated in the control group as compared to the thalassemia minor group are primarily anaerobic bacteria affected by iron availability like members of *Coriobacteriaceae* family (*Coriobacteriaceae* ($padj < 0.05$), *Olsenella* ($padj < 0.001$), *Denitrobacterium* ($padj < 0.05$), *Eggerthella* ($padj < 0.05$), *Enterorhabdus* ($padj < 0.05$)) and *Rikenellaceae* family (*Alistipes* ($padj < 0.001$), *Rikenellaceae* ($padj < 0.001$)), *Bacteroides* ($padj < 0.05$), *Acintobacter* ($padj < 0.05$) and *Treponema* ($padj < 0.001$). Contrary to control group, genera belonging to family *Selenomonadaceae* (*Megamonas* ($padj = 4.959e-05$), *Mitsuokella* ($padj = 0.007$)) and different groups of *Eubacterium* ($padj < 0.05$,) and *Lachnospiracaceae* ($padj < 0.05$) showed increased abundance in the thalassemia minor group probably due to iron overload (Figure4A).

Figure 4B showed the differential abundance analysis of microbial taxa at the ASV level between the control and thalassemia major group. Different microbes exhibited high abundance in the control group like the families belonging to *Bacteroidetes* (*Muribaculaceae* ($padj < 0.05$), *Prevotellaceae* ($padj < 0.001$)), *Phascolarctobacterium* ($padj < 0.05$), and different ASVs of *Rikenellaceae* ($padj < 0.001$). Facultative anaerobic bacteria like *Weissella* ($padj < 0.05$), *Sarcina* ($padj < 0.05$), *Megamonas* ($padj < 0.001$), *Megasphaera* ($padj < 0.05$), *Succinivibrio*

($\text{padj} < 0.05$) and *Akkermansia* ($\text{padj} < 0.05$) showed predominant abundance in thalassemia major samples.

Figure 4C showed the differential abundance analysis between the thalassemia minor and major group. Different Gram positive anaerobic microbial taxa including *Olsenella* ($\text{padj} < 0.05$), *Alistipes* ($\text{padj} < 0.05$), *Sarcinia* ($\text{padj} < 0.001$), and *Lactobacillus* ($\text{padj} < 0.05$) were predominantly found in thalassemia major group. On the contrary Gram-negative anaerobic bacterial taxa *Prevotellaceae* ($\text{padj} < 0.001$), and *Butyrivibrio* ($\text{padj} < 0.001$) showed abundance in thalassemia minor group.

Key changes in potential metabolic function of microbial communities

DESeq2 analysis was performed in order to assess variability in MetaCyc pathways with β -thalassemia. Figure 5A shows the pathways like KdO2-lipidA biosynthesis pathway (LPSSYN-PWY, $\text{padj} < 0.001$) associated with lipopolysaccharide (LPS) biogenesis in Gram-negative bacteria, and L-glutamate degradation pathway (PWY-5088, $\text{padj} < 0.001$), significantly upregulated in thalassemia major group as compared to normal. L-glutamate degradation is highly sensitive to iron overload and generates α -ketoglutarate to feed cellular energy cycles. When iron overload occurs, it causes severe biochemical disruptions that directly link to L-glutamate degradation. However, in normal group pathways like Flavin biosynthesis II (PWY-6167, $\text{padj} < 0.001$), Mevalonate pathway (PWY-6174, $\text{padj} < 0.001$) and Isoprene biosynthesis pathway (PWY-7391, $\text{padj} < 0.05$) exhibited significant abundance. These three metabolic pathways intersect primarily in archaea (specifically methanogens and halophiles) and certain eubacteria (such as specific Gram-positive pathogens and environmental bacteria(29)).

Similar to the thalassemia major group, the thalassemia minor group showed abundance of the KdO2-lipidA biosynthesis pathway (LPSSYN-PWY, $\text{padj} < 0.001$). In comparison to thalassemia minor group, pathways like Super pathways fatty acid and unsaturated fatty acid biosynthesis (PWY-6285, $\text{padj} < 0.05$, PWY-6284, $\text{padj} < 0.05$), Super pathway of geranyl-geranyl-diphosphate biosynthesis I (via mevalonate) (PWY-5910, $\text{padj} < 0.056307$), and Mevalonate pathway (PWY-6174, $\text{padj} < 0.05$), showed the significant association to the normal group. These pathways do not just run parallel; they are metabolically and stoichiometrically dependent on one another. Acetyl-CoA, NADPH, and cellular redox state (FMN/FAD) link the mevalonate-isoprenoid pathways, the fatty acid biosynthesis pathways, and flavin biosynthesis (Figure 5B).

Figure 5C depicted the differential abundance of the metabolic pathways between the thalassemia major and minor group. The key metabolic pathways enriched in thalassemia minor group are Methanogenesis pathway ($\text{padj} < 0.001$), Coenzyme B biosynthesis pathway (PWY-P241, $\text{padj} < 0.001$), Factor 420 biosynthesis II pathway (PWY-5198, $\text{padj} < 0.001$), Flavin biosynthesis pathway (PWY-6167, $\text{padj} < 0.001$), Mevalonate pathway II (PWY-6174, $\text{padj} < 0.001$) whereas trans, octa-cis deca prenyl phosphate biosynthesis pathway (PWY-6383, $\text{padj} < 0.05$) was predominant in thalassemia major group.) trans, octa-cis deca prenyl phosphate biosynthesis pathway is a highly specialized isoprenoid lipid synthesis route exclusively present in Mycobacteria (such as *Mycobacterium tuberculosis*) and related Actinobacteria (30).

Microbiome associated metabolic signatures in β -thalassemia

A multi-omic approach using Data Integration Analysis for Biomarker discovery (DIABLO) was employed for integrating the 16S rRNA data and SCFA data in order to identify coherent microbiome associated metabolic signatures related to Thalassemia major, Thalassemia minor groups. Figure 6 panel A compares the beta-diversity/clustering of the sample groups based on two different data layers: 16S rRNA gene sequencing (microbial composition) and Short-Chain Fatty Acids (SCFAs) profiles (metabolic output). There is clear separation along Dimension 1 between the beta-thalassemia Major group (red circles) and the beta-thalassemia Minor group (green triangles). Both the Major and Minor groups form relatively distinct, well-defined clusters with minimal overlap between them. The Normal controls (blue crosses) show substantial dispersion along Dimension 2 and spread across a wider coordinate space, indicating higher inter-individual variation in their baseline gut microbial profile.

Unlike the bacterial taxonomic composition, the SCFA metabolomic profiles show significant overlap among all three clinical groups (Major, Minor, and Normal) indicating overlapping metabolic features. The clustering suggests that despite distinct differences in taxonomic bacterial composition (seen in the 16S plot), the overall short-chain fatty acid metabolic output remains relatively conserved or shifted only subtly across the cohorts.

Panel B displayed a scatter plot showing the relationship between the 16S rRNA and SCFA data for each sample, with lines connecting the corresponding data points in both datasets. Asterisks represent centroids for each group. Thalassemia minor group (shown in orange) clusters tightly in the upper-left quadrant, displaying relative consistency in both taxonomic composition and SCFA profiles. Thalassemia major group (shown in grey) occupied the upper-right region with longer, highly variable vector trajectories, reflecting elevated heterogeneity

in how taxonomic changes map to SCFA outputs in patients with beta-thalassemia major. The normal group (shown in blue) separates towards the lower-left and shows moderate vector lengths, indicating distinct baseline dynamics compared to diseased/carrier group. Panel B showed distinct clustering of both SCFA and 16SrRNA among all the three groups i.e. Normal, Thalassemia Minor, and Thalassemia Major when taking into account both microbiome and SCFA metabolic shifts.

Loading plots Figure 7 shows distinct microbial signatures and SCFAs within the normal, thalassemia major and thalassemia minor groups. *Dorea* and *Megasphaera* were positively associated with Thalassemia major and valeric acid (pentanoic acid C5) SCFA abundance (Figure 10A). Loading plots (Figure 10B) displayed negative association of butyrate (C4), isobutyrate (IC4) and heptanoic acid with the Thalassemia minor trait with no significant microbial shifts indicating a shift only at metabolite level. Meanwhile, *Elusimicroblum*, *Oscillospiraceae* were more abundant in normal samples Figure 10C and showed a strong negative correlation of Caprylic acid (C8-octanoic acid) with *Enterorhabdus* genus in the normal samples (Figure 6). DIABOLO correlation Circos plot depicted positive association of *Megasphaera* with Caprylic acid (C8-octanoic acid) whereas *Bifidobacterium* is positively associated with butyrate (C4) and *Rikenellaceae-RC-9* gut group negatively correlated with the acetate (C2) abundance in the normal cohort. *Collinsella* is positively correlated with butyrate (C4) abundance in the Thalassemia major group. *Anaerostipes* negatively correlated with acetate (C2) abundance in the Thalassemia minor cohort (Figure 8).

Discussion

The key findings of this study provide a comprehensive outlook on how the severity of beta-thalassemia from minor to major can remodel the gut microbiome along with its associated metabolic output. This shift seems to be driven by a combination of factors like host physiological changes like iron overload followed by the subsequent competitive advantage given to certain taxa. The higher microbial richness and Shannon diversity observed in the Thalassemia Minor group reflects a unique ecological transition. The slight shift in iron homeostasis and compensatory erythropoiesis prevented the dominance of single species, thereby increasing overall richness. Conversely, in the case of Thalassemia Major group, a significant decline in microbial diversity due to severe iron overload was observed. The significant enrichment of *Prevotella* in the Thalassemia Minor group suggests an intact and robust mucosal barrier. In contrast, the shift towards *Subdoligranulum* and *Dorea* in the Thalassemia Major group, with overall reduction in microbial diversity can be linked to

metabolic dysfunction. *Subdoligranulum*, is a butyrate producer and its enrichment can serve as a compensatory mechanism against oxidative stress in the Thalassemia Major group (31). The enrichment of *Akkermansia* in the Thalassemia Major group suggests a compensatory reliance on host mucins as carbon source when primary fibre degrading pathways are compromised(32). Iron seems to be the key player in regulating the microbial diversity. In Normal group, predominance of anaerobic bacteria like *Alistipes* was observed. β -thalassemia or high-dose iron supplementation—demonstrate a direct negative correlation between elevated systemic/luminal iron levels (high ferritin) and the relative abundance of *Alistipes*(33). However, in the case of Thalassemia Minor and Major group increase in abundance of *Megamonas* and *Sarcinia* was observed. The presence of high systemic and luminal iron in the Thalassemia Major group results in increase in abundance of facultative anaerobes and iron tolerant taxa that are capable of sequestering or utilizing iron for their growth(34). The metabolic pathway analysis highlights a transition from healthy to pro-inflammatory pathways. The significant upregulation in the KdO2-lipidA biosynthesis pathway (LPSSYN-PWY) in both the Thalassemia Minor and Major group can lead to heightened systemic inflammation, often termed as “metabolic endotoxemia” as Lipid A is an active component of the potent endotoxin Lipopolysaccharides (LPS)(35).

This endotoxin enrichment in the thalassemic gut may drive chronic, low grade inflammation as the endotoxins may translocate across the weak mucosal barriers. However, the normal group displayed robust nutritional and metabolic homeostasis by its unique association with the Mevalonate and Flavin biosynthesis pathways that are critical for the production of isoprenoids and vitamins like B2 (36). Depletion of these pathways in the thalassemic cohorts may lead to systemic oxidative stress. Moreover, the integration of the 16SrRNA and short chain fatty acid The positive correlation between *Bifidobacterium* and butyrate (C4) points towards intestinal barrier integrity and well-regulated gut environment (37).

Negative association of butyrate and iso-butyrate with the Thalassemia Minor group, indicates compromised metabolic health. Since Butyrate is arguably the most critical molecule for maintaining a healthy gastrointestinal tract produced by beneficial anaerobic bacteria (such as *Faecalibacterium prausnitzii*, *Roseburia*, and *Eubacterium rectale*) through the fermentation of dietary fibers and resistant starches(38). In β -thalassemia, the relationship between butyrate and iron overload represents a critical metabolic axis. Patients suffer from systemic iron overload driven by two concurrent pressures: lifelong blood transfusions and chronically hyper-activated intestinal iron absorption caused by the suppression of the hormone

hepcidin. The unabsorbed free iron floods the large intestine, creating an environment that depletes butyrate-producing bacteria and drives systemic complications. This vicious cycle dumps excess iron in gut initiating a pathophysiological cascade that reduces abundance of sensitive obligate anaerobes involved in butyrate production. The identification of a *Dorea-Megasphaera*-Valeric acid axis in Thalassemia Major provides a distinct multi-omic signature that could potential be used for therapeutic intervention. *Megasphaera* and *Dorea* are two distinct genera of gut bacteria that interact through a process called microbial cross-feeding to drive the production of valeric acid(39). Valeric acid is an emerging ‘postbiotic’ molecule. It plays a critical role in gut health by serving as a potent histone deacetylase inhibitor. This specific epigenetic action offers unique therapeutic advantages for β -thalassemic pateints by mitigating iron-induced inflammation and preventing lipid peroxidation and ferroptosis (40).

Based on these findings, the therapeutic interventions should address the Iron-redox inflammation triad. Strains of *Bifidobacterium* and *Lactobacillus* with low iron scavenging potential may help restore the SCFA barrier (41). Management of the oxidative stress can be accomplished by supplementing the Flavin/Mevalonate pathways. The Lipid A payload can be reduced by targeting prebiotics that can outcompete *Megasphaera* (42). However, further experimental work is required to prove the efficacy of these plausible therapies.

Conclusion

Our results reveal that β -thalassemia was associated with significant changes in the configuration of the intestinal bacterial axis, leading to an altered metabolic profile, which partially contributes to the inflammatory status characteristic of the disease. In particular, although thalassemia minor was characterized by a decreased abundance of several bacterial genera, the overall bacterial diversity and functionality remained relatively high in these individuals. In contrast, thalassemia major was linked with diminished bacterial diversity and a shift in metabolic profiles, possibly due to iron-related mechanisms. Thus, there was a reduced abundance of SCFAs and an increased presence of pro-inflammatory pathways, such as those contributing to lipopolysaccharide production. The altered microbiota profile is strongly linked with dysbiosis, which can be reversed by a carefully designed probiotic or prebiotic treatment. Such treatment could restore the integrity of the intestinal barrier by promoting the growth of valeric acid-producing and butyrate-producing bacteria while decreasing the abundance of iron-dependent pathogens. It could serve as an effective adjunctive therapy to standard iron chelation treatment, diminishing the inflammatory burden, as well as the severity of iron-overload-induced tissue damage

and barrier dysfunction. Furthermore, restoring the butyrate-producing capacity of the microbiota via iron-resistant probiotics would diminish the pathogenicity of iron-overload associated tissue damage and leaky gut.

Declarations

Ethical Approval

The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Review Board at COMSATS University Islamabad (CUI/Bio/ERB/2021/60). Written informed consent of the participants (guardians) was obtained before data collection. Clinical histories and blood transfusion records were obtained for all participants, who were aged between 4 and 17 years old.

Acknowledgements

We would like to thank Mr. Aftab Hussain AVM(R), Director and the staff at Sundas Foundation Pakistan for helping with acquiring and handling the samples. We acknowledge all the participants of the study.

Data availability statement

The survey data for Figure 2 is given as Supplementary_Data_Table_S1.csv whilst the raw sequence files supporting the results of this article are available in the European Nucleotide Archive under the project accession number PRJEB77861 available at <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJEB77861> with details given in Supplementary_Data_Table_S2.csv

Competing Interest

The authors declare that they have no competing interests.

Funding

Umer Zeeshan Ijaz is supported by UKRI MRC Grant No. MR/Z50628X/1.

Author Contribution

M.B.J.: Data Curation; Methodology; Sample Collection and Processing; Investigation; Formal Analysis; Writing - Review & Editing

H.K.: Investigation; Sample Collection; Analysis; Writing - Review & Editing

I.A.: Investigation; Sample Collection; Writing - Review & Editing

J.E.: Investigation; Sample Collection; Writing - Review & Editing

A.A.: Data Curation; Writing - Review & Editing

K.G.: Resources; Writing - Review & Editing

Z.N.: Data Curation; Writing - Review & Editing

S.J.: Data Curation; Writing - Review & Editing

U.Z.I.: Data Curation; Methodology, Conceptualization; Visualization; Software; Formal Analysis; Supervision; Resources; Funding acquisition; Writing - Review & Editing

A.T.: Data Curation; Methodology; Conceptualization; Visualization; Supervision; Project administration; Resources; Writing - Original Draft

References

1. Shafique F, Ali S, Almansouri T, Van Eeden F, Shafi N, Khalid M, et al. Thalassemia, a human blood disorder. *Braz J Biol.* 2021 Sep 3;83:e246062. doi:<https://doi.org/10.1590/1519-6984.246062>
2. Nonejuie P, Wilantho A, McDonald D, Htoo HH, Chalerm J, Tripathi A, et al. Differential gut microbiota composition in β -Thalassemia patients and its correlation with iron overload. *Sci Rep.* 2024 Oct 11;14:23858. doi:10.1038/s41598-024-75456-4 PubMed PMID: 39394230; PubMed Central PMCID: PMC11470119.
3. Fucharoen S, Winichagoon P. New updating into hemoglobinopathies. *Int J Lab Hematol.* 2012;34(6):559–65. doi:10.1111/j.1751-553X.2012.01446.x
4. Forni GL, Grazzini G, Boudreaux J, Agostini V, Omert L. Global burden and unmet needs in the treatment of transfusion-dependent β -thalassemia. *Front Hematol.* 2023 Jun 20;2. doi:10.3389/frhem.2023.1187681
5. Ansari AH, Ansari SH, Salman MJ, Ansari MUH, Jabeen R. A scoping review on the obstacles faced by beta thalassemia major patients in Pakistan- Matter of policy investment. *AIMS Public Health.* 2024 Nov 12;11(4):1105–24. doi:10.3934/publichealth.2024057 PubMed PMID: 39802567; PubMed Central PMCID: PMC11717546.
6. Chauhan W, Shoaib S, Fatma R, Zaka-ur-Rab Z, Afzal M. Beta-thalassemia and the advent of new interventions beyond transfusion and iron chelation. *Br J Clin Pharmacol.* 2022;88(8):3610–26. doi:10.1111/bcp.15343
7. Guo X, Lin S, Zhang X, Li M, Wang Z, Peng Y, et al. Integrated metabolomic and microbiome analysis identifies *Cupriavidus metallidurans* as a potential therapeutic target for β -thalassemia. *Ann Hematol.* 2024 Dec 1;103(12):5169–79. doi:10.1007/s00277-024-06016-z
8. Jaeggi T, Kortman GAM, Moretti D, Chassard C, Holding P, Dostal A, et al. Iron fortification adversely affects the gut microbiome, increases pathogen abundance and induces intestinal inflammation in Kenyan infants. *Gut.* 2015 May 1;64(5):731–42. doi:10.1136/gutjnl-2014-307720 PubMed PMID: 25143342.
9. Paganini D, Uyoga MA, Zimmermann MB. Iron Fortification of Foods for Infants and Children in Low-Income Countries: Effects on the Gut Microbiome, Gut Inflammation, and Diarrhea. *Nutrients.* 2016 Aug 12;8(8):494. doi:10.3390/nu8080494 PubMed PMID: 27529276; PubMed Central PMCID: PMC4997407.
10. Visitchanakun P, Panpetch W, Saisorn W, Chatthanathon P, Wannigama DL, Thim-uam A, et al. Increased susceptibility to dextran sulfate-induced mucositis of iron-overload β -thalassemia mice, another endogenous cause of septicemia in thalassemia. *Clin Sci.* 2021 Jun 16;135(12):1467–86. doi:10.1042/CS20210328
11. Yang W, Cong Y. Gut microbiota-derived metabolites in the regulation of host immune responses and immune-related inflammatory diseases. *Cell Mol Immunol.* 2021 Apr;18(4):866–77. doi:10.1038/s41423-021-00661-4 PubMed PMID: 33707689; PubMed Central PMCID: PMC8115644.

12. Kelly CJ, Zheng L, Campbell EL, Saeedi B, Scholz CC, Bayless AJ, et al. Crosstalk between Microbiota-Derived Short-Chain Fatty Acids and Intestinal Epithelial HIF Augments Tissue Barrier Function. *Cell Host Microbe*. 2015 May 13;17(5):662–71. doi:10.1016/j.chom.2015.03.005 PubMed PMID: 25865369.
13. Obed FA, Omran AM, Jebur KS. Effect of iron overload on prevalence of common bacterial infection in thalassemia patients. *Iraqi J Hematol*. 2024 Jun;13(1):38. doi:10.4103/ijh.ijh_84_23
14. Salvi PS, Cowles RA. Butyrate and the Intestinal Epithelium: Modulation of Proliferation and Inflammation in Homeostasis and Disease. *Cells*. 2021 Jul;10(7):1775. doi:10.3390/cells10071775
15. Taher AT, Weatherall DJ, Cappellini MD. Thalassaemia. *The Lancet*. 2018 Jan;391(10116):155–67. doi:10.1016/S0140-6736(17)31822-6
16. Al-Abidy LH, Al-Aidy SRA. Molecular Study and Detection of The Causative Bacteria of Urinary Tract Infection of Transfusion Dependent Thalassemia Patients Who Are Being Treated with Iron Chelators: *Acad Open*. 2025 Feb 4;10(1):1. doi:10.21070/acopen.10.2025.10567
17. Chen G, Li Y, Wei S, Wang X, Kuang Z, Guo W, et al. Role of gut microbiota in thalassemia: a review of therapeutic prospects. *Front Physiol*. 2025 Mar 19;16. doi:10.3389/fphys.2025.1523448
18. Sadiq IZ, Abubakar FS, Usman HS, Abdullahi AD, Ibrahim B, Kastayal BS, et al. Thalassemia: Pathophysiology, Diagnosis, and Advances in Treatment. *Thalass Rep*. 2024 Dec;14(4):81–102. doi:10.3390/thalassrep14040010
19. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform. *Appl Environ Microbiol*. 2013 Sep;79(17):5112–20. doi:10.1128/AEM.01043-13 PubMed PMID: 23793624; PubMed Central PMCID: PMC3753973.
20. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High resolution sample inference from Illumina amplicon data. *Nat Methods*. 2016 Jul;13(7):581–3. doi:10.1038/nmeth.3869 PubMed PMID: 27214047; PubMed Central PMCID: PMC4927377.
21. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol*. 2019 Aug;37(8):852–7. doi:10.1038/s41587-019-0209-9 PubMed PMID: 31341288; PubMed Central PMCID: PMC7015180.
22. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*. 2013 Jan;41(Database issue):D590–6. doi:10.1093/nar/gks1219 PubMed PMID: 23193283; PubMed Central PMCID: PMC3531112.
23. Douglas GM, Maffei VJ, Zaneveld JR, Yurgel SN, Brown JR, Taylor CM, et al. PICRUSt2 for prediction of metagenome functions. *Nat Biotechnol*. 2020 Jun;38(6):685–8.

doi:10.1038/s41587-020-0548-6 PubMed PMID: 32483366; PubMed Central PMCID: PMC7365738.

24. Dixon P. VEGAN, a package of R functions for community ecology. *J Veg Sci.* 2003;14(6):927–30. doi:10.1111/j.1654-1103.2003.tb02228.x
25. McMurdie PJ, Holmes S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. Watson M, editor. *PLoS ONE.* 2013 Apr 22;8(4):e61217. doi:10.1371/journal.pone.0061217
26. Zhang Y, Jing G, Chen Y, Li J, Su X. Hierarchical Meta-Storms enables comprehensive and rapid comparison of microbiome functional profiles on a large scale using hierarchical dissimilarity metrics and parallel computing. *Bioinforma Adv.* 2021 May 12;1(1):vbab003. doi:10.1093/bioadv/vbab003 PubMed PMID: 36700101; PubMed Central PMCID: PMC9710644.
27. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550. doi:10.1186/s13059-014-0550-8 PubMed PMID: 25516281; PubMed Central PMCID: PMC4302049.
28. Rohart F, Gautier B, Singh A, Lê Cao KA. mixOmics: An R package for ‘omics feature selection and multiple data integration. Schneidman D, editor. *PLOS Comput Biol.* 2017 Nov 3;13(11):e1005752. doi:10.1371/journal.pcbi.1005752
29. Miziorko HM. Enzymes of the mevalonate pathway of isoprenoid biosynthesis. *Arch Biochem Biophys.* 2011 Jan 15;505(2):131–43. doi:10.1016/j.abb.2010.09.028
30. Crick DC, Mahapatra S, Brennan PJ. Biosynthesis of the arabinogalactan-peptidoglycan complex of *Mycobacterium tuberculosis*. *Glycobiology.* 2001 Sep;11(9):107R-118R. doi:10.1093/glycob/11.9.107r PubMed PMID: 11555614.
31. Van Hul M, Le Roy T, Prifti E, Dao MC, Paquot A, Zucker JD, et al. From correlation to causality: the case of Subdoligranulum. *Gut Microbes.* 12(1):1849998. doi:10.1080/19490976.2020.1849998 PubMed PMID: 33323004; PubMed Central PMCID: PMC7744154.
32. Suparan K, Trirattanapa K, PiriyaKhuntorn P, Sriwichaiin S, Thonusin C, Nawara W, et al. Exploring alterations of gut/blood microbes in addressing iron overload-induced gut dysbiosis and cognitive impairment in thalassemia patients. *Sci Rep.* 2024 Oct 23;14(1):24951. doi:10.1038/s41598-024-76684-4
33. Nonejuie P, Wilantho A, McDonald D, Htoo HH, Chalerm J, Tripathi A, et al. Differential gut microbiota composition in β -Thalassemia patients and its correlation with iron overload. *Sci Rep.* 2024 Oct 11;14(1):23858. doi:10.1038/s41598-024-75456-4
34. Putri GSA, Ethica SN, Saptaningtyas R. Proceedings of the 2nd Lawang Sewu International Symposium on Health Sciences: Medical Laboratory Technology (LSISHS-MLT 2023). Springer Nature; 2024. 181 p.
35. Yan Y, Zhang W, Wang Y, Yi C, Yu B, Pang X, et al. Crosstalk between intestinal flora and human iron metabolism: the role in metabolic syndrome-related comorbidities and its

potential clinical application. *Microbiol Res.* 2024 May;282:127667. doi:10.1016/j.micres.2024.127667

36. Averianova LA, Balabanova LA, Son OM, Podvolotskaya AB, Tekutyeva LA. Production of Vitamin B2 (Riboflavin) by Microorganisms: An Overview. *Front Bioeng Biotechnol.* 2020 Nov 12;8. doi:10.3389/fbioe.2020.570828
37. Chen J, Vitetta L. The Role of Butyrate in Attenuating Pathobiont-Induced Hyperinflammation. *Immune Netw.* 2020 Feb 4;20(2):e15. doi:10.4110/in.2020.20.e15 PubMed PMID: 32395367; PubMed Central PMCID: PMC7192831.
38. Leonel A, Alvarez-Leite J. Leonel AJ, Alvarez-Leite JL. Butyrate: implications for intestinal function. *Curr Opin Clin Nutr Metab Care* 15: 474-479. *Curr Opin Clin Nutr Metab Care.* 2012 Jul 12;15:474–9. doi:10.1097/MCO.0b013e32835665fa
39. Yuille S, Reichardt N, Panda S, Dunbar H, Mulder IE. Human gut bacteria as potent class I histone deacetylase inhibitors in vitro through production of butyric acid and valeric acid. *PLOS ONE.* 2018 Jul 27;13(7):e0201073. doi:10.1371/journal.pone.0201073
40. Ma X, Shi W, Wang Z, Li S, Ma R, Zhu W, et al. Butyric acid and valeric acid attenuate stress-induced ferroptosis and depressive-like behaviors by suppressing hippocampal neuroinflammation. *J Transl Med.* 2025 Sep 2;23:974. doi:10.1186/s12967-025-06950-0 PubMed PMID: 40898211; PubMed Central PMCID: PMC12403447.
41. Rusu IG, Suharoschi R, Vodnar DC, Pop CR, Socaci SA, Vulturar R, et al. Iron Supplementation Influence on the Gut Microbiota and Probiotic Intake Effect in Iron Deficiency—A Literature-Based Review. *Nutrients.* 2020 Jul;12(7):1993. doi:10.3390/nu12071993
42. Toe LC, Kerckhof FM, De Bodt J, Morel FB, Ouedraogo JB, Kolsteren P, et al. A prebiotic-enhanced lipid-based nutrient supplement (LNSp) increases *Bifidobacterium* relative abundance and enhances short-chain fatty acid production in simulated colonic microbiota from undernourished infants. *FEMS Microbiol Ecol.* 2020 Jul 1;96(7):fiae105. doi:10.1093/femsec/fiae105

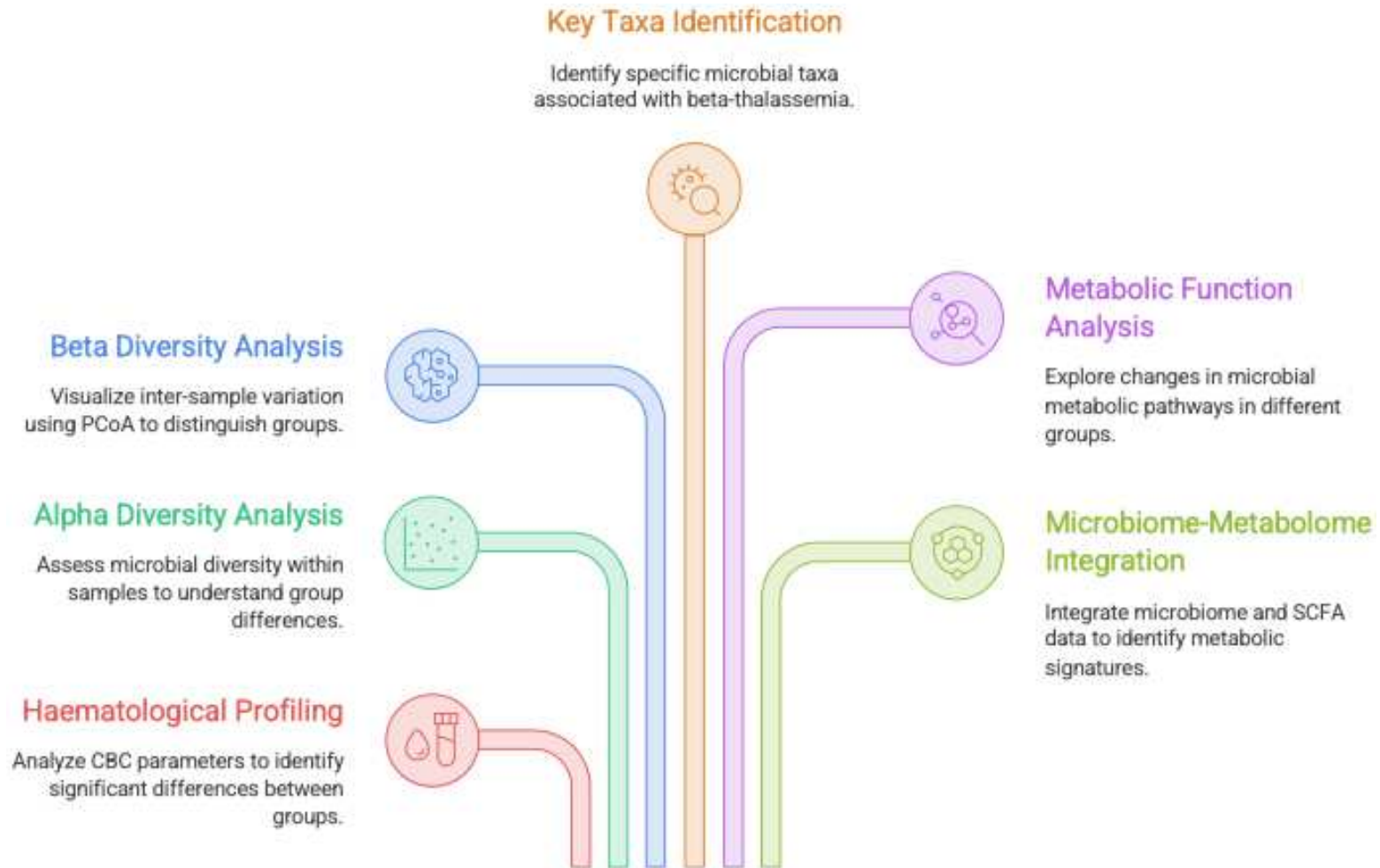


Figure 1: Technical road map of the study is illustrated using the free version of Napkin.AI (<https://app.napkin.ai/>).

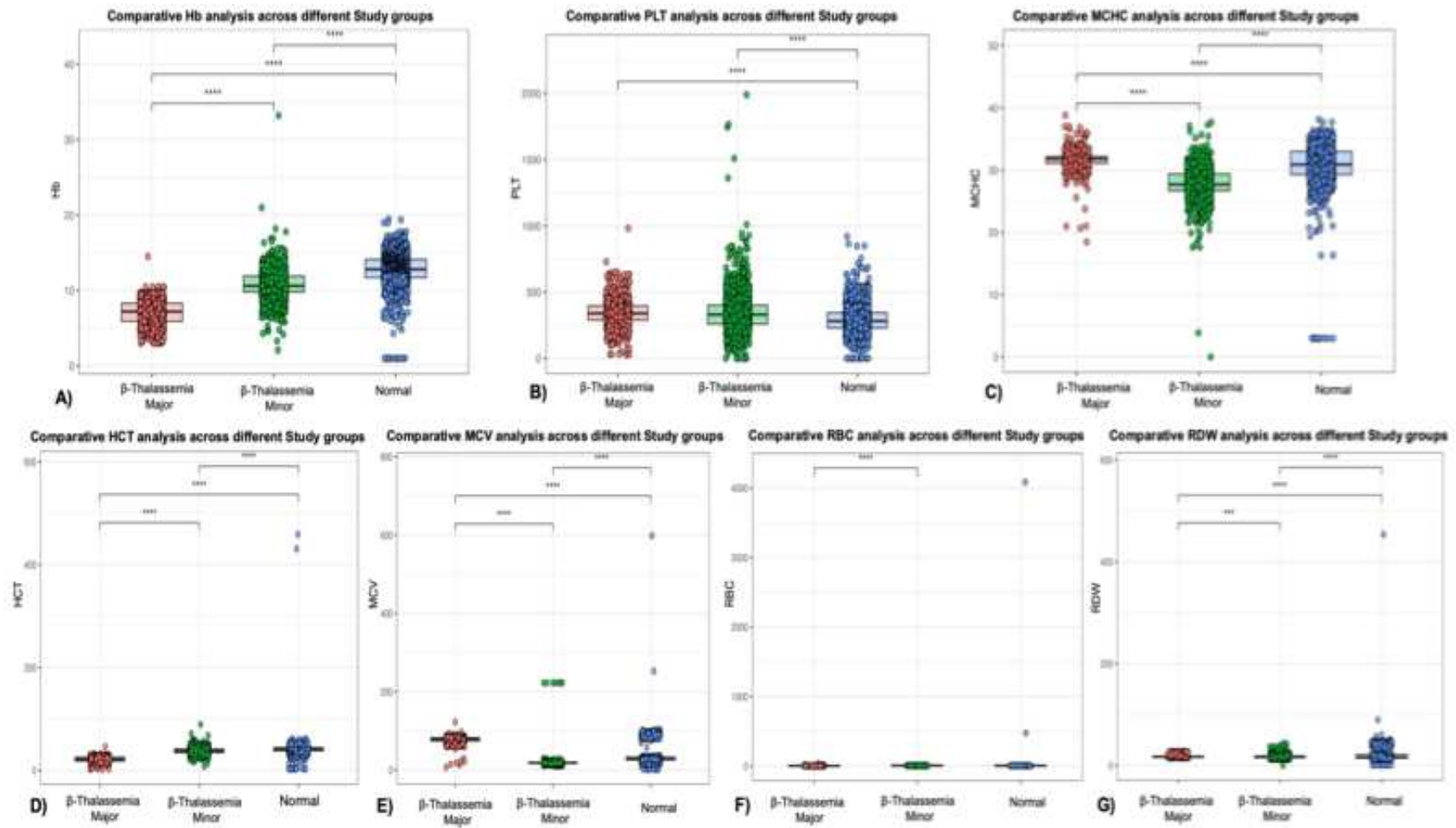


Figure 2: Comparative hemotological parameters (A) haemoglobin (Hb), (B) platelet count (PLT), (C) mean corpuscular haemoglobin concentration (MCHC), (D) haematocrit (HCT), (E) mean corpuscular volume (MCV), (F) mean red blood cell (RBC) count (G) and red blood cell distribution width (RDW) among Normal (N=4), β thalassemia major (N=14) and β thalassemia minor patient samples (N=7).

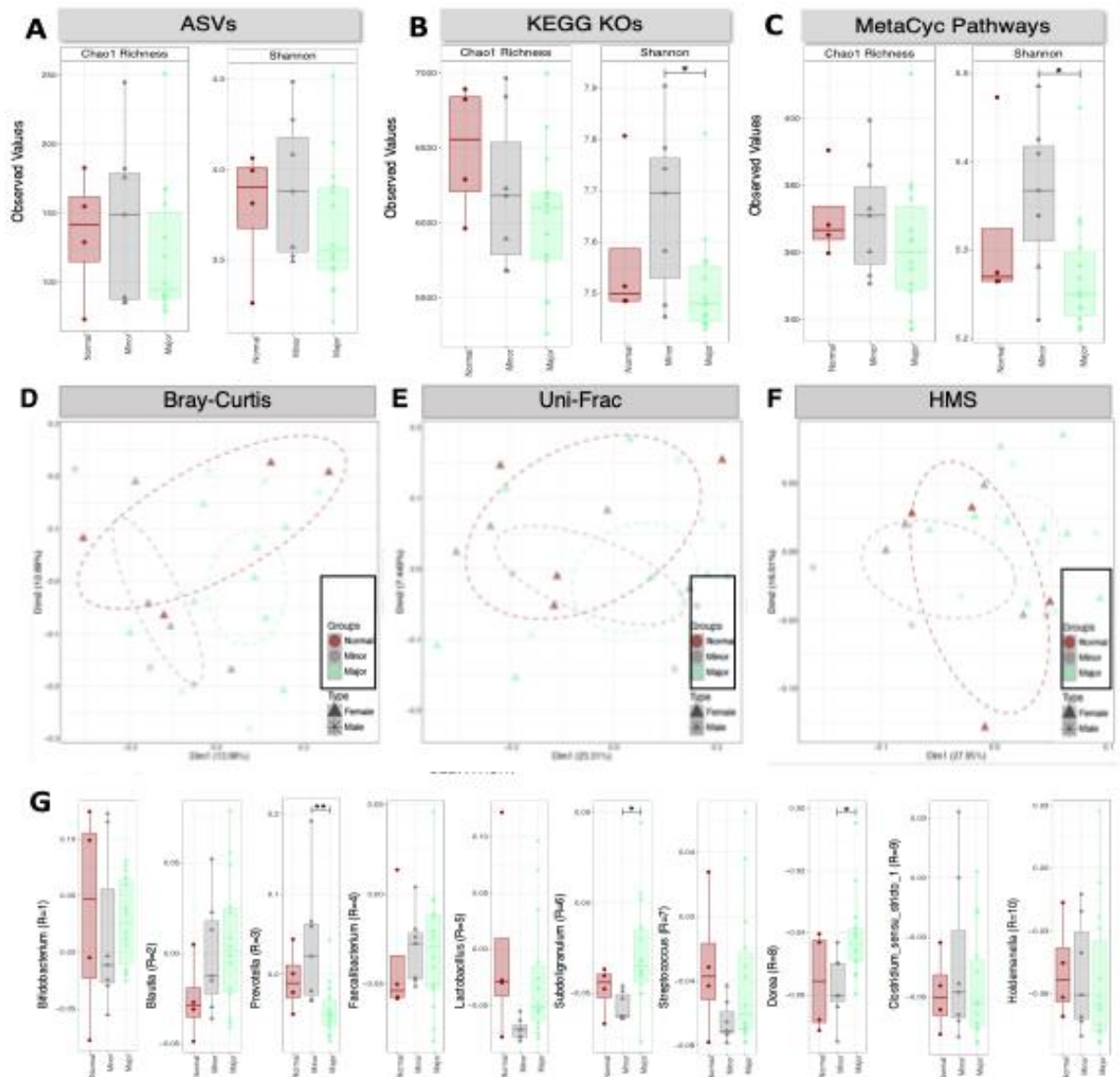


Figure 3: Alpha diversity (Chao 1 Richness and *Shannon* entropy) comparison of (A) ASVs, as well as (B) KEGG KOs, and (C) MetaCyc pathways predicted from the PICRUSt2 software. Beta diversity (D-F) represented by principal coordinate analysis (PCoA) plots with each axis showing the percentage variability explained by that axis, and where ellipses represent 95% confidence interval of the standard error for a given group. PERMANOVA statistics utilising these distance measures are shown underneath the PCoA plots to suggest if the beta diversity differences are significant. (G) shows the TSS+CLR normalized expressions of top 10 most abundant taxa observed in this dataset. The solid lines in panels with boxplots connect groups if ANOVA is significant with significance values as: * $p < 0.05$, ** $p < 0.01$, or *** $p < 0.001$.

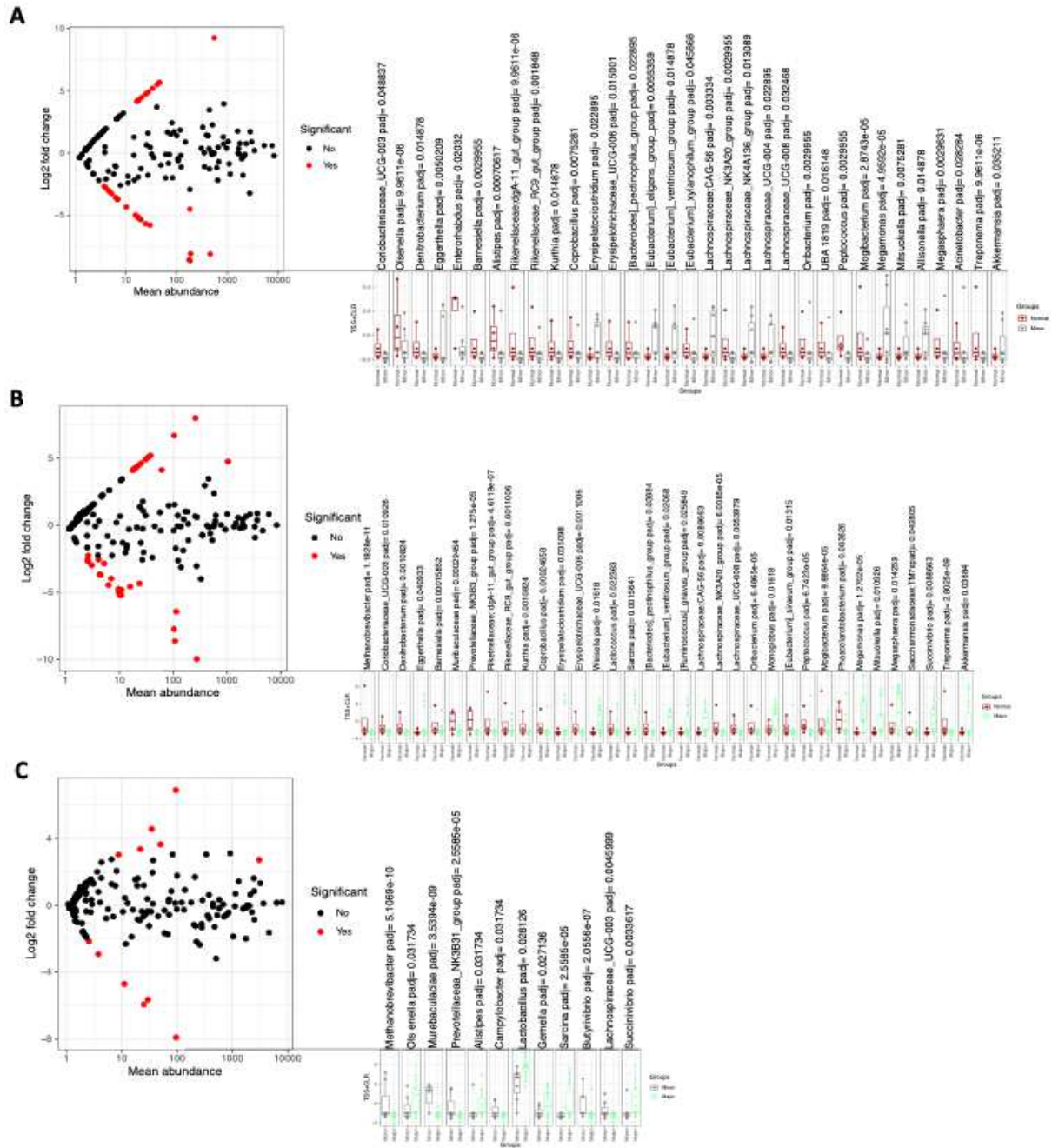


Figure 4: Differential Abundance Analysis of Microbial Genera using DESeq2. (A) Comparison between the Normal and the Thalassaemia minor cohort, (B) Comparison between the *Normal* and the Thalassaemia major cohort, and (C) Comparison between the Thalassaemia *Minor* and the Thalassaemia *Major* cohort. In these panels left side shows the MA plot where the significant features are represented as red dots along with the expression plot (TSS+CLR normalised) of the significant features shown on the right side.

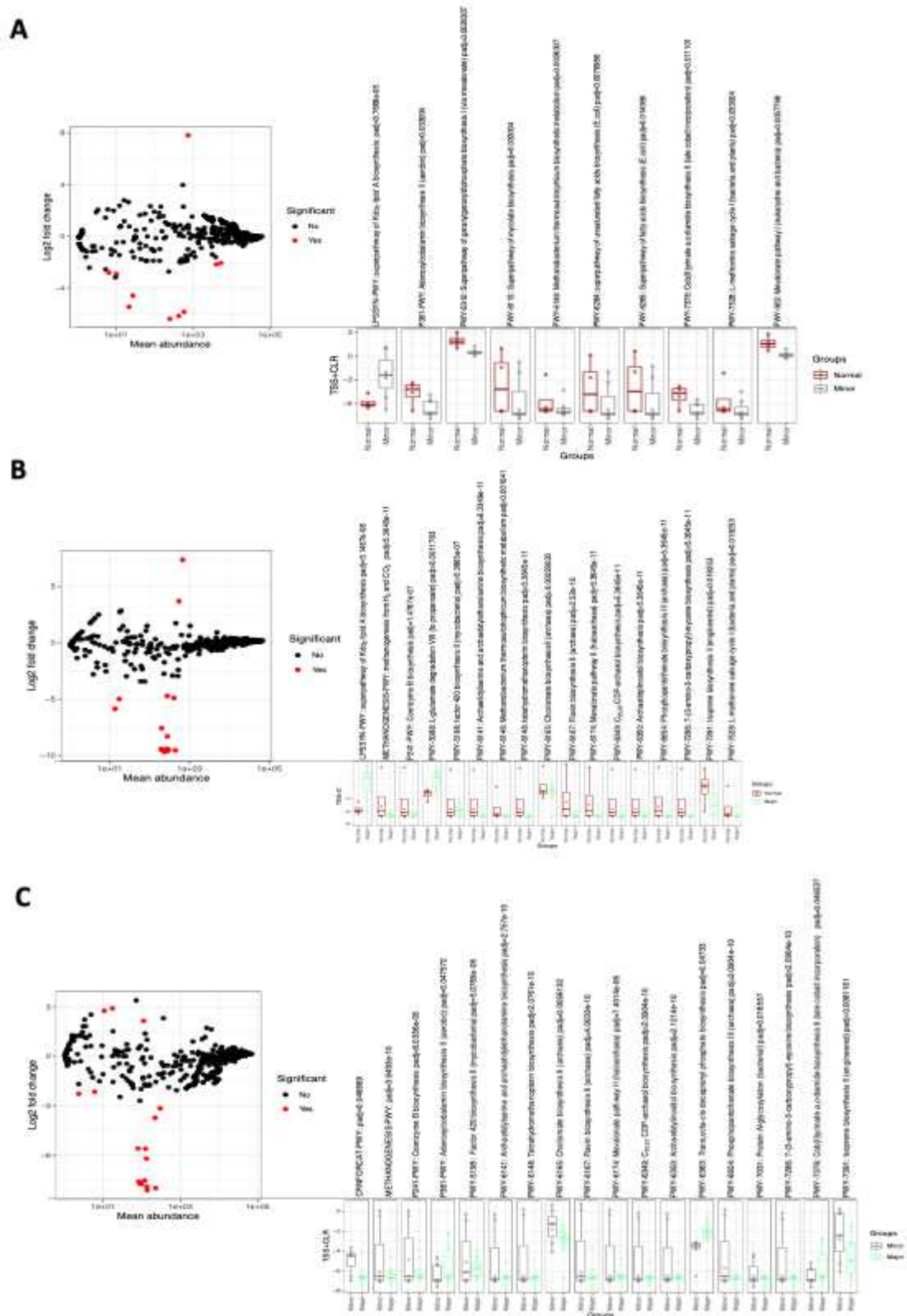


Figure 5: Differential Abundance Analysis of Microbial Metabolic Potential using DESeq2. (A) Comparison between the Normal and the Thalassemia minor cohort, (B) Comparison between the *Normal* and the Thalassemia major cohort, and (C) Comparison between the Thalassemia *Minor* and the Thalassemia *Major* cohort. In these panels left side shows the MA plot where the significant features are represented as red dots along with the expression plot (TSS+CLR normalised) of the significant features shown on the right side.

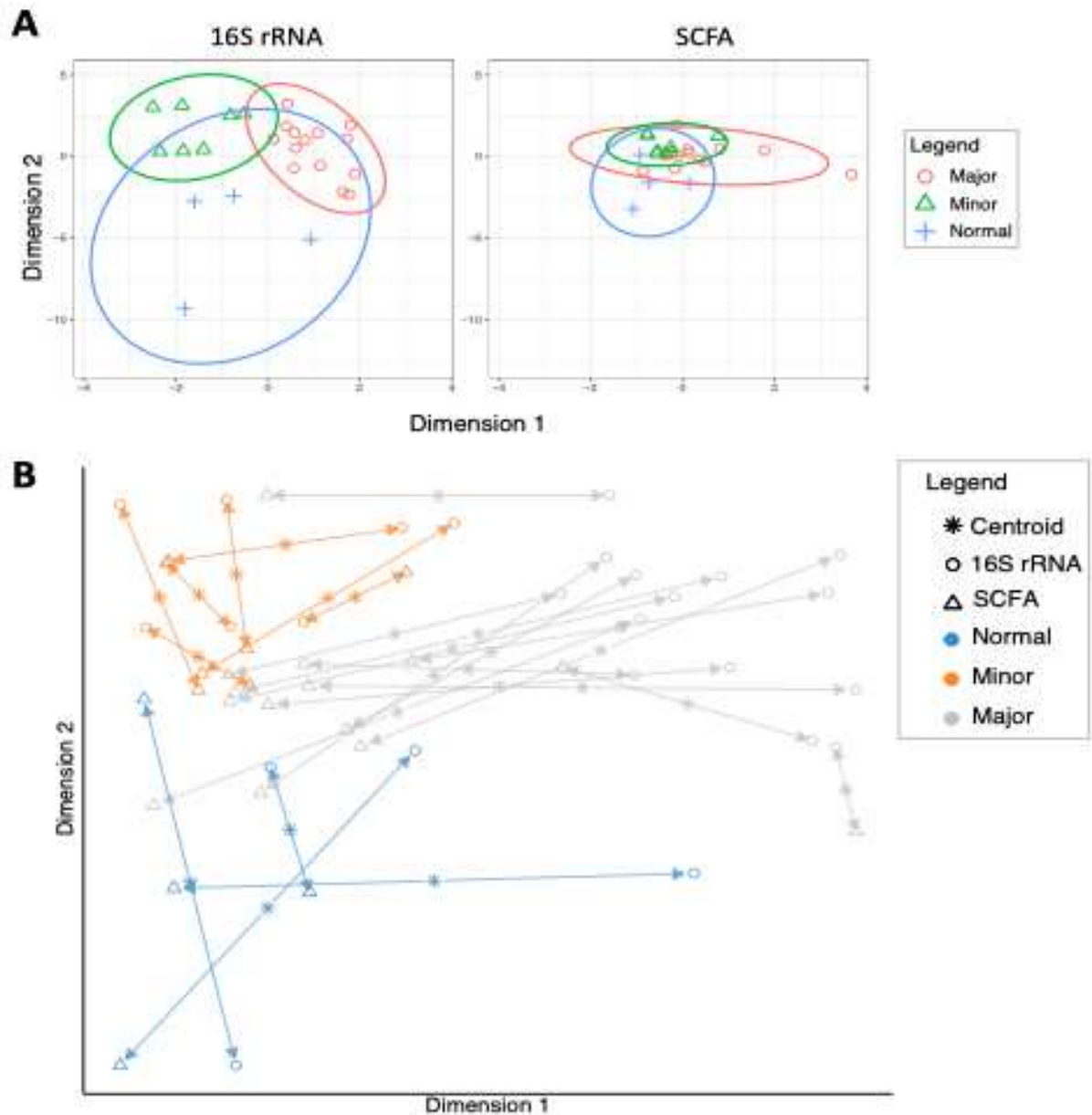


Figure 6: DIABLO analysis comparing changes in the stool microbiota and SCFAs between the *Normal*, Thalassemia *Minor* and the Thalassemia *Major* cohorts. The algorithm found three components, and of these, the samples are shown in the first two components, both separately (A), and then together (B) on a reduced discriminating feature set for both 16S rRNA and SCFA tables. The ellipses represent 95% confidence interval in (A). In (B), the asterisks * indicate the centroid between 16S rRNA and SCFA data for a given sample and the tip of the arrow represent the location of each same sample whether in the 16S rRNA or the SCFA space.

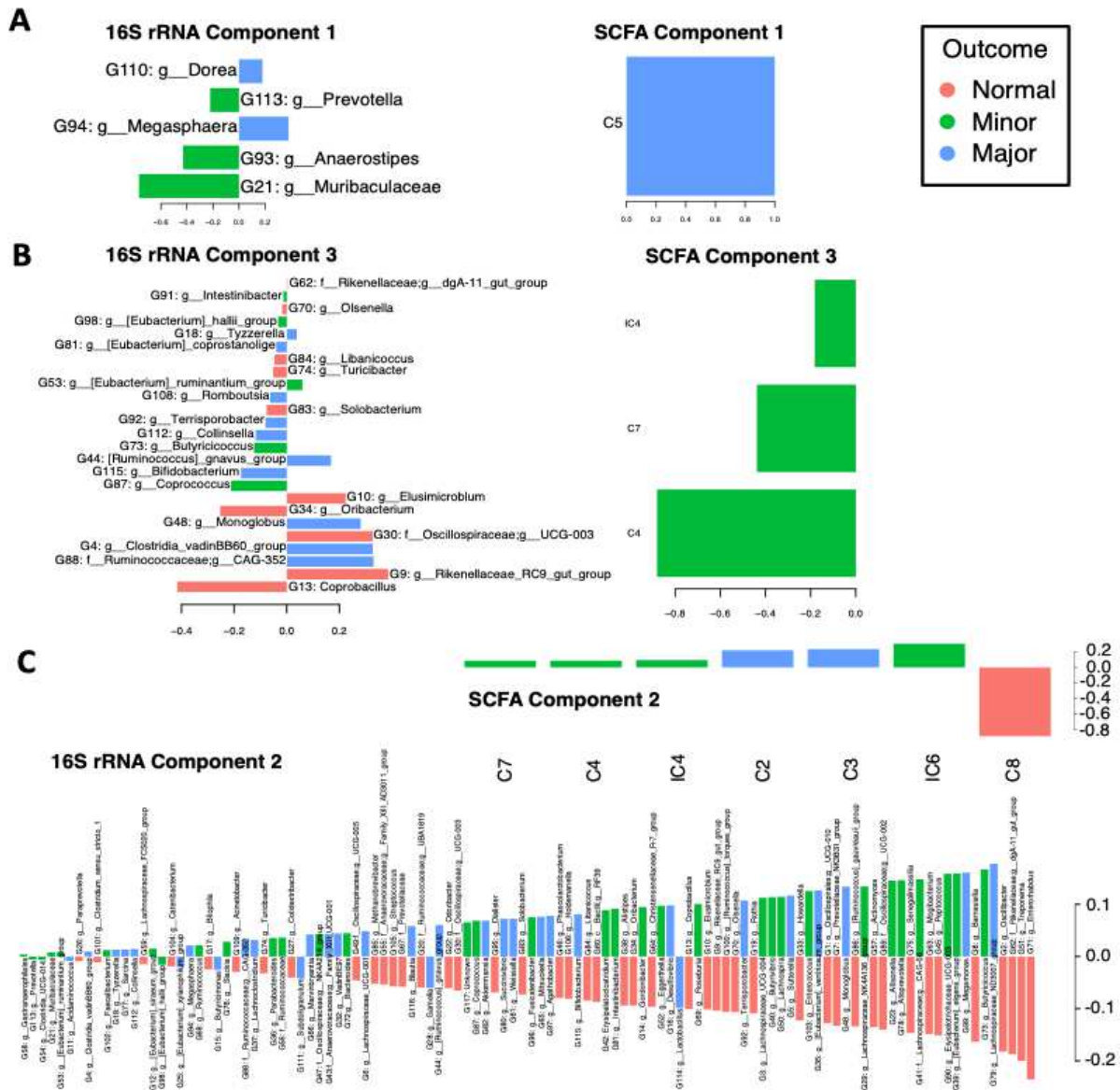


Figure 7: The coefficients of the non-zero (significant) loading components returned from the DIABLO algorithm. Note that the colour represents the groups (Normal, Thalassemia *Minor*, and Thalassemia *Major*) in which the features have the maximum expression/abundance with the length of the bars representing how important these features are.

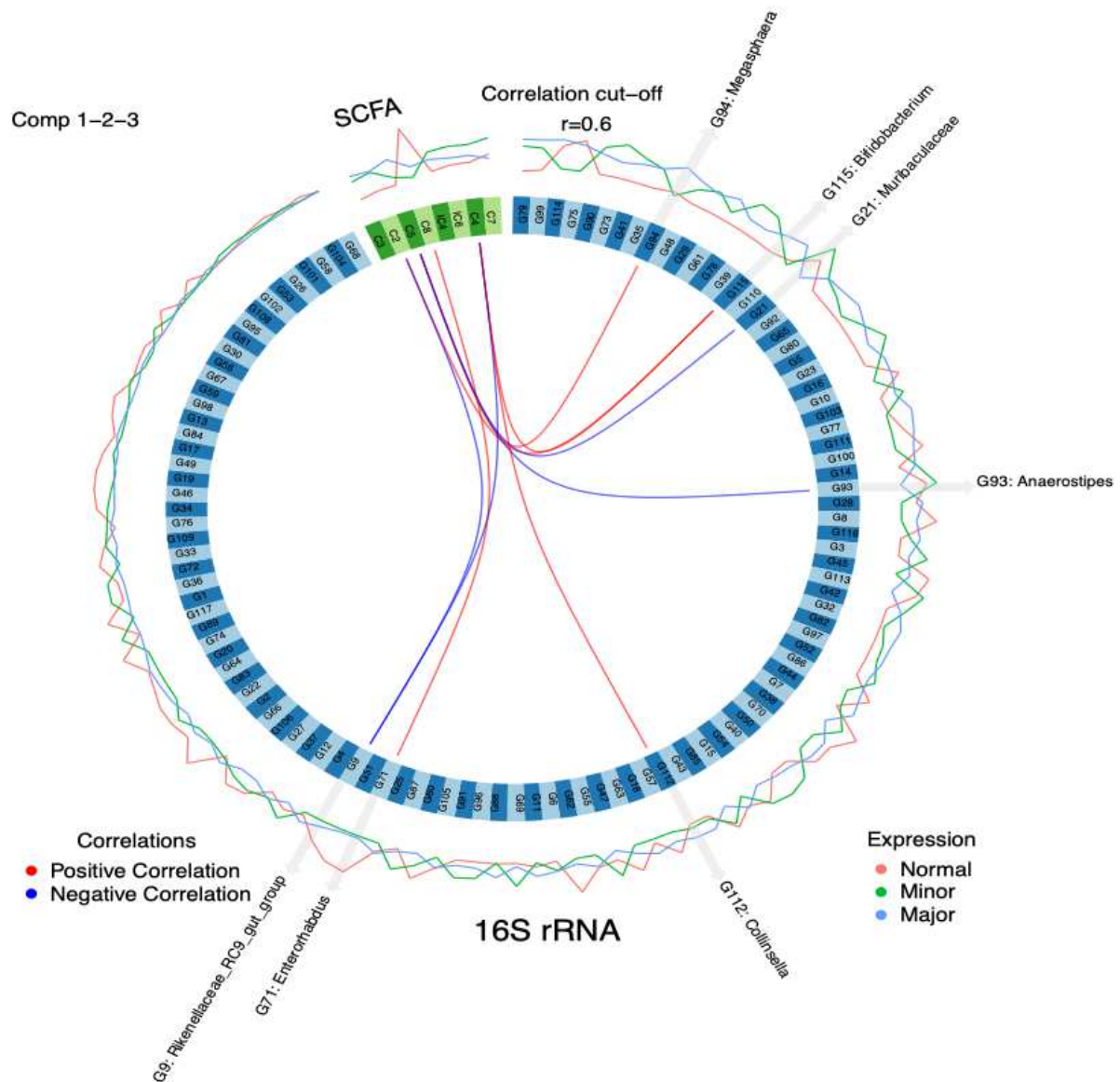


Figure 8: Significant correlations ($|r|>0.6$) whether positive (red) or negative (blue) between microbiome and SCFAs returned from the algorithm as a CIRCOS plot. The expression levels of both 16SrRNAs and SCFAs for different categories are drawn around the CIRCOS plot as line plots. The annotation of main correlating genera is shown alongside.

Table 1: The different haematological parametric comparative analysis showed that significant differences among the Hb, HCT, MCV, MCHC, PLT and RWD (adjusted $P < 0.05$) . Hb, HCT, MCV showed significant difference among the group (adjusted $p = 0$) which indicate the changing in the erythrocyte's properties in the group. The other adjusted p values for MCHC (adjusted $p < 0.001$), PLT (adjusted $p < 0.001$), RWD (adjusted $p < 0.001$) that is strongly related with group status. In contrast the MCH, WBC and RBC are not significantly differed among the group. These results showed that beta thalassemia firstly affect the haemoglobin concentration and the RBCs indices and then secondly it will effect MCH, WBC, RBC count across the study group.

	Feature	Term	P_Value	Adjusted_P_Value
1	HB	Group	$p < 0.05$	$p < 0.05$
2	HCT	Group	$p < 0.05$	$p < 0.05$
3	MCV	Group	$p < 0.05$	$p < 0.05$
4	MCHC	Group	$p < 0.001$	$p < 0.001$
5	PLT	Group	$p < 0.001$	$p < 0.001$
6	RDW	Group	$p < 0.001$	$p < 0.001$
7	MCH	Group	0.187	0.26
8	WBC	Group	0.240	0.30
9	RBC	Group	0.230	0.30

Table 2: Redundancy analysis with both forward and reverse selection was performed to select the most important parameters that explain variation in the community matrices. The initial set of variables considered are as follows (with those selected in the final PERMANOVA models in bold case): **Status** (Normal, Minor Thalassemia, Major Thalassemia), C2, **C3**, IC4, **C4**, IC5, C5, **IC6**, **C6**, C7, C8. Here df, SS, and F are Degree of Freedom, Sum of Squared Errors, and F Statistic, respectively.

	Covariates	df	SS	R ²	F	p
Bray-Curtis Distance						
	Status	2	0.6268	0.1230 0	1.623 3	0.001***
	C6	1	0.3051	0.0598 7	1.580 1	0.033 *
	IC6	1	0.3027	0.0594 0	1.567 8	0.039*
	Residual	20	3.8614	0.7577 3		
	Total	24	5.0960	1		
Unweighted Uni-Frac						
	C3	1	0.3528	0.0974 3	2.648 1	0.001***
	C4	1	0.2742	0.0757 3	2.058 2	0.016*
	IC6	1	0.1963	0.0542 1	1.473 4	0.084
	Residual	21	2.7981	0.7726 3		
	Total	24	3.6215	1		
Weighted Uni-Frac						

	Status	2	0.05517	0.1666	2.299	0.006**
				1	2	
	C8	1	0.02401	0.7252	2.001	0.051
					4	
	Residual	21	0.25196	0.7608		0
				7		
	Total	24	0.33115	1		
Hierarchical Meta-Storms						
	Status	2	0.01124	0.1504	2.278	0.028*
			6	5	3	
	C6	1	0.00801	0.1072	3.249	0.027 *
			9	8	2	
	C8	1	0.00612	0.0819	2.481	0.041
			4	2	2	
	Residual	20	0.04935	0.6603		
			9	5		
	Total	24	0.07474	1		
			7			
Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1						

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryDataTableS1.csv](#)
- [SupplementaryDataTableS2.csv](#)
- [BetaThalassemiaQuestionnaire.docx](#)