scientific reports

Check for updates

OPEN

Distinct functional and compositional properties in the gut microbiome of children with acute lymphoblastic leukaemia identified by shotgun metagenomics

Edoardo Muratore[1,13], Gabriele Conti[2,13], Marco Fabbrini[2]✉, Daniele Zama[3,4], Nunzia Decembrino[5], Paola Muggeo[6], Rosamaria Mura[7], Katia Perruccio[8], Davide Leardini[1], Monica Barone[2], Marco Zecca[9], Simone Cesaro[10], Patrizia Brigidi[2,11], Silvia Turroni[11,12] & Riccardo Masetti[1,4]

Acute lymphoblastic leukaemia (ALL) represents the most common childhood malignancy, and emerging evidence underscores the impact of the gut microbiome (GM) on its pathogenesis. In this study, we used shotgun metagenomics to investigate the GM of 30 ALL patients at diagnosis—19 with B-ALL and 11 with T-ALL—and compared them to 176 healthy controls (HCs). When considered as a single ALL group versus HCs, clear compositional differences emerged: ALL patients exhibited higher relative abundances of *Enterococcus faecium*, oral commensals such as *Rothia dentocariosa*, and multiple opportunistic species, whereas HCs were enriched in short-chain fatty acid producers like *Anaerostipes hadrus* and *Intestinibacter bartlettii*. Functionally, the ALL GM relied more on protein and amino acid catabolism, while HCs possessed enhanced pathways for carbohydrate and folate metabolism. These findings broadly align with 16S rRNA-based analyses from previous publications, though some discrepancies highlight differences in technique-driven resolution. In contrast, comparing the two major molecular phenotypes—B-ALL and T-ALL—revealed only minimal taxonomic and functional differences, primarily confined to BAs metabolism pathways. Overall, our results indicate that children with ALL at the time of diagnosis already display a dysbiotic signature, bolstering the notion that a disturbance in GM development during childhood may be linked to the multistep pathogenesis model of ALL.

Acute lymphoblastic leukaemia (ALL) is the most common childhood malignancy, accounting for approximately one third of all childhood cancers[1,2]. Advances in chemotherapeutic regimens and supportive care have markedly increased survival, with cure rates often approaching 90% in high income countries[3]. Despite these advancements, the risk of mortality due to relapse and therapy-related toxicities, along with the burden of long-term treatment complications, highlights the crucial role of research aimed at deepening our understanding of its aetiology, with the ultimate goal of achieving primary prevention.

[1]Pediatric Hematology and Oncology, IRCCS Azienda Ospedaliero-Universitaria di Bologna, Bologna, Italy. [2]Human Microbiomics Unit, Department of Medical and Surgical Sciences, University of Bologna, Bologna, Italy. [3]Pediatric Unit, IRCCS Azienda Ospedaliero-Universitaria di Bologna, Bologna, Italy. [4]Department of Medical and Surgical Sciences (DIMEC), University of Bologna, Bologna, Italy. [5]Neonatal Intensive Care Unit Unit-AOU Policlinico "Rodolico-San Marco", University of Catania, Catania, Italy. [6]Pediatric Hematology and Oncology Department, University Hospital of Policlinico of Bari, Bari, Italy. [7]Pediatric Hematology and Oncology, Azienda Ospedaliera Brotzu, Cagliari, Italy. [8]Pediatric Hematology and Oncology Department, "Santa Maria Della Misericordia" Hospital, Perugia, Italy. [9]Pediatric Haematology/Oncology IRCCS San Matteo Hospital Foundation, Pavia, Italy. [10]Pediatric Hematology and Oncology, Department of the Mother and the Child, Azienda Ospedaliera Universitaria Integrata Verona, Verona, Italy. [11]IRCCS Azienda Ospedaliero-Universitaria di Bologna, Bologna, Italy. [12]Unit of Microbiome Science and Biotechnology, Department of Pharmacy and Biotechnology, University of Bologna, Bologna, Italy. [13]Edoardo Muratore and Gabriele Conti contributed equally to this work. ✉email: m.fabbrini@unibo.it

The current multistep pathogenesis of pediatric ALL arises from genetic susceptibility through secondary genetic mutations driven by environmental triggers[4]. Some early-life epidemiological factors have been already hypothesized as risk indicators for childhood ALL, including caesarean section birth, infections, diminished breast feeding and paucity of social contacts, all of which are known to have a strong impact on gut microbiome (GM) development[4,5]. These factors could act as triggers for both altered immune system training and changes in the microbiota's composition and diversity, ultimately contributing to the complex pathogenesis of childhood ALL.

The GM is a complex ecosystem that contributes to the host's metabolism, immunity, and mucosal integrity[6]. Disruptions of GM balance have been implicated in the pathogenesis and progression of numerous diseases, including inflammatory disorders, and various malignancies[7]. In haematological cancers specifically, several studies point to the GM as a crucial moderator of both disease course and treatment outcomes[8–12].

Early investigations into the relationship between the GM and ALL primarily utilized 16S rRNA gene sequencing to catalog microbial taxonomic composition which are inherently limited by their resolution, typically identifying only genus-level taxa, and cannot robustly characterize functional pathways or gene-specific mechanisms[13–15]. Despite these limitations, initial studies consistently indicate shifts in alpha and beta diversity, along with a decline in potentially beneficial commensal taxa, in ALL patients compared to healthy populations[16], with lower abundance of short-chain fatty-acid-producing bacterial taxa, potentially being linked to the speculated dysregulated immune responses to common infectious triggers of children with ALL[4,5].

Shotgun metagenomics, by contrast, enables comprehensive profiling of the entire microbial gene repertoire, thereby providing species- and sometimes strain-level taxonomic resolution while revealing a broad spectrum of metabolic and functional pathways[17]. This approach is increasingly employed in oncology research to uncover mechanistic links between the GM and host processes, such as nutrient metabolism, immune modulation, and drug metabolism[18]. Importantly, shotgun metagenomics can illuminate microbial drivers of therapeutic responses, antimicrobial resistance patterns, and metabolic pathways that may interact with the host immune system and cooperate in ALL pathophysiology. Yet, to date, there has been minimal application of this high-resolution technique in the specific context of paediatric ALL.

Materials and methods

Patient enrolment and sample collection

The present study was conducted through the AIEOP network, approved by the ethics committees of the Fondazione IRCCS Policlinico San Matteo, Pavia, Italy (Protocol number 20160001176), and ratified by each participating institution. Parents or guardians provided written informed consent, according to the declaration of Helsinki. Between 1 January 2017, and 30 June 2020, eligible paediatric patients older than 1 month undergoing first-line induction chemotherapy for ALL with a faecal sample collected at the time of diagnosis from paediatric oncohaematology in Bologna, Verona, Pavia, Perugia, Bari, and Cagliari were included in this analysis. During the study period, every centre provided the diagnostic and therapeutic process according to AIEOP national protocols, and supportive therapy following local guidelines, including antibiotic therapy and prophylaxis. Clinical and laboratory data, as well as antibiotic administration, were recorded in an electronic CRF. Faecal samples were considered for this study if taken before the start of chemotherapy. Samples were stored at − 80 °C and shipped in dry ice to the University of Bologna for GM analysis. ALL immunophenotypes were used to further stratify the patients, specifically the B-ALL and T-ALL groups refer to B-cell and T-cell acute lymphoblastic leukaemia immunophenotypes, as determined by hospital routine flow cytometry at diagnosis.

The healthy control (HC) cohort comprised 176 publicly available shotgun metagenomes from children across Europe, including 52 Italian individuals. Each HC sample was annotated with metadata, including age and geographic origin, to allow proper stratification and interpretation (Supplementary Table 1). The age distribution encompassed the full pediatric range (infants to adolescents).

Library preparation and shotgun sequencing

DNA libraries were prepared using the QIAseq FX DNA Library Kit (Qiagen) per the manufacturer's protocol. Microbial DNA was quantified with a Qubit fluorometer (Invitrogen), ensuring ≥ 100 ng input per sample. DNA was fragmented to ~ 450 bp, end-repaired, and A-tailed using the FX enzyme mix (4 °C, 1 min; 32 °C, 8 min; 65 °C, 30 min). Illumina adapter barcodes were ligated at 20 °C for 15 min, followed by double purification with AMPure XP beads (Beckman Coulter). Libraries were PCR-amplified (10 cycles) and purified again. Individual libraries were quantified and pooled at 4 nM equimolar concentration. Sequencing was performed on an Illumina NovaSeq platform (2 × 150 bp) according to the manufacturer's instructions. The sequencing yielded an average depth of 3.70 ± 0.67 Gb (mean ± SD) bases per sample.

Bioinformatic reads processing

The raw sequencing reads generated in this study, along with the 176 publicly HC metagenomes specified in Supplementary Table 1 (retrieved from NCBI or ENA), were processed through a uniform bioinformatics pipeline. Raw sequencing reads were first assessed for overall quality using FastQC[19]. Adapter trimming, quality filtering, and error correction were then performed with fastp[20] using the following parameters: --detect_adapter_for_pe --dont_eval_duplication --length_required 50 --correction --overlap_len_require 20. Quality-filtered reads were subsequently processed with KneadData v0.7.2[21]using the Bowtie2 aligner[22] to remove potential human host contamination by mapping against the hg38 human reference genome. For taxonomic profiling, MetaPhlAn 4[23] was employed, with alpha diversity metrics derived using the accompanying MetaPhlAn utility scripts. Functional pathway characterization was carried out using HUMAnN 3[24], following normalization of pathway counts in copies per million (CPM). Both the ALL-derived shotgun data and the downloaded HC datasets underwent these identical steps to ensure methodological consistency. Separately, 16S rRNA datasets were obtained from

NCBI (Rajagopala et al., 2020: NCBI PRJNA421352, De Pietri et al., 2020: ENA PRJEB35526, Chua et al., 2020: NCBI PRJNA533024) and analysed using a containerized 16S-KrakenBracken-Pipeline v2.2 (publicly available on GitHub at: https://github.com/FabbriniMarco/16S-KrakenBracken-Pipeline), which utilizes the SILVA 138.1 SSuRef database[25]. To align 16S-based genus-level assignments with shotgun metagenomic taxonomy calls, the corresponding NCBI TaxIDs were extracted from the SILVA taxmap and full metadata files and matched against MetaPhlAn TaxIDs, enabling direct comparisons and facilitating an evaluation of the shared taxonomic fraction.

Statistical analyses

Statistical analyses were performed using R (version 4.4.2; R Core Team, Vienna, Austria). Continuous variables and alpha diversity indices were compared using t-tests or Wilcoxon rank-sum tests, depending on Shapiro–Wilk normality results. Categorical variables were analyzed via chi-squared tests. Beta diversity (Unweighted/Weighted UniFrac) was visualized via PCoA and tested using PERMANOVA (9999 permutations, adonis function). Differential taxonomic and functional features were assessed with pairwise Wilcoxon tests and Benjamini–Hochberg FDR correction (significance: FDR < 0.05; trend: FDR < 0.1). Correlations between pathways and species abundances were evaluated using Kendall's tau in the microbAIDeR package[26] with FDR-adjusted *p*-values. The resulting corrplot was made using the ggcorrplot R-package[27].

Results

Study cohort description

A total of 30 patients diagnosed with ALL were enrolled in this study; 13 from Bari, 10 from Bologna, 3 from Perugia, 2 from Verona, 1 from Pavia and 1 from Cagliari. 57% of children were male, with a median age at diagnosis of 3.8 years (range 1.5–17.3). Median weight at ALL onset was 17 kg (range 12–67). Antibiotics were administered in 27% (8/30) of patients in the 30 days before ALL diagnosis.

The study population includes 19 children with the B-ALL phenotype and 11 with the T-ALL phenotype, and no recurrent genetic and cytogenetic aberrations were observed in the cohort. No significant differences were observed between the B-ALL and T-ALL groups in terms of prior antibiotic use, presence of fever, onset of mucositis, future development of bloodstream infections during induction therapy, or biological sex distribution ($p > 0.05$). Additionally, no significant differences existed between the two groups with respect to age or body weight at enrolment.

ALL microbiome show compositional and functional differences compared to HCs

We first sought to compare the GM profiles of children with ALL to those of a healthy population. To this end, we included faecal shotgun metagenomes from 176 healthy controls (HCs) across Europe, sourced from publicly available datasets (Supplementary Table 1). The HC cohort encompassed a wide range of ages, facilitating robust comparisons with the 30 ALL patients in terms of age-related microbial signatures and lifestyle factors.

Although alpha diversity did not differ significantly between the two groups (Supplementary Fig. 1), beta diversity analyses based on both Weighted and Unweighted UniFrac distances revealed clear compositional differences (Fig. 1A, PERMANOVA $p = 0.001$). Notably, Unweighted UniFrac indicated age-related clustering (Supplementary Fig. 2), with infant cohorts (< 1 year old) forming distinct clusters compared to adolescents. This observation suggests that the control cohort is relatively homogeneous overall, and that most compositional differences are age-related rather than cohort-specific.

To explore these differences further, we examined species-level GM composition (Fig. 1B). ALL patients exhibited significantly higher abundances of *Enterococcus faecium*, *Rothia dentocariosa*, and *Lautropia mirabilis*. *E. faecium* is commonly associated with repeated antibiotic exposure in hospital settings, while *R. dentocariosa* and *L. mirabilis* are typical oral commensals that may have translocated to the gut—a phenomenon frequently observed in immunocompromised individuals. Conversely, signatures enriched in HCs included short-chain fatty acid (SCFA) producers such as *Anaerostipes hadrus*, *Blautia wexlerae*, and *Intestinibacter bartlettii*, as well as the commensals *Haemophilus parainfluenzae* and *Romboutsia timonensis*. Interestingly, patient-specific signatures were identified in the ALL cohort, with markedly higher abundances of opportunistic pathogens, including *Isoptericola variabilis*, *Trueperella pyogenes*, *Rothia mucilaginosa*, *Rothia aeria*, and *Abiotrophia defectiva*.

No significant differences in species-level relative abundances were observed based on antibiotic use in the 30 days before sampling, whether considering all antibiotics (8 exposed vs. 22 not exposed) or only impactful antibiotics (2 ceftazidime, 1 piperacillin/tazobactam, 1 cefotaxime + teicoplanin vs. 26 not exposed to these strong antibiotics), as determined by FDR-corrected Wilcoxon tests.

Functionally, significant differences were detected in pathway-level CPMs between ALL patients and HCs (Fig. 2). HCs exhibited enriched pathways associated with canonical energy and nucleotide metabolism, including glycolysis and nucleotide biosynthesis. These results suggest that HC-associated microbes are well-adapted for utilizing dietary carbohydrates (e.g., fiber and glucose) to support core anabolic processes such as DNA and RNA synthesis. Enhanced folate metabolism in HCs further underscores a balanced, nutrient-rich microbiome that fosters beneficial metabolic interactions.

In contrast, ALL patients showed a distinct reliance on protein and amino acid catabolism. Enriched pathways in the ALL cohort included ethanolamine utilization (a metabolite derived from phospholipid breakdown), as well as degradation pathways for allantoin, arginine, lysine, and putrescine. This metabolic shift may reflect a gut environment shaped by the immunocompromised state and altered dietary intake. As observed in earlier analyses comparing B-ALL and T-ALL, ALL patients demonstrated significantly higher CPMs in BAs metabolism pathways, including 7β-dehydroxylation, 7α-dehydroxylation, and epimerization processes. These findings suggest a heightened potential for secondary BAs production in ALL patients. However, without direct metabolomic measurements, the downstream effects of this shift—such as pro- or anti-inflammatory impacts—

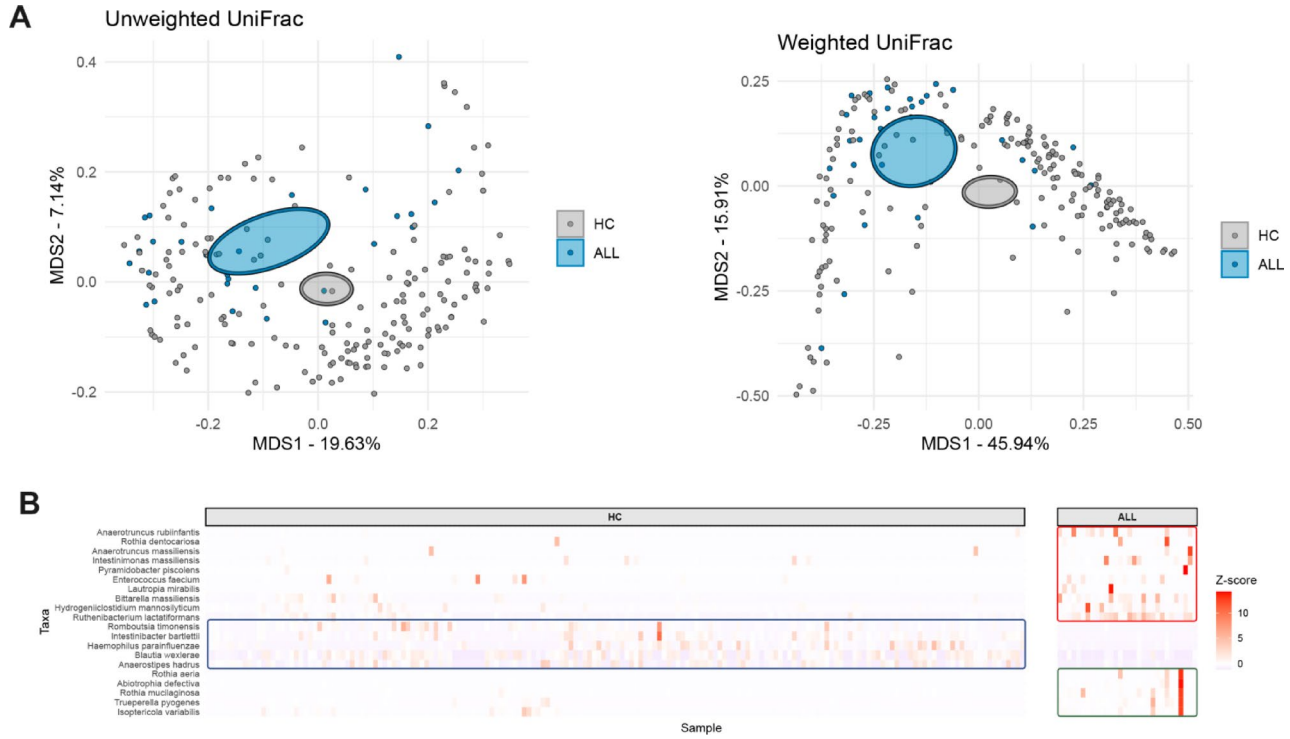

Fig. 1. Beta diversity analyses and species-level differences between HCs and ALL patients. (**A**) Principal coordinates analysis (PCoA) plots based on unweighted UniFrac (left) and weighted UniFrac (right) distances, illustrating distinct clustering of HCs (grey) and ALL (blue). Ellipses represent 95% confidence intervals for each group's distribution. (**B**) Heatmap showing species with significantly different relative abundances between the two groups (FDR-adjusted $p < 0.05$, Wilcoxon test). Z-score scaling was applied to each species across all samples. Species labels are color-coded to indicate higher abundance in ALL (red), higher abundance in HCs (blue), or unique presence in specific patients (green).

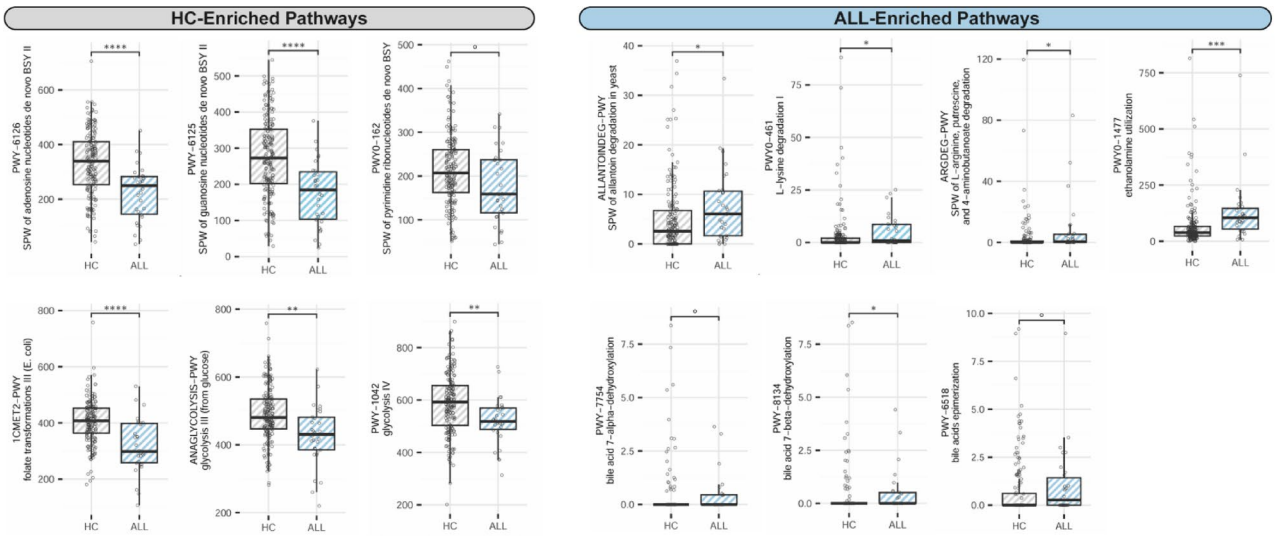

Fig. 2. Functional potential differences in microbiomes of HCs and ALL patients. Pathways showing significantly different abundance (in copies per million reads, CPM) between healthy controls and acute lymphoblastic leukaemia patients. Boxplots on the left represent pathways enriched in HCs, while those on the right illustrate pathways enriched in ALL. Statistical significance was determined by Wilcoxon tests with false discovery rate (FDR) correction ($p < 0.05$): ****, $p < 0.0001$; *** $p, < 0.001$, **, $p < 0.01$, *, $p < 0.05$; °, $p < 0.1$. Pathways are annotated according to the MetaCyc nomenclature. BSY, biosynthesis; PWY, pathway; SPW, superpathway.

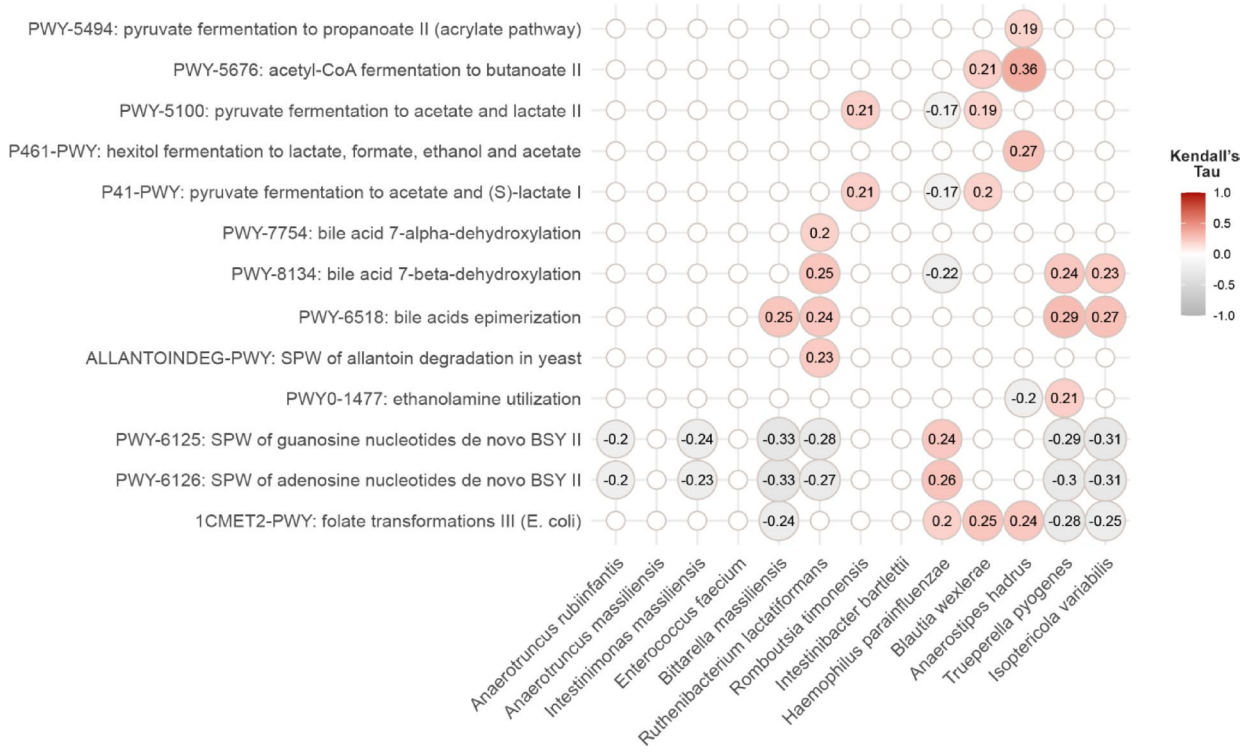

Fig. 3. Associations between microbiome composition and function in ALL patients. Correlation matrix linking functional pathways to microbial species in acute lymphoblastic leukaemia (ALL) patients. The plot displays Kendall's tau correlation coefficients (≥0.2) for pathway-species pairs with a false discovery rate (FDR)–adjusted $p < 0.05$. Warmer (red) or cooler (grey) colours indicate positive or negative correlations, respectively, underscoring key associations between the microbial taxa and functional pathways.

remain unclear. Compared to HCs, this altered BAs metabolism may indicate a rearrangement of inflammatory control mechanisms mediated by the microbiome.

Correlation analysis between functional pathway counts and species abundances (Fig. 3) highlighted key drivers of increased BA-transforming potential in the ALL cohort. Notable contributors included *Ruthenibacterium lactatiformans* (7α- and 7β-dehydroxylation, epimerization), *Bittarella massiliensis* (epimerization), and the opportunistic pathogens *Trueperella pyogenes* and *Isoptericola variabilis* (7β-dehydroxylation, epimerization).

To further dissect the drivers of functional differences, we analyzed stratified HUMAnN outputs to identify the main taxa contributing to each pathway (Supplementary Fig. 3). Bile acid–transforming pathways (PWY-7754, PWY-8134, PWY-6518) were primarily driven by *Clostridium scindens*, *Ruthenibacterium lactatiformans* and *Bacteroides spp.* Pathways enriched in HCs, such as folate transformations, nucleotide de novo biosynthesis, and SCFA production (butanoate, lactate fermentation), were mostly associated with *Bacteroides vulgatus*, *Veillonella seminalis*, *Anaerostipes hadrus* and *Bifidobacterium longum*. Conversely, ALL-enriched pathways (ethanolamine utilization, putrescine biosynthesis, allantoin degradation) were dominated by *Enterobacteriaceae* including Escherichia coli, *Klebsiella pneumoniae* and *Raoultella spp.* These findings indicate that specific pathobionts fuel ALL-associated dysbiosis, while HC-associated pathways are maintained by SCFA-producing commensals.

Comparing shotgun metagenomics results with 16S data

To contextualize and compare the differential findings between shotgun metagenomics and previously published 16S rRNA datasets[13–15], we examined whether both techniques revealed consistent microbial patterns between ALL patients and healthy controls, while acknowledging their intrinsic methodological differences. Alpha and beta diversity analyses revealed differences primarily driven by the sequencing technique, with higher diversity observed in metagenomics datasets compared to 16S amplicon datasets, as expected (Supplementary Fig. 4). However, no significant differences in alpha diversity between ALL and HC groups were observed within either technique group. Both 16S sequencing and shotgun metagenomics highlighted significant differences in microbiota composition at the genus level.

Examining the relative abundances of the 50 most abundant genera across cohorts (Supplementary Fig. 5), we observed notable inconsistencies among the 16S datasets, both within and between HC and ALL groups. In our shotgun metagenomics analysis, differences in HC composition were largely attributable to age-related patterns. Specifically, the infant (1-year-old) HC cohort exhibited higher abundances of *Bifidobacterium*, while adolescent HCs were enriched in *Bacteroides*, consistent with age-related microbiome development.

When comparing genus-level differences between ALL and HC groups, it became evident that the most striking variations were technique-driven. These discrepancies likely reflect the limited taxonomic resolution of

the variable regions targeted in 16S rRNA sequencing and the smaller size of 16S reference databases compared to the more comprehensive shotgun metagenomics approach.

B-ALL and T-ALL showed minimal differences in gut microbiome

Next, we investigated whether any of the observed microbiome differences might vary between the two predominant molecular phenotypes of ALL: B-ALL and T-ALL. Although these subtypes are clinically distinct, our aim was to see if they exhibited similar or divergent gut microbiome profiles at diagnosis, potentially providing additional insight into the disease's underlying biological heterogeneity.

Alpha diversity metrics, including Shannon and Inverse Simpson indices, did not reveal significant differences among the B-ALL, T-ALL, and HC groups (Wilcoxon test, FDR-adjusted $p > 0.05$). Regarding gut microbiome composition, no significant differences were observed between the B-ALL and T-ALL groups at either the genus or species level ($p > 0.05$). Functionally, however, B-ALL patients exhibited higher CPM counts for bile acids (BAs) metabolism pathways, specifically 7β-dehydroxylation, 7α-dehydroxylation, and epimerization, when compared to HCs ($p < 0.05$, Supplementary Fig. 6). In contrast, no significant differences were detected in the T-ALL group relative to either HCs or B-ALL patients. These findings suggest that the gut microbiome of B-ALL patients may exhibit a greater propensity for secondary BAs production compared to T-ALL patients.

Overall, these data suggest that while B-ALL patients may exhibit a slightly enhanced capacity for secondary BAs production, the T-ALL subtype shows no distinct microbial or functional signatures when compared with B-ALL or HCs. Consequently, aside from these modest variances in BAs metabolism, the gut microbiome profiles of B-ALL and T-ALL patients at diagnosis appear largely comparable, suggesting that immunophenotype exerts only a limited influence on GM composition and function.

Discussion

In this study, we investigated the GM community structure of newly diagnosed ALL patients, evaluating both compositional and functional differences versus a large healthy control (HC) cohort. Using shotgun metagenomics, we generated high-resolution taxonomic and functional profiles, complemented by published 16S rRNA datasets. This enabled characterization of baseline GM in treatment-naïve patients with minimal antibiotic exposure. Our findings reveal novel insights into ALL-associated GM, highlighting marked contrasts with healthy states and subtle B-ALL vs T-ALL differences.

Comparing ALL patients with HCs revealed a distinctly altered gut microbiome. While alpha diversity did not differ significantly, beta diversity analyses clearly separated the groups. Species-level differences were notable: despite early disease stage, ALL patients showed enrichment of organisms linked to hospital environments and immunocompromised states, including *Enterococcus faecium* and oral commensals like *Rothia dentocariosa* and *Lautropia mirabilis*, possibly translocated before chemotherapy. In contrast, HCs were enriched in SCFA-producing species such as *Anaerostipes hadrus* and *Intestinibacter bartlettii*, along with taxa associated with gut homeostasis. This suggests that a disrupted GM, characterized by increased pathobionts and reduced SCFA producers, precedes therapy and is not solely a consequence of immunosuppression or antibiotics. The early presence of opportunists may predispose to translocation and infection, highlighting the importance of GM monitoring from diagnosis.

Functional analyses mirrored compositional patterns, indicating a GM shift toward protein and amino acid metabolism in ALL. Pathways for ethanolamine utilization and degradation of allantoin, arginine, lysine, and putrescine were enriched, suggesting catabolism of nitrogen-rich compounds. This shift is particularly remarkable given that patients had not yet undergone therapy, suggesting that the disease context itself—rather than therapeutic interventions—may be driving these metabolic adaptations. In contrast, HCs exhibited metabolic pathways associated with carbohydrate fermentation, folate biosynthesis, and other anabolic processes, reflecting a more balanced interaction with dietary inputs and the host.

The contributional pathways analysis revealed that the increased bile acid–transforming potential in ALL was not generic but specifically linked to *Clostridium scindens* and *Ruthenibacterium lactatiformans*, two species with established roles in secondary bile acid metabolism. Such taxa may contribute to an accumulation of deoxycholic acid and other secondary bile acids, with potential implications in mucosal inflammation and immune modulation. Conversely, HCs showed strong contributions from SCFA-producing commensals, suggesting that reduced butyrate- and lactate-producing capacity may further exacerbate gut barrier dysfunction in ALL. These results, together with recent evidence linking specific taxa to functional pathways[28] provide a more mechanistic understanding of how microbiome alterations may impact ALL biology.

The potential immunomodulatory role of BAs metabolic changes is particularly intriguing. Secondary BAs, for instance, can act as signalling molecules influencing T-cell differentiation and regulatory pathways. A gut microbiome equipped with heightened BA-transforming capabilities might therefore have implications for inflammation and immune surveillance in ALL.

Comparison with 16S rRNA studies contextualized methodological effects. Although results were broadly consistent, differences in diversity estimates and resolution reflect technique-specific limitations. Shotgun metagenomics enabled strain-level and functional insights, underscoring its value despite 16S utility. Together, these findings stress the need for thoughtful methodological integration in microbiome research.

Notably, a recent study by Wang et al.[29], investigated the gut microbiota of 49 children with ALL during induction chemotherapy using shotgun metagenomics. While our study focuses on treatment-naïve patients at diagnosis, Wang and colleagues reported pronounced microbial shifts during chemotherapy, with progressive decreases in α-diversity, enrichment of pathobionts such as *Escherichia coli* and *Klebsiella pneumoniae*, and depletion of beneficial taxa including *Bifidobacterium longum*. Functionally, they observed alterations in carbohydrate and sulphur metabolism pathways, with baseline *B. longum* abundance emerging as a predictor of infectious complications. Taken together, these results reinforce our observations of an early dysbiotic signature

in ALL, marked by loss of short-chain fatty acid producers and enrichment of opportunists, and suggest that this imbalance is exacerbated by chemotherapy, potentially contributing to infection risk. Thus, our dataset complements previous evidence by capturing the pre-treatment microbial landscape, providing a baseline context for interpreting subsequent therapy-driven dysbiosis.

Despite this overall divergence compared to healthy children, B-ALL and T-ALL phenotypes displayed remarkably similar gut microbial communities, as reflected by alpha and beta diversity metrics, as well as by taxonomic analyses at the genus and species levels. The only appreciable differences emerged in certain functional pathways related to BAs metabolism. B-ALL patients exhibited a higher abundance of genes involved in 7α- and 7β-dehydroxylation and epimerization steps, suggesting a greater capacity for transforming primary into secondary BAs. Although this finding highlights a potential metabolic distinction between the two subtypes, it must be approached cautiously in the absence of direct metabolomic data. Whether these functional pathways lead to pro- or anti-inflammatory effects is not entirely clear. Nevertheless, the observed differences invite further inquiry into the role of BAs metabolism in shaping the immune status and inflammatory tone in ALL patients.

Several limitations of this study warrant consideration. First, despite the inclusion of a relatively large HC cohort, the inherent heterogeneity of ALL populations—driven by factors such as age and early antibiotic usage—challenge the identification of robust microbiome signatures. Second, our analysis did not include direct metabolomic or immunological readouts, limiting interpretation of the biological consequences of shifts in BAs metabolism and protein catabolism. Lastly, cross-sectional sampling provides only a snapshot of a highly dynamic microbial community.

Another limitation concerns the heterogeneity of age in our ALL cohort (range 1.5–17.3 years). Given the strong influence of age on gut microbiome development, this represents a potential confounder. However, we selected a healthy control cohort spanning the same pediatric range and annotated each sample with age metadata. Our beta diversity analyses revealed age-related clustering among controls, but the ALL-associated dysbiotic signatures persisted across age groups. While stratification or covariate adjustment was not feasible given the limited sample size, these observations support the robustness of our conclusions despite age heterogeneity.

While our age-stratified visualizations indicate that ALL-associated signatures persist across pediatric age groups, we acknowledge that more formal modeling frameworks could further delineate age-related from disease-specific effects. In particular, supervised "microbiota-age" prediction and age-discriminatory-taxa approaches have been effectively used to characterize early-life microbial succession and identify taxa most strongly associated with age[30–32]. Applying such frameworks—once adequately powered and ideally with longitudinal sampling—would allow future studies to first adjust for age-discriminatory signals and then test for residual, ALL-specific differences.

In conclusion, our findings reveal early gut microbiome alterations in children with acute lymphoblastic leukaemia, characterized by an overabundance of opportunistic and oral-associated species, a depletion of short-chain fatty acid–producing commensals, and a functional shift toward amino acid, ethanolamine, and secondary bile acid metabolism. These results, derived from shotgun metagenomics, provide a high-resolution view of the microbial and metabolic landscape preceding treatment, supporting the hypothesis that gut dysbiosis may contribute to ALL pathogenesis and offering a framework for future integrative studies combining metagenomics, metabolomics, and immune profiling.

Data availability

High-quality paired-end sequences generated from this study can be found in the NCBI Sequence Read Archive, under the Project Accession code PRJNA1226574. Processed tables can be found in Figshare under https://doi.org/10.6084/m9.figshare.30074665.

Received: 7 July 2025; Accepted: 3 November 2025
Published online: 03 December 2025

References

1. Pui, C. H. et al. Clinical impact of minimal residual disease in children with different subtypes of acute lymphoblastic leukemia treated with response-adapted therapy. *Leukemia* **31**, 333–339 (2016).
2. Public Health England. *Children, Teenagers and Young Adults UK Cancer Statistics Report 2021*. (2021).
3. Hunger, S. P. & Mullighan, C. G. Acute lymphoblastic leukemia in children. *N. Engl. J. Med.* **373**, 1541–1552 (2015).
4. Greaves, M., Cazzaniga, V. & Ford, A. Can we prevent childhood leukaemia?. *Leukemia* **35**, 1258–1264 (2021).
5. Hauer, J., Fischer, U. & Borkhardt, A. Toward prevention of childhood ALL by early-life immune training. *Blood* **138**, 1412–1428 (2021).
6. Sommer, F. & Bäckhed, F. The gut microbiota—Masters of host development and physiology. *Nat. Rev. Microbiol.* **11**, 227–238 (2013).
7. Garrett, W. S. Cancer and the microbiota. *Science* **1979**(348), 80–86 (2015).
8. Taur, Y. et al. The effects of intestinal tract bacterial diversity on mortality following allogeneic hematopoietic stem cell transplantation. *Blood* **124**, 1174–1182 (2014).
9. Peled, J. U. et al. Microbiota as predictor of mortality in allogeneic hematopoietic-cell transplantation. *N. Engl. J. Med.* **382**, 822–834 (2020).
10. Leardini, D. *et al.* Levofloxacin prophylaxis in pediatric and young adult allogeneic hematopoietic stem cell transplantation recipients does not prevent infective complications and infections-related deaths. In *Open Forum Infect Dis* https://doi.org/10.1093/OFID/OFAE707
11. Masetti, R. et al. Gut microbiota diversity before allogeneic hematopoietic stem cell transplantation as a predictor of mortality in children. *Blood* **142**, 1387–1398 (2023).
12. Fabbrini, M. et al. Levofloxacin prophylaxis and parenteral nutrition have a detrimental effect on intestinal microbial networks in pediatric patients undergoing HSCT. *Commun. Biol.* **6**, 36 (2023).
13. Chua, L. L. et al. Temporal changes in gut microbiota profile in children with acute lymphoblastic leukemia prior to commencement-, during-, and post-cessation of chemotherapy. *BMC Cancer* **20**, 1–11 (2020).

14. De Pietri, S. et al. Gastrointestinal toxicity during induction treatment for childhood acute lymphoblastic leukemia: The impact of the gut microbiota. *Int. J. Cancer* **147**, 1953–1962 (2020).
15. Rajagopala, S. V. et al. Persistent gut microbial dysbiosis in children with acute lymphoblastic leukemia (all) during chemotherapy. *Microb. Ecol.* **79**, 1034–1043 (2020).
16. Peppas, I., Ford, A. M., Furness, C. L. & Greaves, M. F. Gut microbiome immaturity and childhood acute lymphoblastic leukaemia. *Nat. Rev. Cancer* **23**, 565–576 (2023).
17. Franzosa, E. A. et al. Species-level functional profiling of metagenomes and metatranscriptomes. *Nat. Methods* **15**, 962–968 (2018).
18. Alexander, J. L. et al. Gut microbiota modulation of chemotherapy efficacy and toxicity. *Nat. Rev. Gastroenterol. Hepatol.* **14**, 356–365 (2017).
19. Andrews, S. FastQC: *A Quality Control Tool for High Throughput Sequence Data*. Preprint at http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (2010).
20. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
21. The Huttenhower Lab. KneadData—Quality control tool on metagenomic and metatranscriptomic sequencing data, especially data from microbiome experiments. Preprint at https://github.com/biobakery/kneaddata.
22. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
23. Blanco-Míguez, A. et al. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nat. Biotechnol.* **41**, 1633–1644 (2023).
24. Beghini, F. et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with biobakery 3. *Elife* **10**, e65088 (2021).
25. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**, D590–D596 (2013).
26. Fabbrini, M. & Conti, G. microbAIDeR—An ensemble of functions for easier and quicker preliminary microbiome analyses. Preprint at https://github.com/FabbriniMarco/microbAIDeR/ (2024).
27. Kassambara, A. ggcorrplot: *Visualization of a Correlation Matrix using 'ggplot2'*. Preprint at https://CRAN.R-project.org/package=ggcorrplot (2022).
28. Chen, S. et al. Consistent signatures in the human gut microbiome of longevous populations. *Gut. Microbes* **16**, 2393756 (2024).
29. Wang, H. et al. Microbial metagenomic shifts in children with acute lymphoblastic leukaemia during induction therapy and predictive biomarkers for infection. *Ann. Clin. Microbiol. Antimicrob.* **23**, 52 (2024).
30. Shao, Y. et al. Primary succession of Bifidobacteria drives pathogen resistance in neonatal microbiota assembly. *Nat. Microbiol.* **9**, 2570–2582 (2024).
31. David, L. A. et al. Gut microbial succession follows acute secretory diarrhea in humans. *MBio* **6**, 10 (2015).
32. Fahur Bottino, G. et al. Early life microbial succession in the gut follows common patterns in humans across the globe. *Nat. Commun.* **16**, 660 (2025).

Acknowledgements

We acknowledge CINECA for the availability of ADA Cloud high performance computing resources and support.

Author contributions

EM: Conceptualization, Writing—Original Draft; GC: Investigation; MF: Formal Analysis, Visualization, Writing—Original Draft; DL: Data Curation; MB: Writing—Review & Editing; ND: patients enrolment; KP: patients enrolment; RM: patients enrolment; PM: patients enrolment; SC: patients enrolment, supervision; MZ: patients enrolment, supervision; DZ: Resources; PB: Funding acquisition; ST: Writing—Review & Editing; RM: Project administration, Supervision.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-27280-7.

Correspondence and requests for materials should be addressed to M.F.