

Reviewed Preprint

v1 • July 21, 2026

Not revised

✉ For correspondence:

dmanwani@montefiore.org

libusha.kelly@einsteinmed.edu

Competing interests:

No competing interests declared

Reviewing editor:

Peter J Turnbaugh, University of California, San Francisco, United States

© 2026, Flamholz et al. This article is distributed under the terms of the

[Creative Commons Attribution](#)

[License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

Distinct prokaryotic gut microbiome and proviral-immune axes of pathophysiology in Sickle Cell Disease

Zachary N Flamholz¹, Jennifer De Los Santos², Karen Ireland², Janine Keenan², Jacob S Kazmi³, Aakash Mahant Mahant⁴, Anayeli Correa⁴, Paul S Frenette³, Betsy C Herold^{2,4}, Deepa Manwani²✉, Libusha Kelly^{1,4}✉

¹Department of Systems and Computational Biology, Albert Einstein College of Medicine, Bronx, United States •

²Department of Pediatrics, Montefiore Health System and Albert Einstein College of Medicine, Bronx, United States •

³Department of Cell Biology, Albert Einstein College of Medicine, Bronx, United States • ⁴Department of Microbiology and Immunology, Albert Einstein College of Medicine, Bronx, United States

eLife Assessment

This study presents a **valuable** metagenomic analysis of the gut microbiome in sickle cell disease (SCD) patients, revealing associations between bacteriophage, host immunity, and SCD pathophysiology. While these data are interesting and helpful for hypothesis generation, they are deemed **incomplete**; additional experiments would be needed to test causality and to provide mechanistic insight. Despite these limitations, this work will be of broad interest to researchers studying SCD, immunology, phage biology, and the microbiome, adding to the small but growing literature suggesting a microbial component to SCD.

<https://doi.org/10.7554/eLife.111630.1.sa4>

Abstract

Sickle cell disease (SCD) is a chronic, inherited condition rising across the globe. Prior studies revealed a direct link between the gut microbiome and disease micropathology via aged-like (ANs) neutrophils in mouse models. In SCD patients community-level shifts in the gut microbiome included decreases in diversity and the Firmicutes/Bacteroides (F:B) ratio, coupled to a loss of short chain fatty acid producing microbes and a shift to non-canonical butyrate production and aerobic fatty acid oxidation pathways. ANs and the proviral microbiome associate with multiple blood cytokines, while bacterial gut microbiome features largely do not. Prophages depleted of genes related to lysis, transcriptional regulation, and host takeover were enriched in SCD patient guts, pointing to domestication of these elements, and 25% of prophages were shared at high identity between study patients. In sum, we identify a viral-immune axis in SCD pathophysiology and targetable functional alterations to the gut microbiome in a heterogeneous chronic disease both affected by and effecting microbiome composition and function.

Introduction

Sickle cell disease (SCD) is a chronic, inherited hematologic disorder causing significant morbidity and mortality across the globe. SCD is characterized by pathologic red blood cell hemolysis leading to complications stemming from vaso-occlusion and vascular stress, ultimately resulting in end organ damage. In a 2018 study, it was estimated 1,950 children are born with SCD in the U.S. annually, and the life expectancy of a child with SCD is 54 years compared to 76 years for the age- and race-matched U.S. population¹. While there has been a steady increase in life expectancy for SCD patients resulting from diagnostic screening, early intervention, and better evidence-based

treatment guidelines, SCD hospitalization rates have increased². Pain crises, the clinical manifestation of acute vaso-occlusion, were the number one reason for SCD-related hospital admissions from 2004-2012². While gene therapy will play an important role in SCD treatment, there are significant barriers to widespread utilization³ and with SCD births on the rise globally⁴ alternative avenues of treatment must be pursued. In addition, novel biomarkers are valuable to better stratify SCD patients for targeted early intervention and more precise treatment decisions.

The gut microbiome is a physiological system with complex interactions with SCD. It was shown in a murine model of SCD that microbial antigens transit from the gut to the bloodstream where they activate neutrophils to a disease-promoting phenotype, termed “aged”, or “aged-like”, neutrophils (hereafter, “ANs”), and that depletion of gut microbes using broad-spectrum antibiotics decreased the AN population and improved inflammation-induced organ damage⁵. Human SCD patients were also found to have higher levels of the microbial antigen lipopolysaccharide in their blood and an increased AN percentage when compared to a control group of iron deficiency anemia patients with similar hemoglobin levels⁶. These findings led to the proposal of gut-targeting antibiotics as a treatment for SCD, and small studies in human subjects have shown antibiotics reduce the fraction of ANs and lipopolysaccharide in SCD patients⁷⁻⁹. The clinical momentum, though, has outpaced investigation of the SCD gut microbiome itself.

Gut microbiome studies that seek to identify microbial biomarkers of disease or disease severity look for abundance differences in individual taxa between case and control cohorts, as well as differences in summary measures related to community diversity. To date, there have been four studies that probed bacterial composition of the SCD gut microbiome compared to controls to determine if there are characteristic features of the SCD gut microbiome. Lim et al. did not observe differences in overall diversity for 35 SCD patients compared to a sickle trait control cohort, though one genus of Bacteroidetes had lower relative abundance in the SCD cohort¹⁰. Similarly, in a cohort of 32 pediatric SCD patients, Mohandas et al. did not observe a difference in alpha diversity or beta diversity compared to an immunocompromised cohort and a non-immunocompromised control cohort¹¹. In contrast, Brim et al. observed patterns of abundance differences at all taxonomic levels below phylum when 14 SCD patients were compared to healthy controls¹². Brim et al. also observed that SCD patients have a lower Firmicutes:Bacteroidetes (F:B) ratio, a gut microbiome community-level metric relevant to human health¹³. Finally, in a cohort of Angolan SCD pediatric patients, Delgado et al. did not observe a difference in alpha diversity compared to healthy siblings but did observe abundance differences at all taxonomic ranks¹⁴, though fewer in number and mostly different than observed by Brim et al.

These prior studies suffer from two limitations in characterizing the gut microbiome of SCD patients. First, all studies suffered from small cohort sizes, limiting their ability to detect statistical differences in microbial taxa and population metrics. Second, the studies utilized 16S sequencing for community profiling, which cannot resolve bacterial species/strain level variation, cannot identify other microbes such as viruses, and underperforms relative to metagenomic sequencing in identifying low abundance taxa¹⁵. Additionally, these studies found different changes in the SCD microbiome, leading to differing conclusions about its relationship with the disease. A robust evaluation of SCD patients’ gut microbiomes, together with clinical data, is necessary to determine whether there exists a relationship between the microbial community and disease pathology.

Microbiome markers have been shown to stratify treatment response in other diseases^{16,17}, providing direction for the development of clinically-meaningful gut microbiome biomarkers. Additionally, gut bacteriophage populations have been recognized for their association with chronic disease¹⁸ and have been shown to shift life-cycle distribution at the population level in inflammatory bowel disease¹⁹. Interrogation of the SCD microbiome, in conjunction with studying the effect of AN levels in SCD patients, could identify novel biomarkers of disease severity in SCD. Thus, we conducted a study of blood and stool samples from patients with SCD compared with age and race matched controls. Using whole community metagenomics, we investigated the gut microbiome on multiple axes, revealing complex yet informative patterns of microbiome change.

By comparing gut microbiome signatures with blood AN and cytokine levels, we begin to characterize the pathological interplay between the gut microbiome and immune system activation in SCD.

Results

Sickle cell disease patient cohort

Patients and controls in the study are followed at a large hospital system in the United States. Cohorts were matched for age, sex, ethnicity, and race (Supplemental Table 1 [↗](#)). For patients, clinical history (Supplemental Table 2 [↗](#)) and treatment history (Supplemental Table 3 [↗](#)) were aggregated from the clinical record. Additionally, laboratory measurements were collected from the clinical record by their most recent reading (Supplemental Table 4 [↗](#)) or assayed as part of the study (Supplemental Table 5 [↗](#)).

Gut microbiome changes in sickle cell disease

To study the SCD gut microbiome, we sequenced fecal samples from patients and controls. Sample sequencing reads were profiled for microbial taxa using a library of taxa-specific marker genes ²⁰. To determine whether there is a shift in the SCD gut microbiome at the community level, we used Shannon diversity to compare the species richness between patients and controls and found that SCD patients have less diverse gut microbiomes (Figure 1a [↗](#)). To identify microbes positively or negatively associated with SCD, we modeled taxa abundance as a function of age, sex, ethnicity, race, and SCD status using a multivariable generalized linear model (Supplemental Table 1 [↗](#)). Overall, 25 taxa are significantly associated with SCD and 6 taxa are associated with age (Figure 1b [↗](#)). We noted that SCD samples had higher abundances of the phylum Bacteroidetes (B) and lower abundances of the phylum Firmicutes (F), the two taxonomic groups that make up the composite metric F:B ratio, and observe a reduction in the ratio in our study (Figure 1c [↗](#)), as had been observed previously ¹². Looking at human gut microbiome indicator species identified in a meta-analysis of human microbiome studies ²¹, we observed a decrease in health indicators ($p=5.6e-9$; Mann–Whitney–Wilcoxon test) and increase in disease indicators ($p=4.4e-3$; Mann–Whitney–Wilcoxon test) in SCD patients compared to controls (Figure 1d [↗](#)), pointing to changes in the SCD gut microbiome that are consistent with changes observed in other human diseases.

We also performed a functional characterization of samples using gene-based pathway annotation ²⁰. We again modeled abundance with the same fixed effects but included an additional effect, the F:B ratio, to both capture community-level changes and identify SCD-specific changes that are independent of the high-level shift (Figure 1e [↗](#)). High F:B ratio had exclusively negative associations that were significant, with the most significant associations with pathways related to nucleotide biosynthesis. SCD status was positively associated with three pathways related to butyrate fermentation, fatty acid oxidation, and vitamin B6 biosynthesis, and negatively associated with a pathway related to anaerobic metabolism. The small number of pathways robustly associated with SCD in our heterogeneous patient population point to specific functional changes that shed light on conserved contributions of the gut microbiome to SCD pathology.

Gut microbiome markers and known SCD biomarkers

We investigated whether gut microbiome changes in SCD are related to the AN population by asking whether microbiome markers correlate with the fraction of ANs of the total neutrophil population in a patient's blood (Supplemental Figure 1 [↗](#)). For this analysis, we utilized a subset of the SCD cohort whose blood was assayed ($n=57$). Controls were not assayed for ANs and are excluded from the following analysis. We evaluated four groups of markers: (1) species positively associated with SCD-*E. lenta*, *F. plautii*, and *C. bolteae*; (2) species negatively associated with SCD-*O. sp. 57* ²⁰, *E. siraeum*, *textitA. shahii*, *B. intestinihominis*, *R. callidus*, *B. angulatum*, *C. catus*, *S. isoflavoniconvertens*, *M. smithii*, *O. sp CAG 241*, and *C. comes*; (3) functions significantly different between patients and controls-succinate fermentation to butanoate, fatty acid beta oxidation VI mammalian peroxisome, and fatty acid beta oxidation II; and (4) system-level summary metrics

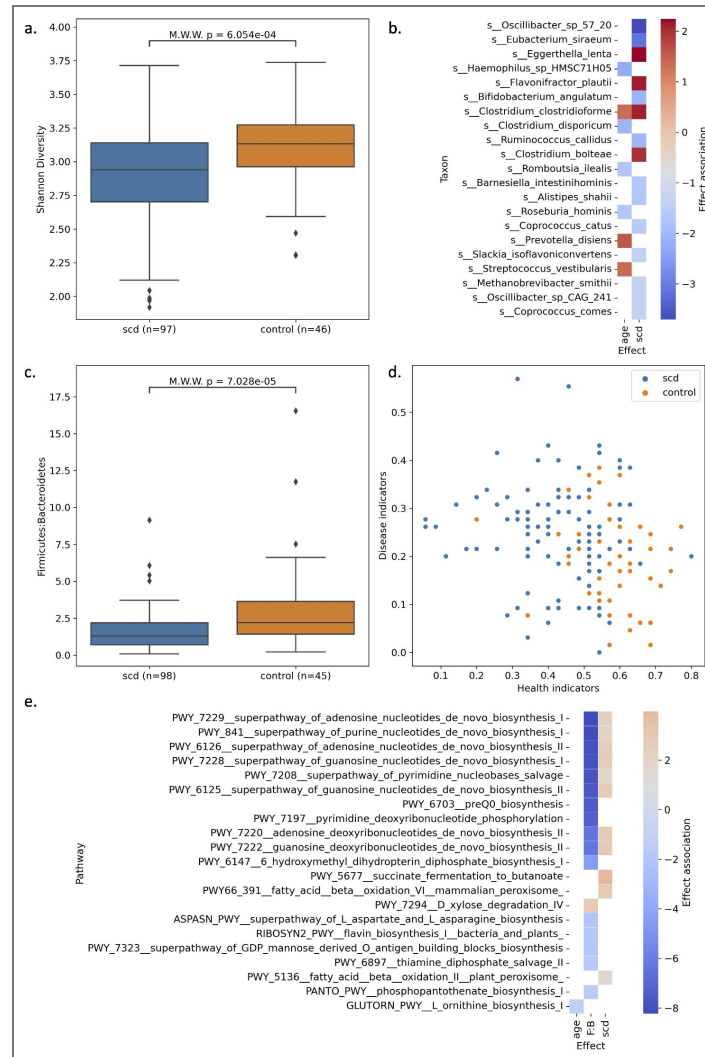


Figure 1. Gut microbiome community changes in SCD patients (n=98) compared to healthy controls (n=46).

Sample whole community sequencing was profiled using the MetaPhlAn taxa marker database²⁰. (a) Distribution of sample alpha diversity measured as Shannon diversity for patients and controls. (b) A generalized linear model was used to determine the effect size and association of bacteria taxa abundance with sample metadata including SCD status, age, race, ethnicity, and gender, all modeled as fixed effects. Taxa with q-value < 0.05 for any effect are shown. (c) Distribution of Firmicutes to Bacteroidetes (F:B) ratio in the gut microbiome of patients and controls. (d) Scatter plot of health and disease indicator scores per sample. (e) Sample whole community sequencing was profiled using the HUMAnN pathway marker database²⁰. A generalized linear model was used to determine the effect size and association of pathway abundance with sample metadata including SCD status, age, race, ethnicity, and gender, all modeled as fixed effects. Additionally, the F:B was also included as fixed effect. Pathways with q-value < 0.05 for any effect are shown. In a and c, samples with values ± 3 s.d. from mean of all samples were removed. M.W.W.-Mann-Whitney-Wilcoxon test.

shown to be differential between patients and controls-Shannon diversity, F:B ratio, health indicators, and disease indicators. A single marker, the bacterial species *A. shahii*, significantly correlated positively with the AN percent ($\rho = 0.26$, $p=0.048$).

Given the substantial changes in the gut microbiome of SCD patients, we next investigated whether gut microbiome features are associated with SCD clinical and molecular measures of disease severity. We utilized a molecular blood panel to profile SCD patients for levels of cytokines and chemokines (Supplemental Table 5 [\[3\]](#)). Additionally, we included a number of hemolysis-related clinical measurements (Supplemental Table 4 [\[3\]](#)).

We first evaluated the AN fraction given its known role in SCD pathology. Neutrophil activation assay measures had significant correlations with white blood cell measures collected from the health record (Supplemental Figure 2 [\[3\]](#)), providing confidence in comparing AN data with clinical data. ANs were significantly correlated with four inflammatory cytokines: interferon gamma, interleukin-1 β , interleukin-10, and interleukin-17A, and two chemokines: monocyte chemoattractant protein-1 and macrophage inflammatory protein-1B (Figure 2 [\[3\]](#), top). In contrast, most of the gut microbiome metrics and all of the community metrics did not correlate with inflammatory markers (Figure 2 [\[3\]](#), bottom). Exceptions to this pattern are *A. shahii* and *B. angulatum*, both negatively associated with SCD and inversely correlated with multiple marker levels. Gut microbiome features did have significant correlations with clinical disease markers, including absolute reticulocyte count (ARC), total bilirubin, and indirect bilirubin.

Gut microbiome viruses

Having observed limited correlation between bacterial gut microbiome markers and ANs, we interrogated the viral population in metagenomic samples. Assembled metagenomes were profiled for viral sequences using a hybrid approach that combines viral marker genes and sequence features to distinguish chromosomal, plasmid, and viral sequences ²². Overall, there was a decrease in the number of viral sequences in SCD compared to controls (Figure 3a [\[3\]](#)), likely reflecting the loss in diversity observed in SCD samples given its correlation with sample Shannon diversity and health indicators ($p=2.6e-5$; Mann-Whitney-Wilcoxon test) (Figure 3c [\[3\]](#)). Proviruses are lysogenic virus sequences integrated into host genomes and can be labeled in assembled viral genomes by identifying sequence features such as viral integrase proteins and integration sites ²³. Proviruses constitute a small fraction of the viral sequences present in samples, yet there was a significant enrichment of provirus sequences in SCD compared to controls ($p=4.7e-4$; Mann-Whitney-Wilcoxon test) (Figure 3b [\[3\]](#)), which was robust to provirus or lysogenic viral sequence calling method (Supplemental Figure 2 [\[3\]](#)). While the provirus fraction was not correlated with bacterial community-level metrics Shannon diversity and F:B, it was inversely correlated with health indicators and positively correlated with disease indicators (Figure 3c [\[3\]](#)), providing evidence that provirus enrichment mirrors changes in the bacterial species present in the sample.

From assembled provirus sequences it is not possible to determine whether a provirus is active, meaning replicating independently of its host via a lytic lifecycle, or inactive. Using a method developed previously ²⁴, we analyzed the sequencing read distribution of provirus sequences and their flanking host sequences to determine if there was increased read coverage of proviruses, which would indicate independent replication. Across SCD samples, >90% of proviruses were inactive (Figure 3d [\[3\]](#)), and the active provirus fraction distribution in SCD was not different from controls ($p=3.208e-1$; Mann-Whitney-Wilcoxon test). Finally, we performed the same correlation experiment for viral-based metrics as we performed for bacterial markers. While the provirus fraction did not significantly correlate with AN fraction ($\rho = 0.02$, $p=0.91$), the provirus fraction had a number of statistically significant positive correlations with molecular cytokines (Figure 3e [\[3\]](#)), which was not true of non-integrated temperate phages (Supplemental Figure 4 [\[3\]](#)).

Integrated viruses

After finding that the provirus fraction in the gut microbiome correlated with the level of multiple inflammatory cytokines in SCD patients, we decided to further investigate the provirus sequences. A total of 4,962 proviruses were identified across the SCD patient gut microbiomes. Provirus

Figure 2. Aged neutrophil percentage (n=57) and gut microbiome feature (n=98) correlations with clinical measures and blood molecular immune markers.

Correlations were measured with Spearman ρ and significance was measured using permutation testing. Value of correlation is colored for strength of correlation where $p < 0.05$, otherwise comparison cell is gray. For the number of samples assayed for each clinical and molecular measure, see Supplemental Tables 4 [4](#) [5](#).

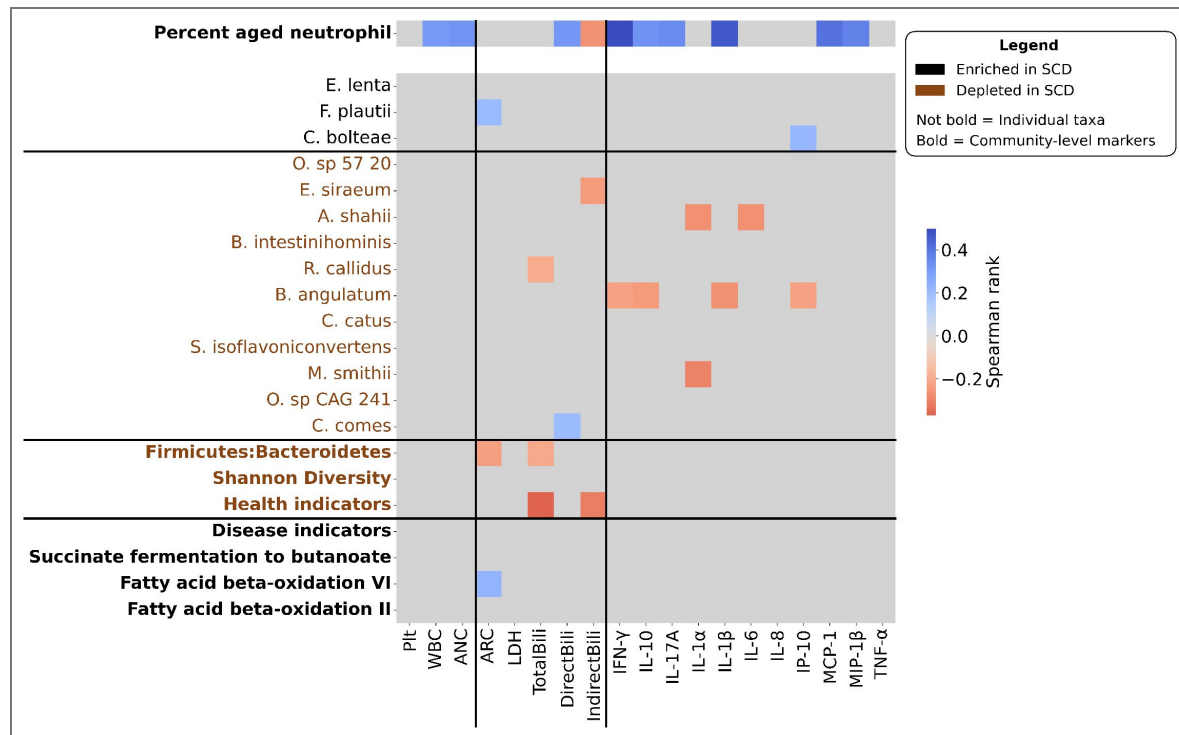
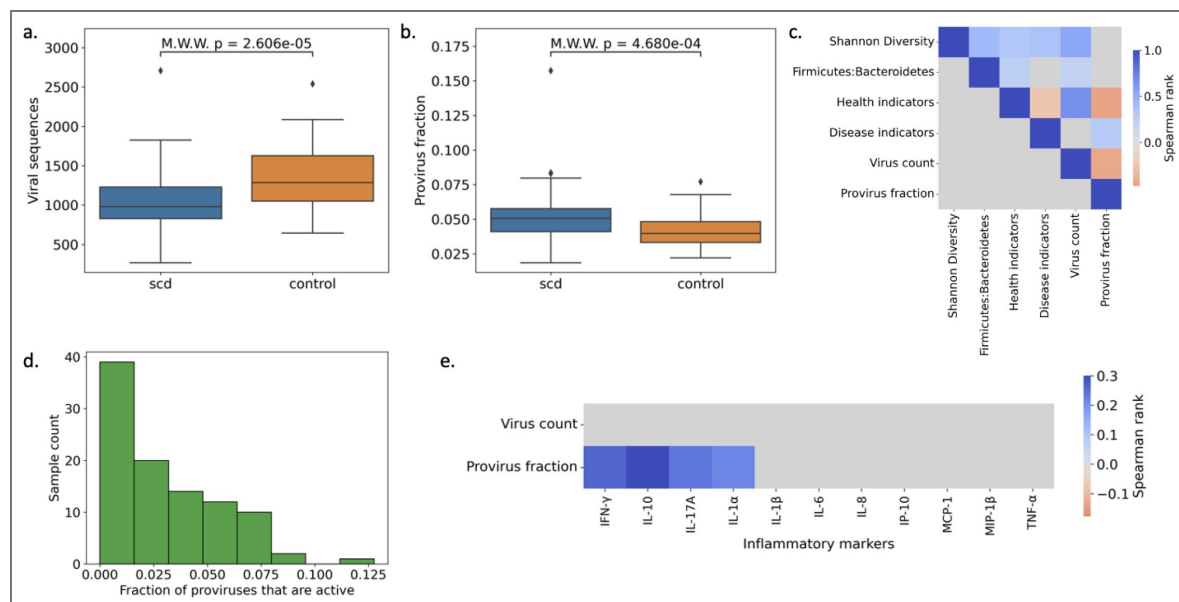


Figure 3. Gut microbiome viral population is altered in SCD.

Sample whole community sequencing was assembled and viral sequences were identified using marker genes and sequence features. (a-b) The number of viral sequences (a) and the fraction of provirus sequences (b) between patients and controls. (c) Comparison of viral population and bacterial population community-level metrics. (d) Histogram of active provirus fraction for SCD samples. (e) Correlation of viral community-level metrics with clinical and molecular markers as in 3b and 3c. Value of correlations is shown for comparisons with $p < 0.05$ otherwise comparison cell is gray. M.W.W = Mann-Whitney-Wilcoxon test.



lengths had a bimodal distribution with a median size of 23,775 base pairs (Supplemental Figure 5 [↗](#)). All proviruses with predicted taxonomy (99.5%) were bacteriophages of the class Caudoviricetes (99.3%), meaning the proviral sequences are prophages.

Because the vast majority of phages were categorized as inactive in host bacterial genomes, we wondered whether these sequences would be similar between patients. We clustered all prophage sequences at 99% identity over at least 70% bi-directional coverage and found 395 clusters of sequences with median size of 2 (Figure 4a [↗](#)). In total, 1,249 (25%) sequences had a nearly identical sequence in another sample in the dataset. These clustered prophages were homologous to sequences in common gut commensals, however *Bacteroides* stood out with 113 clusters matching species in the genus (Supplemental Figure 6 [↗](#)). The decreased F:B ratio in our patient population may also represent an increase in *Bacteroides* strains carrying these conserved prophages, although at the community level, the prophage fraction did not inversely correlate with the F:B ratio as would be expected if it were such strains driving the community shift. To determine whether clustered prophages differed from singletons, we compared the high-level functional content of sequences in each group (Figure 4b [↗](#)). Interestingly, clustered prophage sequences had lower fractions of genes with functions related to lysis, transcriptional regulation, and host takeover and a higher fraction of genes with unknown function compared to non-clustered prophages. A reduction in proteins that function in host lysis was also observed in active prophages compared to dormant prophages (Supplemental Figure 7 [↗](#)).

Discussion

The primary goal of our study was to ask whether there exist consistent, significant interactions between the gut microbiome and disease pathology in SCD. Such interactions, when observed across heterogeneous patients with varying clinical histories and burdens of disease, could lead to novel treatment modalities and diagnostics to improve patient care and monitoring in this disease of increasing importance around the world. With orthogonal computational and experimental approaches, we can begin to build a model of the interactions between gut microbiome features, immune system activity, and pathology in SCD.

Our study reveals a two-tiered dialogue between the gut ecosystem and systemic inflammation in SCD (Figure 5 [↗](#)). Shotgun metagenomics of 98 patients showed a disease-wide contraction of bacterial diversity, a drop in the F:B ratio, and characteristic changes in pan-disease indicator taxa. These shifts parallel patterns seen across disparate chronic illnesses ^{21,25}, suggesting that SCD provokes a generic disease signature at the microbiome community level.

Neutrophil biology and the microbiome appear to occupy distinct, complementary niches. The AN subset correlated with an overlapping cytokine panel but not with bacterial community metrics. Conversely, bacterial α -diversity and indicator taxa associated with indirect bilirubin and reticulocyte count—clinical proxies for hemolysis, yet show little relationship to the immune milieu. This dissociation implies that neutrophil activation and microbial changes represent parallel arms of SCD pathophysiology, each informative for a different disease facet.

Beyond bacteria, dormant prophages emerge as a previously unrecognized axis of SCD biology. We detect a 1.7-fold increase in the prophage fraction, yet >90% of these elements are predicted to be replicatively silent. Strikingly, the proportion of prophages, but not total phage population or lytic phage fraction, tracks with IL-1 α , IL-10, IL-17A and IFN- γ levels. Immune signaling molecules are known to be elevated at baseline in SCD patients ^{26–28}, and 3/4 cytokines correlated with prophage level are observed to be lower in patients in vaso-occlusive crisis compared to SCD patients not in vaso-occlusive crisis, including IL-10, IL-17A, IFN γ ²⁶. Interestingly, the gut microbiome in SCD patients harbors prophage sequences that are highly conserved between unrelated individuals, with one such sequence present in over 50% of metagenomes. These conserved, shared prophages were an unexpected feature of SCD patient microbiomes, and stand in contrast to studies of lytic phages, that found non-integrating bacteriophage populations are highly individual-specific ²⁹. We found that clustered sequences had a higher fraction of genes coding for proteins involved in host takeover, gene expression manipulation, and lysis. We speculate that lysogens bearing ‘domesticated’ prophages maintain a survival advantage in the setting of a disease-associated gut

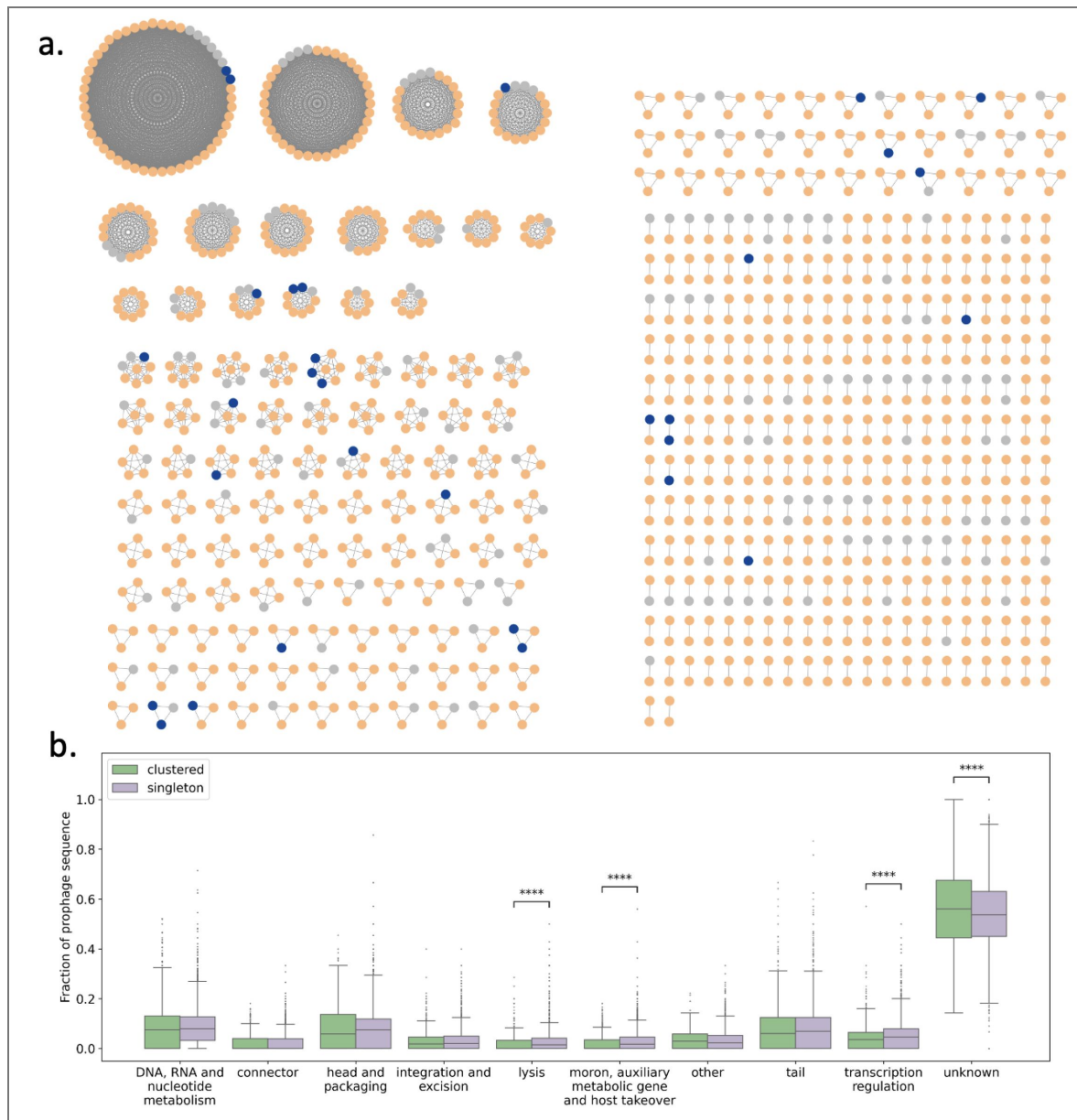


Figure 4. Highly conserved prophage sequences are shared by gut microbiomes of SCD patients.

(a) Patient microbiomes sharing a highly conserved prophage sequence visualized as fully connected networks. Conserved prophages had 99% sequence identity over at least 70% of bidirectional sequence coverage. Node color indicates predicted prophage activity: active (blue), dormant (orange), and undetermined (gray). (b) Prophage sequence distribution of high-level viral protein functions compared between clustered (n=1,249) and singleton (n=3,713) prophages. Significance tested with a Mann-Whitney-Wilcoxon test: **** = $p < 0.0001$.

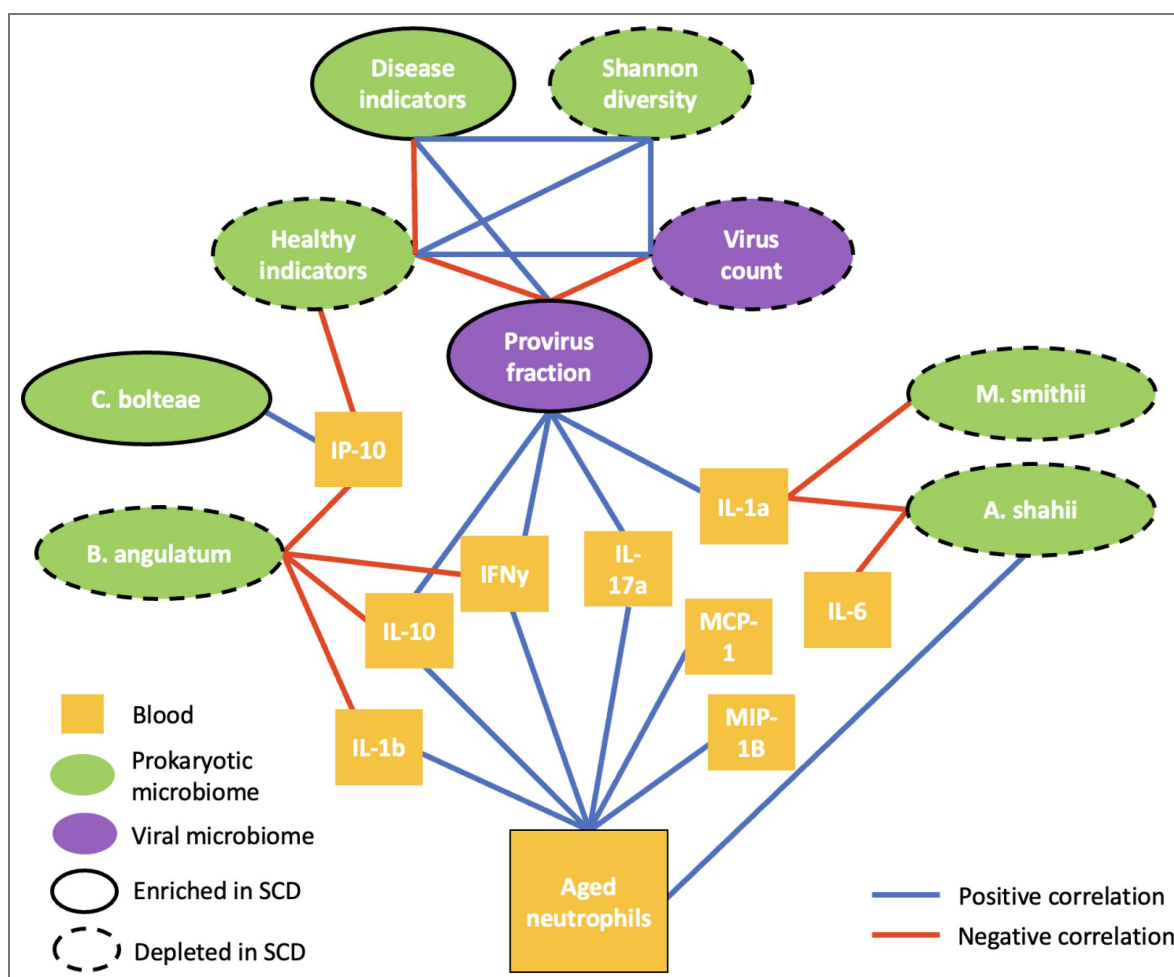


Figure 5. A model of interactions between the gut microbiome and markers of inflammation in the blood of SCD patients.

The gut microbiome can influence and is influenced by the immune system. Using multiple data modalities, we begin to understand the complex interaction of gut microbiome changes and immune system activation in the pathology of SCD.

microbiome. Enrichment of lysogens may modulate host immunity by altering bacterial surface antigens or metabolite flux. Experimental induction assays and longitudinal sampling will be needed to test whether prophage excision precedes cytokine spikes or merely registers host stress.

Functional signature investigation of the gut microbiome, enabled by metagenomics, can have therapeutic implications³⁰. Among the functional module changes observed in SCD, three merit particular attention: (i) enrichment of nucleotide biosynthesis pathways, (ii) succinate to butanoate fermentation, a non-canonical route for microbial butyrate generation, and (iii) fatty-acid β -oxidation pathways, which typically require molecular oxygen and are not common in the anaerobic gut environment. The communal shift toward a low F:B ratio was tightly coupled to an expansion of nucleotide-biosynthesis modules, offering a narrow-spectrum antimicrobial target set that could re-balance the ecosystem more precisely than broad antibiotics trialed to date^{7–9}. Looking at functional modules enriched in SCD independent of community shift, we see the succinate to butanoate butyrate production pathway; a departure from the primary gut butyrate acetyl-CoA pathway³¹. Butyrate is the primary energy source for gut epithelial cells and has been shown to be anti-inflammatory and supportive of epithelial-barrier integrity³². This shift in butyrate production pathway in the SCD gut may reflect the loss of members of taxonomic groups that produce butyrate via complex carbohydrate fermentation, such as *C. catus*, *C. comes* and *E. siraeum*^{31,33}. Conversely, elevated β -oxidation genes suggest greater flux through short- and medium-chain fatty-acid catabolism, a shift that might deplete beneficial metabolites and warrants metabolic follow-up. Notably, the SCD enriched β -oxidation pathways II and IV are mitochondrial and peroxisomal, and therefore generally aerobic, pointing to the availability of molecular oxygen in the gut, a hallmark of gut barrier disruption^{34,35}.

There are a number of recognized limitations for our study. First, while we collectively grouped SCD patients into a single cohort, the disease can vary in its presentation and clinical course, which may affect the gut microbiome. Our cross-sectional design precludes causal inference and we acknowledge that as a chronic disease, SCD presentation can be highly variable. Second, flow-cytometry assays for neutrophil activation markers were performed within 8 hours of venipuncture, and while we have correlative evidence that our results reflect accurate measures, we did not formally benchmark marker stability to this time window. Next, prophage activity was inferred from coverage ratios, as is common in the field, but with this method asynchronous induction events could evade detection. Finally, findings derive from a single U.S. center and may not generalize to global SCD populations with different diets or genotypes.

In summary, we propose that the gut microbiome bacterial shift mirrors hemolytic burden while ANs and dormant prophages mirror immunological activation tone. Longitudinal and interventional studies, particularly those integrating metatranscriptomics and targeted bacteriotherapy, are now warranted to unravel causality and translate these markers into clinical tools.

Methods

Study design and subjects

The study is a single-center outpatient cohort study of SCD patients and age- and race-matched controls. Study participants were recruited from the Children's Hospital at Montefiore and Montefiore Medical Center (IRB NUM-BER: 2018-9080). Patients were enrolled with the following inclusion criteria: 1) diagnosis of sickle cell anemia (sickle cell disease SS genotype or sickle beta zero thalassemia, $S\beta^0$ -thal), or sickle cell trait in association with hereditary persistence of fetal hemoglobin (S-HPFH); 2) age ≥ 4 ; and 3) currently in a usual state of health. Exclusion criteria included: 1) malignancy; 2) inflammatory bowel disease or other gastrointestinal disorder; 3) immunocompromised due to additional underlying disease or immunosuppressive medication; 4) history of hematopoietic stem cell transplant; 5) history of *C. difficile* infection in the preceding 2 months; 6) receipt of chemotherapy in the preceding 2 months; 7) receipt of systemic antibiotic other than Pen VK in the preceding 2 months; 8) Hospitalization within the preceding 2 weeks; 9)

intercurrent febrile illness or sickle cell related pain episode requiring opioids within the preceding 2 weeks. Controls were included with age ≥ 4 and excluded with the same criteria in addition to any systemic antibiotic use in the preceding 2 months.

In total, 101 SCD patients and 66 controls were enrolled in the study from 2018-2020. Of these participants, 98 SCD patients and 46 controls provided stool samples. One patient sample did not have enough material for metagenomic sequencing. Fecal samples were collected by study participants into DNA/RNA Shield Fecal Collection tube (Zymogen Research) and samples were split into 6 tubes and stored at -80°C . Blood was also drawn from study participants around the time of fecal collection.

Clinical and molecular data

Clinical data was collected for SCD patients based on chart review. Information included: (i) demographics (age (years), sex, ethnicity, race, hemoglobin genotype (SS, $\text{S}\beta^0$, SS with high F, AS, AA, where 'S' indicates a sickle cell allele, ' β^0 ' indicates a thalassemia allele coding for no wild type hemoglobin, 'F' indicates a fetal hemoglobin allele, and 'A' indicates a wild-type allele); (ii) patient history-hydroxyurea (on/off), folic acid (on/off), glutamine (on/off), stroke (yes/no), asthma (yes/no), allergies (yes/no), bacteremia (yes/no), silent infarct (yes/no), history of meningitis, osteomyelitis, or urinary tract infection (yes/no), number of acute chest syndrome past year, number of pain admissions past year, number of 30 day readmissions past year, number of emergency room visits past year, number of transfusions in the past year, number of exchanged transfusions in the past year; and (iii) clinical measurements-weight (kg), body mass index (BMI), urine microalbumin (MA/Creatinine; mg/gm), lactate dehydrogenase (LDH; U/L), platelet (PLT; k/uL), white blood cells (WBC; k/uL), absolute neutrophil count (ANC; k/uL), hemoglobin (Hb; g/dL), sickle hemoglobin (HbS; %), fetal hemoglobin (HbF; %), wild type hemoglobin (HbA; %), absolute reticulocyte count (ARC; k/uL), total bilirubin (mg/dL), direct bilirubin (mg/dL), indirect bilirubin (mg/dL), alanine aminotransferase (ALT; U/L), Creatinine (mg/dL), iron studies: iron level (ug/dL), transferrin (ug/dL), saturation (%), total iron binding capacity (TIBC; ug/dL), ferritin (ng/dL). All yearly measures were annualized to 1 year if patient records are < 1 year in duration. Acute care in the past year is the sum of pain admissions and emergency department visits.

Venous blood in Acid Citrate Dextrose was collected from study participants. Inflammatory cytokine and chemokine levels were measured from serum using the MILLIPLEX Human Cytokine/Chemokine/Growth Factor Panel A kit (MilliporeSigma, HCYTA-60K). Luminex assay data were acquired on a Luminex MAGPIX and analyzed with the MILLIPLEX Analyst program (MilliporeSigma). The following cytokines and chemokines were profiled: G-CSF (pg/mL), IFN γ (pg/mL), IL-10 (pg/mL), IL-17A (pg/mL), IL-1 α (pg/mL), IL-1 β (pg/mL), IL-6 (pg/mL), IL-8 (pg/mL), IP-10 (pg/mL), MCP-1 (pg/mL), MIP-1 β (pg/mL), TNF- α (pg/mL). Cytokines and chemokines with $< 50\%$ of samples assayed at the lower limit of detection were excluded from further analysis.

Neutrophil activation biomarker assessment

Blood was also evaluated by flow cytometry for neutrophil adhesion and activation markers. Flow cytometry was performed using LSRII equipped with FACS Diva 8.0.1 software (BD Biosciences) and analyzed with FlowJo software (Tree Star). Neutrophils were identified by forward and side scatter characteristics combined with CD16b expression, and the AN subset evaluated by CD62L^{lo}CXCR4^{hi} expression within the neutrophil population. Whole blood from the same sample used for flow cytometry assays was diluted 1:10 in PBS for complete blood count with differential counts on ADIVA 120 (Siemens Healthcare Diagnostics). Total WBC and absolute neutrophil counts along with percent AN from flow cytometry were used to calculate the absolute aged neutrophil count. Samples were excluded if analyses were not performed within 8 hours of blood draw.

DNA extraction and sequencing

Sample preparation and sequencing was done by the Molecular Microbiology Facility of the Integrated Genomics Operation at Memorial Sloan Kettering Cancer Center. DNA was extracted from samples using a custom phenol chloroform extraction optimized for fungal and bacterial

isolation. Metagenomic sequencing was performed using next generation sequencing with the Illumina HiSeq platform using 2x150bp paired end reads. For some samples, multiple sequencing runs were performed to achieve targeted sequencing depth. Base calling was performed by platform software resulting in paired FASTQ files for each sample. Samples with multiple runs were concatenated to single paired files.

Metagenomic sequence profiling

For bacterial taxa and function analysis, the bioBakery 3 suite of metagenomic sequence tools was used²⁰. First, raw reads were first processed for quality and removal of human contamination using kneaddata (v0.10.0) with the following parameters -t 8 -p 12 -cat-final-output²⁰. Bacterial taxa were profiled with MetaPhlAn3 (v3.0) with default parameters and -no map -nproc 12 and utilized the mpa v30 CHOCOPhlan 201901 clade-specific marker gene database²⁰. MetaPhlan output for all samples were collapsed to a single table using the merge metaphlan tables.py utility script²⁰. Function pathways were profiled with HUMAnN3 (v3.7)²⁰. Paired read files were merged to a single table and HUMAnN was run with default parameters and -threads 96²⁰ and database versions mpa vOct22 CHOCOPhlanSGB 202212 and uniref90 201901b full.dmnd. For comparison between samples, pathway profiles were normalized for read sequence depth using the script humann renorm table to copies per million with parameters -units cpm -update-snames; stratified profiles were split using the humann split stratified table script; and merged to a single table using the humann join tables script²⁰.

For viral profiling, raw sequences were processed with the nextflow nf-core/mag (v2.4.0) metagenomics pipeline^{36,37} for quality and contamination removal, assembly with MEGAHIT³⁸, and gene calling with prodigal³⁹. The pipeline was run with-profile singularity and parameters -skip binning -skip spades. Viral sequences were profiled from MEGAHIT assemblies using geNomad (v1.7.1) with the end-to-end command and default parameters²² and using VIBRANT (v1.2.1) using the VIBRANT run.py script with -f nucl -t 8 -no plot parameters⁴⁰. Both methods predict provirus sequences but only VIBRANT predicts lytic vs. lysogenic bacteriophage life cycle. To calculate the non-integrated lysogenic virus count in the VIBRANT data, the count of provirus sequences was subtracted from the count of lysogenic viruses per sample. PropagAte (v1.1.0) with default parameters was used to determine the activity of provirus sequences²⁴ from raw sequence reads, MEGAHIT assemblies, and provirus coordinates called by both geNomad and VIBRANT. All fraction metrics were calculated by dividing the count of interest by the total number of viruses called by the respective method (e.g. the provirus fraction for geNomad is the count of proviruses divided by the total count of viruses per sample).

SCD metagenome analysis

Multiple metrics were calculated for the bacterial population in each sample using its MetaPhlan profile. Shannon diversity was calculated using the MetaPhlan utility script calculate diversity.R script with parameters -d alpha -m shannon -s s. The F:B ration was calculated by dividing the abundance of 't Firmicutes' by the abundance of 't Bacteroidetes. Health and disease associated taxa were collected from a gut microbiome meta-analysis²¹, and the number of taxa in each set was counted based on the presence of the taxa at any abundance in the sample, as performed in²¹.

To determine bacterial taxa and pathways differentially abundant in SCD patient samples compared to controls, we utilized a multivariate generalized linear model approach to model profiles as log-linear with MaAsLin 2 (R; v1.8)⁴¹. For taxa, MaAslin2 was run with min prevalence = 0.1 and min abundance = 0.0001. For pathways, MaAslin2 was run with min prevalence = 0.1 and normalization = 'NONE'. We included a number demographic variables as fixed effects, modeling

$$\begin{aligned} \text{taxa} &\sim \text{scd} + \text{age} + \text{race} + \text{ethnicity} + \text{sex} \\ \text{pathways} &\sim \text{scd} + \text{age} + \text{race} + \text{ethnicity} + \text{sex} + F : B \end{aligned}$$

where *scd* is an indicator variable for the disease status of a sample and abundance refers to either taxa or pathways. Effect heat maps were generated from MaAslin2 significant result tables, and significant effects were reported for $q < 0.05$.

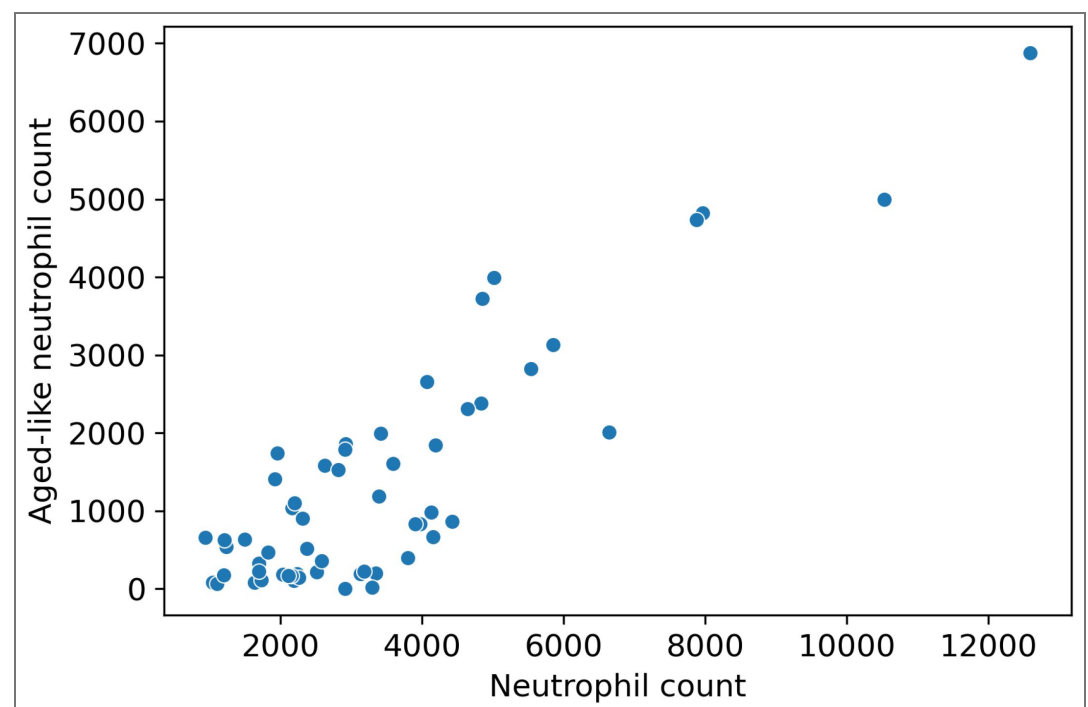
Prophages investigation

Prophage sequences identified using geNomad were clustered at 99% sequence identity over $\geq 70\%$ of bidirectional sequence coverage using *mmseqs2* (v14.7e284) ⁴² with parameters `–cluster-mode 0 –cov-mode 0 –c 0.7 –min-seq-id 0.99`. Clusters were converted to fully connected networks using the python package *networkx* (v3.1) ⁴³ and visualized using *cytoscape* (v3.10.0) ⁴⁴. VPF-PLM ⁴⁵ was used for functional annotation of prophage sequences. Clustered prophage homology to bacterial species sequences was done using *blastn* (v2.16.0+) and the core nt database with parameters `eval 1e-10` and `perc identity 90`. The best hit bacterial species was assigned to each cluster where such a hit was identified.

Statistical analyses and visualizations

Mann–Whitney–Wilcoxon was used to compare distribution means between SCD patients and controls for Shannon diversity, F:B ratio, healthy indicator fraction, and disease indicator fraction. Values were removed in ± 3 standard deviations from mean of all study samples. Spearman's ρ was used for all tested correlations with the exception of flow cytometry and electronic health record white blood cell measures which was tested with Pearson correlation coefficient. Significance of a correlation was measured using a permutation test with a two-sided alternative hypothesis. Mann–Whitney–Wilcoxon test was used for all tested association of categorical variables test with a two-sided alternative hypothesis. All methods were implemented with *scipy stats* package ⁴⁶; *spearmanr*, *pearsonr*, and *mannwhitneyu* with default parameters and permutation test with `alternative='two-sided'`, `permutation type='pairings'`. Significance was assigned for $p < 0.05$ for all comparisons. Visualizations were produced using python packages *seaborn* ⁴⁷ and *statannotations* ⁴⁸. *Pandas* ⁴⁹ and *numpy* ⁵⁰ python packages were used for analysis.

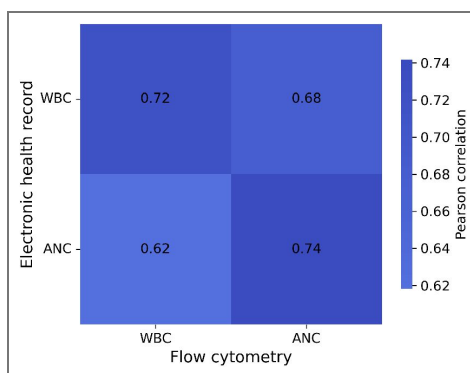
Supplemental material



Supplemental Figure 1. SCD patient blood neutrophil profiling. Neutrophil and AN counts were assayed by flow cytometry.

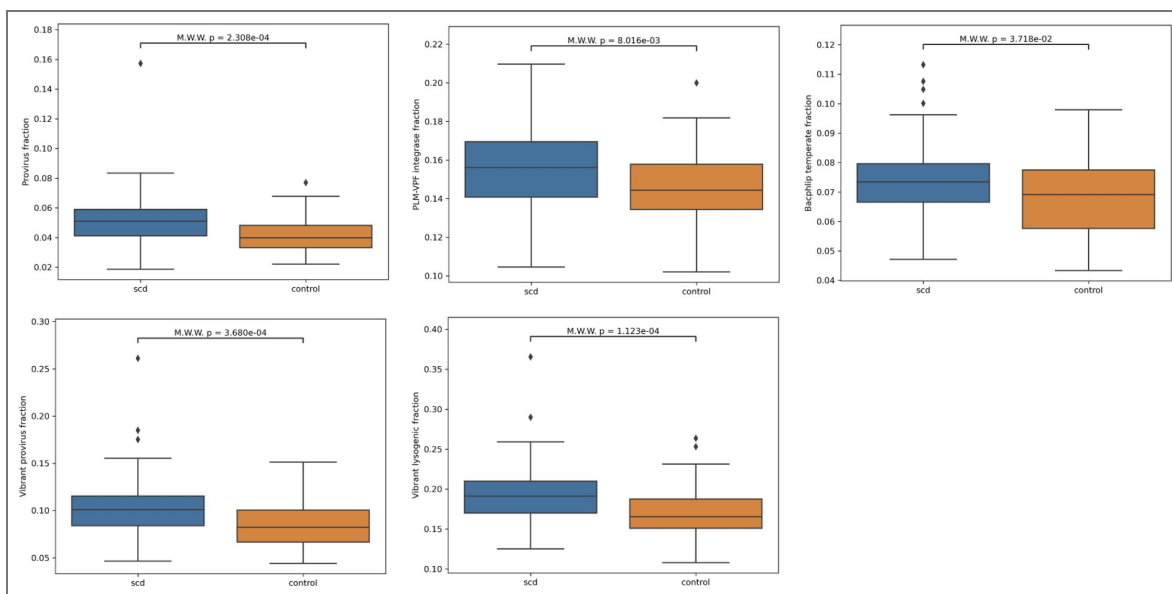
Supplemental Figure 2. Correlation of flow cytometry and EHR measures of patient white blood cell populations.

Correlation measured with Pearson correlation coefficient. Abbreviations: WBC-white blood cells, ANC-absolute neutrophil count.



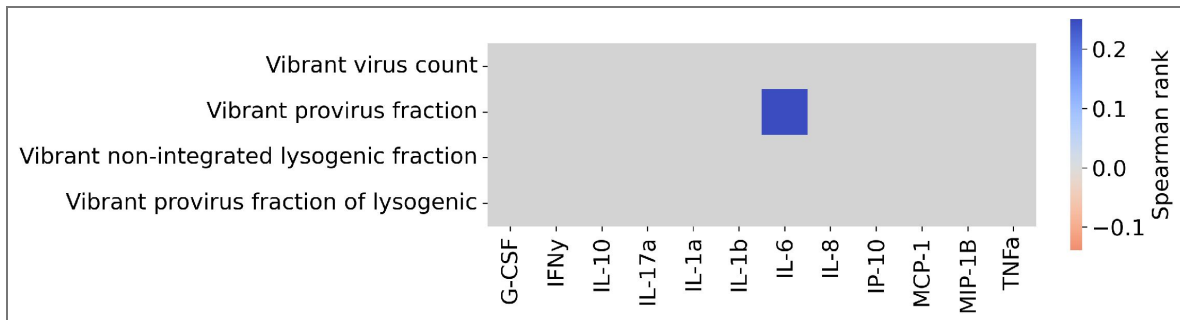
Supplemental Figure 3. Enrichment in provirus or lysogenic virus prediction across multiple labeling methods.

The determination of a lysogenic virus can be made by either looking for the virus sequence integrated into a host genome or by predicting that a viral sequence has lysogenic potential. We utilized multiple methods that rely on different approaches to determine that the provirus enrichment is robust to method.

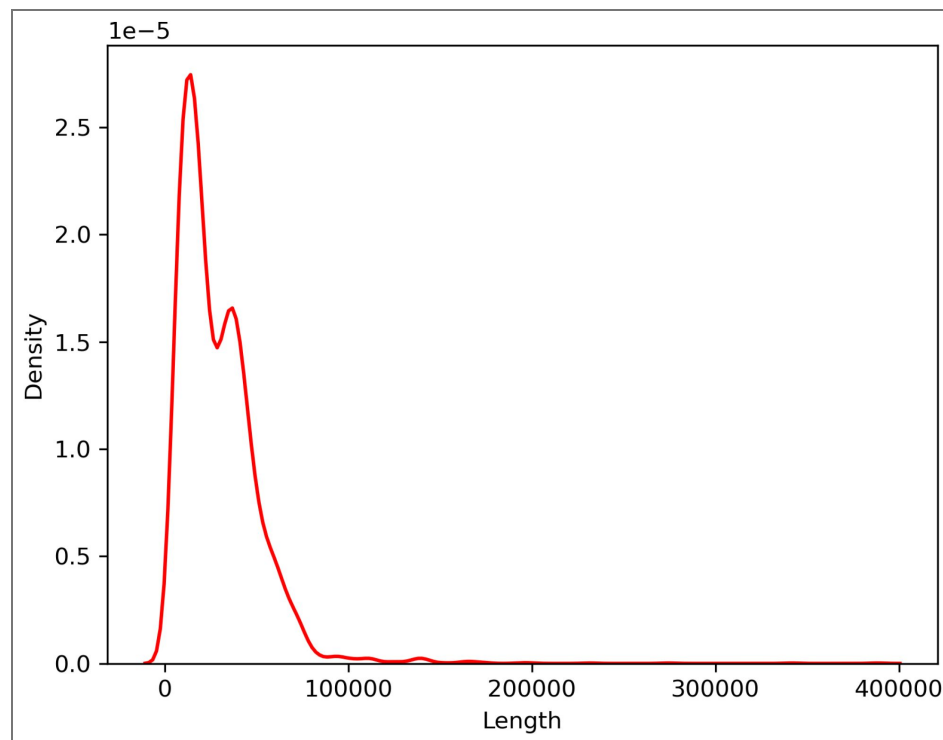


Supplemental Figure 4. Provirus fraction but not non-integrated lysogenic virus fraction correlate with immune cytokines.

Using the vibrant viral calling method, the fraction of non-integrated, lysogenic predicted viruses was quantified and compared to the provirus fraction for correlation with molecular cytokines.

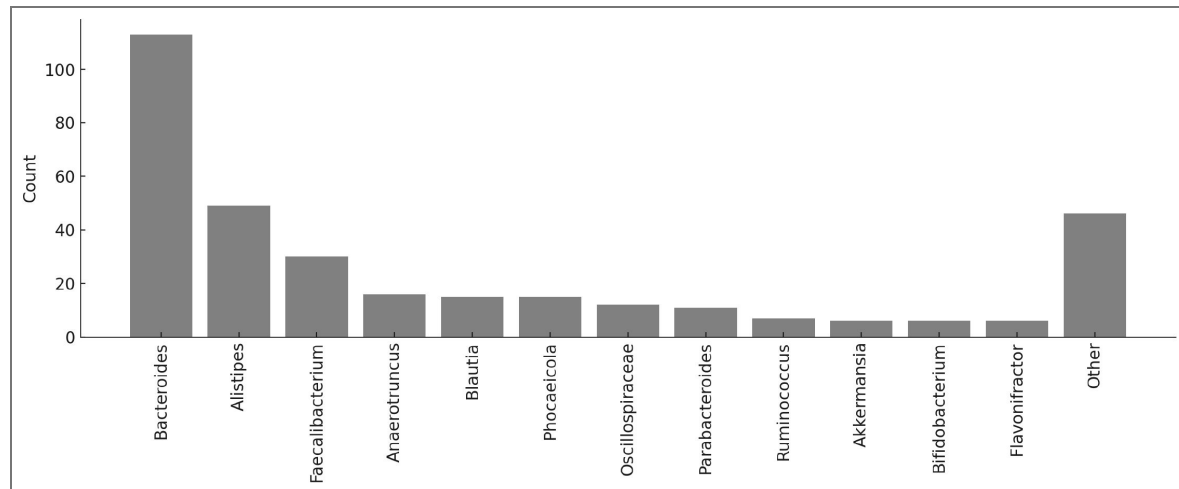


Supplemental Figure 5. Provirus length distribution with kernel density estimation.



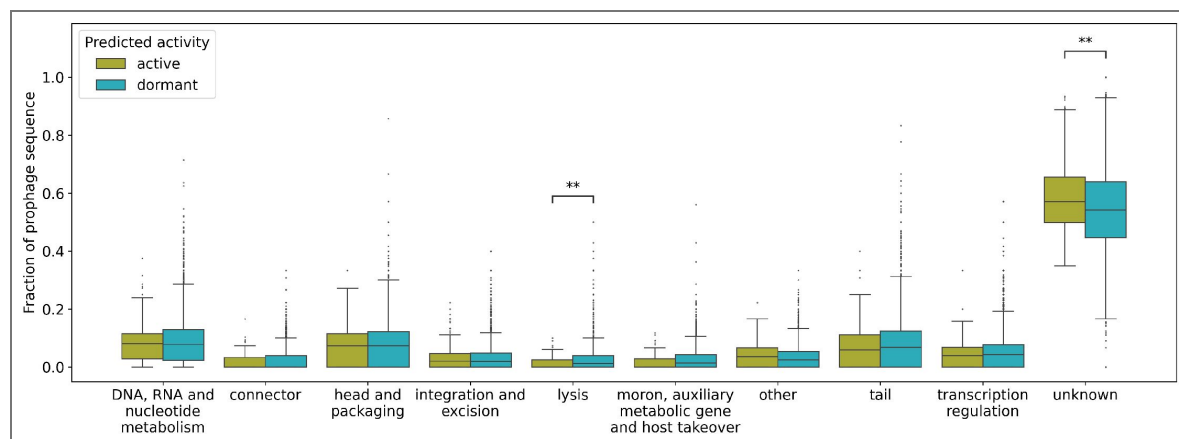
Supplemental Figure 6. Prophage cluster sequence homology with bacterial hosts.

Representative cluster sequences with blast sequence homology to bacterial species sequences were aggregated to the genus level. Count represents the number of clusters homologous to species in the genus.



Supplemental Figure 7. Genome content comparison between predicted active (n=136) and dormant (n=4739) prophages in the SCD gut microbiome.

Significance tested with a Mann-Whitney-Wilcoxon test: * = 0.01 ≤ p < 0.05; ** = 0.001 ≤ p ≤ 0.01.



Supplemental Table 1. Study participant demographics by cohort.

	SCD (n=98)	control (n=46)
Sex (Female) [%]	49	65
Ethnicity (Non-Hispanic; Hispanic; Unknown) [%]	78; 20; 2	74; 26; 0
Race (Black; White; American Indian; Multiple race; Other/Unknown) [%]	80; 1; 1; 1; 17	74; 0; 0; 0; 26
Genotype [Hb type]	SS 92%, Sβ-thalassemia 4%, SCD with high HbF 4%	AS 61%, AA 39%
Age [mean (standard deviation)]	18.2 (12.8)	18.6 (14.9)

Supplemental Table 2. SCD patient (n=98) medical history.

History (number with measure)	SCD patients [%]
Asthma (97)	24
Environmental allergies (97)	43
Stroke (96)	7
Acute chest syndrome (97)	6
Bacteremia (96)	26
History of bacteremia, meningitis, osteomyelitis, or urinary tract infection (96)	30
Acute care visits in past year (97)	45
ED visit in past year (97)	35
Pain admission in past year (97)	22

Supplemental Table 3. SCD patient (n=98) treatment history.

Treatment (number with measure)	SCD patients [%]
HU (98)	62
Folic acid (97)	72
Glutamine (97)	16
Transfusions in past year (97)	37

Supplemental Table 4. Clinical measures for SCD patients (n=98).

Measure (number with measure)	Mean	Standard deviation
BMI [kg/m ²] (92)	19.9	5.0
LDH [U/L] (96)	515.4	190.4
PLT [k/uL] (97)	373.9	123.6
WBC [k/uL] (97)	9.5	3.6
ANC [k/uL] (97)	4.7	2.4
ARC [k/uL] (97)	288.3	136.4
Total bilirubin [mg/dL] (97)	3.1	2.3
Direct bilirubin [mg/dL] (95)	0.4	0.2
Indirect bilirubin [mg/dL] (97)	2.6	2.3
ALT [U/L] (97)	20.1	16.0
Creatinine [mg/dL] (97)	0.6	0.3

Measure (number with measure)	Mean	Standard deviation	Fraction above lower limit of detection
<i>Cytokines</i>			
G-CSF [pg/mL] (70)	18.6	31.9	36%
IFN γ [pg/mL] (85)	20.4	31.7	62%
IL-1 α [pg/mL] (87)	68.4	122.9	99%
IL-1 β [pg/mL] (90)	4.1	8.1	54%
IL-6 [pg/mL] (74)	20.5	33.7	99%
IL-10 [pg/mL] (95)	33.6	82.9	73%
IL-17A [pg/mL] (91)	5.1	8.5	58%
TNF- α [pg/mL] (67)	4.5	5.1	73%
<i>Chemokines</i>			
IL-8 [pg/mL] (94)	49.1	104.6	88%
IP-10 [pg/mL] (91)	329.5	264.8	99%
MCP-1 [pg/mL] (92)	67.5	52.4	99%
MIP-1 [pg/mL] (92)	18.2	9.9	98%

Supplemental Table 5. Molecular inflammatory measures for SCD patients (n=98).

Data availability

The dataset "Sickle cell disease patient gut microbiome study" is available in the repository NCBI and can be accessed via the following accession ID: PRJNA1320713. Patient level clinical, neutrophil, and immune data can be found in the Supplementary Material file.

Acknowledgements

The authors thank Marcel R. M. van den Brink and Jonathan U. Peled at the Molecular Microbiology Facility of Memorial Sloan Kettering Cancer Center for their generous assistance with microbiome sequencing.

Additional information

Ethics approval and consent to participate

This study was approved by the Institutional Review Board of the Albert Einstein College of Medicine and Montefiore Medical Center (IRB 2018-9080). All study participants were recruited from the outpatient clinics at the Children's Hospital at Montefiore and Montefiore Medical Center. Written informed consent was obtained from all participants or their legal guardians prior to enrollment.

Funding

Support for this study was provided in part by the NIH grants R01HL069438 (L.K. and D.M.), 1U01DE035632 (L.K.), P30 AI124414 (B.C.H.), and the Einstein Medical Scientist Training Program 1T32GM149364 (Z.N.F.) and the Price Family Foundation.

Authors' contributions

Z.N.F. developed the hypotheses, analyzed the data, and prepared all figures. Z.N.F. and L.K. wrote the manuscript text with D.M. and B.C.H. contributing. D.M. and P.S.F. designed the study and secured IRB approval. J.D.L.S., K.I., and J.K. supported study participant enrollment, sample collection, and data aggregation. J.S.K. conducted the neutrophil assay. A.M.M. and A.C. conducted the immune profiling assays.

Author ORCID iDs

Betsy C Herold:  <https://orcid.org/0000-0001-9974-0786>

Libusha Kelly:  <https://orcid.org/0000-0002-7303-1022>

References

- [1] **Agodoa I**, et al. (2018) Societal Costs of Sickle Cell Disease in the United States. *Blood* **132**:4706-4706 <https://doi.org/10.1182/blood-2018-99-119420>
- [2] **Ho ATN**, et al. (2019) Trends in hospitalizations for sickle cell disease related-complications in usa 2004 - 2012. *Journal of Hematology* **8**:11-16 <https://doi.org/10.14740/jh475> | [PubMed](#)
- [3] **Tessema FA**, et al. (2022) High-priced sickle cell gene therapies threaten to exacerbate us health disparities and establish new pricing precedents for molecular medicine. *Journal of Law, Medicine and Ethics* **50**:380-384 <https://doi.org/10.1017/jme.2022.66> | [PubMed](#)
- [4] **Thomson AM**, et al. (2023) Global, regional, and national prevalence and mortality burden of sickle cell disease, 2000-2021: a systematic analysis from the global burden of disease study 2021. *The Lancet Haematology* **10**:e585-e599 [https://doi.org/10.1016/s2352-3026\(23\)00118-7](https://doi.org/10.1016/s2352-3026(23)00118-7) | [PubMed](#)
- [5] **Zhang D**, et al. (2015) Neutrophil ageing is regulated by the microbiome. *Nature* **525**:528-532 <https://doi.org/10.1038/nature15367> | [PubMed](#)

- [6] Dutta D, et al. (2019) Intestinal injury and gut permeability in sickle cell disease. *Journal of Translational Medicine* **17**:183 <https://doi.org/10.1186/s12967-019-1938-8> | PubMed
- [7] Lim SH, et al. (2019) Rifaximin for sickle cell disease. *American Journal of Hematology* **94**:E325-E328 <https://doi.org/10.1002/ajh.25637> | PubMed
- [8] Dutta D, et al. (2019) Effects of rifaximin on circulating aged neutrophils in sickle cell disease. *American Journal of Hematology* **94**:E175-E176 <https://doi.org/10.1002/ajh.25467> | PubMed
- [9] Dutta D, Lim SH (2019) Effects of Rifaximin on Intestinal Pathophysiologic Changes Associated with Sickle Cell Disease (SCD). *Blood* **134**:2282-2282 <https://doi.org/10.1182/blood-2019-127043>
- [10] Lim SH, et al. (2018) Intestinal microbiome analysis revealed dysbiosis in sickle cell disease. *American Journal of Hematology* **93**:E91-E93 <https://doi.org/10.1002/ajh.25019> | PubMed
- [11] Mohandas S, et al. (2020) Differences in gut microbiome in hospitalized immunocompetent vs. immunocompromised children, including those with sickle cell disease. *Frontiers in Pediatrics* **8** <https://doi.org/10.3389/fped.2020.583446> | PubMed
- [12] Brim H, et al. (2021) The gut microbiome in sickle cell disease: Characterization and potential implications. *PLOS One* **16**:1-15 <https://doi.org/10.1371/journal.pone.0255956> | PubMed
- [13] Ley RE, et al. (2006) Human gut microbes associated with obesity. *Nature* **444**:1022-1023 <https://doi.org/10.1038/4441022a> | PubMed
- [14] Delgadinho M, et al. (2022) Microbial gut evaluation in an angolan paediatric population with sickle cell disease. *Journal of Cellular and Molecular Medicine* **26**:5360-5368 <https://doi.org/10.1111/jcmm.17402> | PubMed
- [15] Durazzi F, et al. (2021) Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota. *Scientific Reports* **11**:3030 <https://doi.org/10.1038/s41598-021-82726-y> | PubMed
- [16] Gu Y, et al. (2017) Analyses of gut microbiota and plasma bile acids enable stratification of patients for antidiabetic treatment. *Nature Communications* **8**:1785 <https://doi.org/10.1038/s41467-017-01682-2> | PubMed
- [17] Muscolino P, et al. (2023) Potential predictive role of gut microbiota to immunotherapy in hcc patients: a brief review. *Frontiers in Oncology* **13** <https://doi.org/10.3389/fonc.2023.1247614> | PubMed
- [18] Tisza MJ, Buck CB (2021) A catalog of tens of thousands of viruses from human metagenomes reveals hidden associations with chronic diseases. *Proceedings of the National Academy of Sciences* **118**:e2023202118 <https://doi.org/10.1073/pnas.2023202118> | PubMed
- [19] Clooney AG, et al. (2019) Whole-virome analysis sheds light on viral dark matter in inflammatory bowel disease. *Cell Host & Microbe* **26**:764-778.e5, <https://doi.org/10.1016/j.chom.2019.10.009> | PubMed
- [20] Beghini F, et al. (2021) Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with biobakery 3. *eLife* **10**:e65088 <https://doi.org/10.7554/eLife.65088> | PubMed
- [21] Khan S, Kelly L (2023) Oral microbes are a signature of disease in the gut. *Research Square* <https://doi.org/10.21203/rs.3.rs-1631596/v2>
- [22] Camargo AP, et al. (2023) Identification of mobile genetic elements with genomad. *Nature Biotechnology* **42**:1303-1312 <https://doi.org/10.1038/s41587-023-01953-y> | PubMed
- [23] Howard-Varona C, et al. (2017) Lysogeny in nature: mechanisms, impact and ecology of temperate phages. *The ISME Journal* **11**:1511-1520 <https://doi.org/10.1038/ismej.2017.16> | PubMed
- [24] Kieft K, et al. (2022) Deciphering active prophages from metagenomes. *mSystems* **0**:e00084-22 <https://doi.org/10.1128/msystems.00084-22> | PubMed

- [25] Duvallet C, et al. (2017) Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. *Nature Communications* **8**:1784 <https://doi.org/10.1038/s41467-017-01973-8> | PubMed
- [26] Silva-Junior AL, et al. (2021) Immunological hallmarks of inflammatory status in vaso-occlusive crisis of sickle cell anemia patients. *Frontiers in Immunology* **12**:559925 <https://doi.org/10.3389/fimmu.2021.559925> | PubMed
- [27] Garcia NP, et al. (2020) Sickle Cell Anemia patients display an intricate cellular and serum biomarker network highlighted by TCD4+CD69+ lymphocytes, IL-17/MIP-1 β , IL-12/VEGF, and IL-10/IP-10 axis. *Journal of Immunology Research* **2020**:4585704 <https://doi.org/10.1155/2020/4585704> | PubMed
- [28] Gomes ICP, et al. (2023) Levels of inflammatory markers are differentially expressed in sickle cell anemia and sickle cell trait. *eJHaem* **4**:705-709 <https://doi.org/10.1002/jha2.712> | PubMed
- [29] Stockdale SR, et al. (2023) Interpersonal variability of the human gut virome confounds disease signal detection in ibd. *Communications Biology* **6**:221 <https://doi.org/10.1038/s42003-023-04592-w> | PubMed
- [30] Wang Z, et al. (2015) Non-lethal inhibition of gut microbial trimethylamine production for the treatment of atherosclerosis. *Cell* **163**:1585-1595 <https://doi.org/10.1016/j.cell.2015.11.055> | PubMed
- [31] Vital M, et al. (2014) Revealing the bacterial butyrate synthesis pathways by analyzing (meta)genomic data. *mBio* **5** <https://doi.org/10.1128/mbio.00889-14>
- [32] Zheng L, et al. (2017) Microbial-derived butyrate promotes epithelial barrier function through il-10 receptor-dependent repression of claudin-2. *The Journal of Immunology* **199**:2976-2984 <https://doi.org/10.4049/jimmunol.1700105> | PubMed
- [33] Holdeman LV, Moore WEC (1974) New genus, *Coproccoccus*, twelve new species, and emended descriptions of four previously described species of bacteria from human feces. *International Journal of Systematic and Evolutionary Microbiology* **24**:260-277 <https://doi.org/10.1099/00207713-24-2-260>
- [34] Zong W, et al. (2024) Disruption of intestinal oxygen balance in acute colitis alters the gut microbiome. *Gut Microbes* **16**:2361493 <https://doi.org/10.1080/19490976.2024.2361493> | PubMed
- [35] Ramakrishnan SK, Shah YM (2016) Role of intestinal hif-2 α in health and disease. *Annual Review of Physiology* **78**:301-325 <https://doi.org/10.1146/annurev-physiol-021115-105202> | PubMed
- [36] Krakau S, et al. (2022) nf-core/mag: a best-practice pipeline for metagenome hybrid assembly and binning. *NAR Genomics and Bioinformatics* **4**:qac007 <https://doi.org/10.1093/nargab/lqac007> | PubMed
- [37] Krakau S, et al. (2023) nf-core/mag. version: 2.5.1
- [38] Li D, et al. (2015) MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**:1674-1676 <https://doi.org/10.1093/bioinformatics/btv033> | PubMed
- [39] Hyatt D, et al. (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**:119 <https://doi.org/10.1186/1471-2105-11-119> | PubMed
- [40] Kieft K, et al. (2020) VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome* **8**:90 <https://doi.org/10.1186/s40168-020-00867-0> | PubMed
- [41] Mallick H, et al. (2021) Multivariable association discovery in population-scale meta-omics studies. *PLOS Computational Biology* **17**:1-27 <https://doi.org/10.1371/journal.pcbi.1009442> | PubMed
- [42] Steinegger M, Söding J (2017) Mmseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology* **35**:1026-1028 <https://doi.org/10.1038/nbt.3988> | PubMed
- [43] Hagberg AA, et al. (2008) Exploring network structure, dynamics, and function using networkx. In: Proceedings of the 7th Python in Science Conference. pp. 11-15 <https://doi.org/10.25080/tcww9851>

- [44] Shannon P, et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* **13**:2498-2504 <https://doi.org/10.1101/gr.1239303> | PubMed
 - [45] Flamholz ZN, et al. (2024) Large language models improve annotation of prokaryotic viral proteins. *Nature Microbiology* **9**:537-549 <https://doi.org/10.1038/s41564-023-01584-8> | PubMed
 - [46] Virtanen P, et al. (2020) SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **17**:261-272 <https://doi.org/10.17863/cam.73648>
 - [47] Waskom ML (2021) seaborn: statistical data visualization. *Journal of Open Source Software* **6**:3021 <https://doi.org/10.21105/joss.03021>
 - [48] Charlier F, et al. (2022) Statannotations.
 - [49] T. pandas development team (2020) pandas-dev/pandas: Pandas.
 - [50] Harris CR, et al. (2020) Array programming with NumPy. *Nature* **585**:357-362 <https://doi.org/10.1038/s41586-020-2649-2> | PubMed
- Flamholz (2025) Sick cell disease patient gut microbiome study. NCBI BioProject. ID PRJNA1320713 <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA1320713>

Peer reviews

Reviewer #1 (Public review):

Summary:

In this manuscript, Flamholz and colleagues use metagenomic sequencing to profile the microbiome of individuals with sickle cell disease (SCD), the most common genetic blood disorder in the world. To build on previous studies that found dysbiosis in SCD, this manuscript aims to examine whether changes in either bacterial species or bacteriophages correlate with inflammatory hallmarks of the disease. The authors claim that sickle cell dysbiosis does not correlate with inflammatory hallmarks of the disease, but instead, aged neutrophil numbers and bacteriophages do. Appropriate control subjects and additional analyses are needed to support that conclusion.

Strengths:

The primary strength of this paper is the investigation into disease-associated changes in bacteriophages. This is an entirely novel idea in the sickle cell field, and based on the current results, may be an important, under-recognized disease hallmark. It is unclear, however, if phages are "the chicken or the egg" in terms of sickle cell inflammatory profiles; do these increases in phage number simply result from other disease processes, or are they in any way contributing to disease pathophysiology?

Weaknesses:

A primary weakness of the manuscript is the fact that the majority of individuals included in the control group maintain sickle cell trait (HbAS genotype). Although typically asymptomatic, it is unclear if this genotype is associated with microbial changes that would not be observed in a true control group (HbAA genotype). This is a significant limitation that may limit the ability to draw conclusions from the current data set.

Another key weakness is the lack of beta diversity assessment. Although decreased alpha diversity is observed in individuals with SCD, and specific bacterial taxa are differentially abundant following multivariate analyses, there is no overall comparison of bacterial community composition between individuals with SCD and controls. Prior to drawing conclusions about the relationship (or lack thereof) between the SCD microbiome and inflammatory markers, it is important to know if this study did indeed find disease-associated changes in microbiome composition.

It is unclear which individuals were used for aged neutrophil (AN) and molecular data assessments. For example, were children who were still receiving penicillin prophylaxis included in these specific assessments? Given the authors' previous work demonstrating that antibiotic treatment decreases AN pathology, it seems critical to limit all AN/molecular analyses to older subjects who are not on daily penicillin treatment (if possible).

A minor weakness is the continued use of "disease" vs. "healthy" indicators as primary microbiome metrics that are used for molecular correlations. The lack of metric specificity - and lack of discussion regarding which diseases were used to generate these indicators (how similar/different are they to sickle cell?) - could be said to make these metrics essentially meaningless.

<https://doi.org/10.7554/eLife.111630.1.sa3>

Reviewer #2 (Public review):

Summary:

The study analyzes stool metagenomes from 98 SCD patients and 46 controls, with SCD and control groups matched on age, race, sex, and ethnicity. The authors report lower Shannon diversity, lower Firmicutes/Bacteroidetes ratio, loss of health-associated taxa, increased disease-associated indicators, altered butyrate/fatty-acid metabolism pathways, and enrichment of provirus/prophage fractions in SCD. They further correlate aged-like neutrophils and prophage fractions with inflammatory cytokines. The main strength is that this is not just another 16S comparison. The use of whole-community metagenomics, immune profiling, neutrophil assays, and clinical metadata makes the study more biologically interesting than prior small SCD microbiome papers. The main weakness is that the causal and mechanistic interpretation is too strong. The data support an association between SCD status and microbiome/virome features, but they do not yet establish a clear "axis of pathophysiology." The provirus findings are intriguing, but require stronger statistical control, better validation, and more cautious interpretation.

Strengths:

The major strengths of the study include the clinically relevant disease setting, the use of whole-community sequencing, the integration of microbial, immune-cell, cytokine, and clinical measurements, and the novel attention to bacterial virus-related features. A particularly interesting aspect of the work is the analysis of virus-like elements integrated into bacterial genomes. The authors report that these elements are enriched in the gut microbial communities of patients with sickle cell disease and are associated with several inflammatory signals in blood. This observation is potentially important because it suggests that the microbial contribution to inflammation in sickle cell disease may involve not only bacteria but also bacterial virus-related genetic elements.

Weaknesses:

The evidence for this proposed immune-related mechanism is incomplete. The study is cross-sectional and largely based on associations, so it cannot determine whether these virus-like

elements drive immune activation, reflect immune activation, or are linked indirectly through disease severity, treatment history, or other clinical factors. The main limitations are the single-center design, modest sample size for some immune measurements, limited ability to control for treatment and disease heterogeneity, and the need for clearer multiple-testing correction in the correlation analyses. In particular, stronger adjustment for available clinical factors such as hydroxyurea use, transfusion history, pain admissions, genotype, and other markers of disease burden would help readers judge how specific the microbial and viral findings are to sickle cell disease itself.

Overall, the authors largely achieve their descriptive aim of identifying gut microbial differences associated with sickle cell disease. The evidence is solid for the presence of broad microbial community differences, but incomplete for the stronger conclusion that virus-like elements form a pathophysiological immune axis. The work will likely be useful to researchers studying the microbiome, inflammation, and sickle cell disease, especially as a hypothesis-generating dataset. Its impact would be strengthened by more cautious interpretation, stronger control of clinical confounders, clearer statistical correction, and future longitudinal or experimental studies to test causality.

<https://doi.org/10.7554/eLife.111630.1.sa2>

Reviewer #3 (Public review):

Summary:

In this manuscript, Flamholz et al. sought to determine whether consistent and significant interactions exist between the gut microbiome and disease pathology in sickle cell disease (SCD). By sequencing and analysing metagenomes from faecal samples collected from 98 SCD patients and 46 control subjects, they identified community-level shifts in both the bacterial and proviral gut microbiome of SCD patients. They further reported correlations between the proviral microbiome and multiple blood cytokines, whereas similar associations were not observed for the bacterial microbiome. Based on these findings, the authors propose the existence of a viral-immune axis in SCD pathophysiology and targetable functional alterations in the gut microbiome.

Strengths:

This work includes the largest SCD cohort analysed to date, enabling analysis with relatively strong statistical power. In addition to profiling the bacterial microbiome, the study also examines the gut proviral microbiome, thereby providing a more comprehensive investigation of the topic. The newly generated metagenomic dataset will also be valuable for further meta-analysis by the wider community. Overall, the authors have largely achieved their aims.

Weaknesses:

However, this study represents a single-centre cross-sectional investigation, and most findings remain correlative in nature. In particular, the claim that the study identifies targetable functional alterations in the gut microbiome for disease treatment may be somewhat overstated. Although the reported functional module changes in SCD patients are intriguing, additional mechanistic and/or longitudinal evidence would be required before these features can realistically be considered targetable.

<https://doi.org/10.7554/eLife.111630.1.sa1>

Author response:

eLife Assessment:

This study presents a valuable metagenomic analysis of the gut microbiome in sickle cell disease (SCD) patients, revealing associations between bacteriophage, host immunity, and SCD pathophysiology. While these data are interesting and helpful for hypothesis generation, they are deemed incomplete; additional experiments would be needed to test causality and to provide mechanistic insight. Despite these limitations, this work will be of broad interest to researchers studying SCD, immunology, phage biology, and the microbiome, adding to the small but growing literature suggesting a microbial component to SCD.

The authors would like to thank the reviewers for thorough and constructive comments on our manuscript. We will make major updates to the manuscript addressing the following points and suggestions from the three reviewers: (1) assessing HbAS/AA genotype influence on microbiome composition; (2) conducting the requested beta diversity analysis, (3) conducting the requested sensitivity analysis to assess the impact of disease severity and therapy on microbiome and virome features; (4) modifying our language to clearly state that our results do not indicate causality or mechanism of microbiome interactions with sickle cell disease pathophysiology; (5) improved discussion of the phage results and their strengths and limitations; (6) additional changes throughout for clarity and correction of errors. We will change the title to “Bacterial and viral gut microbiome alterations characterize microbiome-immune-pathophysiology axes in Sickle Cell Disease.” These additions will greatly improve our work and presentation and we are grateful to the reviewers and our editors.

We have indicated where specific changes were made in response to the public reviews below.

Public Reviews:

Reviewer #1 (Public review):

Summary:

In this manuscript, Flamholz and colleagues use metagenomic sequencing to profile the microbiome of individuals with sickle cell disease (SCD), the most common genetic blood disorder in the world. To build on previous studies that found dysbiosis in SCD, this manuscript aims to examine whether changes in either bacterial species or bacteriophages correlate with inflammatory hallmarks of the disease. The authors claim that sickle cell dysbiosis does not correlate with inflammatory hallmarks of the disease, but instead, aged neutrophil numbers and bacteriophages do. Appropriate control subjects and additional analyses are needed to support that conclusion.

Strengths:

The primary strength of this paper is the investigation into disease-associated changes in bacteriophages. This is an entirely novel idea in the sickle cell field, and based on the current results, may be an important, under-recognized disease hallmark. It is unclear, however, if phages are "the chicken or the egg" in terms of sickle cell inflammatory profiles; do these increases in phage number simply result from other disease processes, or are they in any way contributing to disease pathophysiology?

Weaknesses:

A primary weakness of the manuscript is the fact that the majority of individuals included in the control group maintain sickle cell trait (HbAS genotype). Although typically

asymptomatic, it is unclear if this genotype is associated with microbial changes that would not be observed in a true control group (HbAA genotype). This is a significant limitation that may limit the ability to draw conclusions from the current data set.

Another key weakness is the lack of beta diversity assessment. Although decreased alpha diversity is observed in individuals with SCD, and specific bacterial taxa are differentially abundant following multivariate analyses, there is no overall comparison of bacterial community composition between individuals with SCD and controls. Prior to drawing conclusions about the relationship (or lack thereof) between the SCD microbiome and inflammatory markers, it is important to know if this study did indeed find disease-associated changes in microbiome composition.

It is unclear which individuals were used for aged neutrophil (AN) and molecular data assessments. For example, were children who were still receiving penicillin prophylaxis included in these specific assessments? Given the authors' previous work demonstrating that antibiotic treatment decreases AN pathology, it seems critical to limit all AN/molecular analyses to older subjects who are not on daily penicillin treatment (if possible).

A minor weakness is the continued use of "disease" vs. "healthy" indicators as primary microbiome metrics that are used for molecular correlations. The lack of metric specificity - and lack of discussion regarding which diseases were used to generate these indicators (how similar/different are they to sickle cell?) - could be said to make these metrics essentially meaningless.

We thank the reviewer for their helpful comments and suggestions. We want to first note that patients on prophylactic penicillin within six months of sample collection were excluded from the study due to the known impact of antibiotics on gut microbiomes, this will be clarified in the main text. We will include an analysis evaluating the influence of control genotype (HbAA/HbAS) on our microbiome and virome results. To evaluate whether control genotype influenced major microbiome and virome features, analyses were restricted to control participants only. Controls were stratified by genotype as HbAA or HbAS. Four significant microbiome and virome features were tested: F:B ratio, Shannon diversity, provirus fraction, and virus count. HbAA and HbAS controls were compared using two-sided Mann-Whitney U tests. Benjamini-Hochberg FDR correction was applied across the four tested features. HbAS and HbAA controls did not differ significantly for F:B ratio, Shannon diversity, provirus fraction, or virus count. The inclusion of HbAA/AS will strengthen our results with respect to the observation that sickle cell disease patient microbiomes remain significantly different from sickle trait (HbAS) controls. These results will be reported in a new Supplemental Table.

We will include a beta diversity analysis using MetaPhlAn species profiles. Beta diversity analyses were performed in Python using pandas and NumPy for data processing, scikit-bio for distance calculations and PERMANOVA, scikit-learn for ordination-related computations, statsmodels for multiple-testing correction where applicable, and matplotlib for visualization.

For the primary disease/control comparison, samples were grouped as control or SCD. For the genotype control sensitivity analysis, samples were restricted to HbAA and HbAS individuals as described above. Species detected in at least 10% of included samples were retained for beta diversity analysis. To account for the compositional structure of metagenomic relative abundance data, species profiles were transformed using a centered log-ratio transformation after addition of a small pseudocount to accommodate zero values. Aitchison distances were calculated from the CLR-transformed species profiles. Statistical significance of group separation was assessed by PERMANOVA using 999 permutations. For the control versus SCD comparison, PERMANOVA was performed between the two disease-status groups. For the HbAA versus HbAS control comparison, PERMANOVA was performed among controls only.

In the SCD cohort, beta diversity differed significantly between controls and SCD participants by Aitchison distance after CLR transformation ($R^2 = 0.030$, $p = 0.001$). In contrast, HbAA and HbAS controls did not differ significantly in beta diversity ($R^2 = 0.024$, $p = 0.282$), supporting the conclusion that the observed SCD/control separation was not driven by control genotype composition. These methods and results will be reported in the revised manuscript.

The manuscript describing the microbiome health and disease indicators was submitted to eLife jointly with this manuscript as a package; eLife declined to review the indicator manuscript. Briefly, this study conducted a cross-disease meta-analysis of 38 studies comprising 8,204 samples and identified 100 bacterial taxa or “indicators” that are weakly but consistently associated with health or disease across diverse conditions, including, but not limited to, inflammatory bowel disease, colorectal cancer, type 2 diabetes. The indicator taxa were validated in an independent cohort of Graves’ disease patients. We currently cite an older version of this work posted as a preprint. The manuscript is currently under review at another journal and we will update this manuscript with the updated citation when it is available.

We will address the other recommendations from this reviewer in the revision.

Reviewer #2 (Public review):

Summary:

The study analyzes stool metagenomes from 98 SCD patients and 46 controls, with SCD and control groups matched on age, race, sex, and ethnicity. The authors report lower Shannon diversity, lower Firmicutes/Bacteroidetes ratio, loss of health-associated taxa, increased disease-associated indicators, altered butyrate/fatty-acid metabolism pathways, and enrichment of provirus/prophage fractions in SCD. They further correlate aged-like neutrophils and prophage fractions with inflammatory cytokines. The strength is that this is not just another 16S comparison. The use of whole-community metagenomics, immune profiling, neutrophil assays, and clinical metadata makes the study more biologically interesting than prior small SCD microbiome papers. The main weakness is that the causal and mechanistic interpretation is too strong. The data support an association between SCD status and microbiome/virome features, but they do not yet establish a clear “axis of pathophysiology.” The provirus findings are intriguing, but require stronger statistical control, better validation, and more cautious interpretation.

Strengths:

The major strengths of the study include the clinically relevant disease setting, the use of whole-community sequencing, the integration of microbial, immune-cell, cytokine, and clinical measurements, and the novel attention to bacterial virus-related features. A particularly interesting aspect of the work is the analysis of virus-like elements integrated into bacterial genomes. The authors report that these elements are enriched in the gut microbial communities of patients with sickle cell disease and are associated with several inflammatory signals in blood. This observation is potentially important because it suggests that the microbial contribution to inflammation in sickle cell disease may involve not only bacteria but also bacterial virus-related genetic elements.

Weaknesses

The evidence for this proposed immune-related mechanism is incomplete. The study is cross-sectional and largely based on associations, so it cannot determine whether these virus-like elements drive immune activation, reflect immune activation, or are linked indirectly through disease severity, treatment history, or other clinical factors. The main

limitations are the single-center design, modest sample size for some immune measurements, limited ability to control for treatment and disease heterogeneity, and the need for clearer multiple-testing correction in the correlation analyses. In particular, stronger adjustment for available clinical factors such as hydroxyurea use, transfusion history, pain admissions, genotype, and other markers of disease burden would help readers judge how specific the microbial and viral findings are to sickle cell disease itself.

Overall, the authors largely achieve their descriptive aim of identifying gut microbial differences associated with sickle cell disease. The evidence is solid for the presence of broad microbial community differences, but incomplete for the stronger conclusion that virus-like elements form a pathophysiological immune axis. The work will likely be useful to researchers studying the microbiome, inflammation, and sickle cell disease, especially as a hypothesis-generating dataset. Its impact would be strengthened by more cautious interpretation, stronger control of clinical confounders, clearer statistical correction, and future longitudinal or experimental studies to test causality.

We thank the reviewer for their helpful comments and suggestions. We want to first note that patients on prophylactic penicillin within six months of sample collection were excluded from the study due to the known impact of antibiotics on gut microbiomes, this will be clarified in the main text. We will temper our interpretation of our results, making clear that we are not arguing that either prophages or bacteria are causal or mechanistically associated with SCD biology and pathology. We will strengthen our control of clinical confounders, and add clearer statistical correction in the revision. We look forward to conducting future studies to test causality and understand mechanism.

We will address the other recommendations from this reviewer in the revision.

Reviewer #3 (Public review):

Summary:

In this manuscript, Flamholz et al. sought to determine whether consistent and significant interactions exist between the gut microbiome and disease pathology in sickle cell disease (SCD). By sequencing and analysing metagenomes from faecal samples collected from 98 SCD patients and 46 control subjects, they identified community-level shifts in both the bacterial and proviral gut microbiome of SCD patients. They further reported correlations between the proviral microbiome and multiple blood cytokines, whereas similar associations were not observed for the bacterial microbiome. Based on these findings, the authors propose the existence of a viral-immune axis in SCD pathophysiology and targetable functional alterations in the gut microbiome.

Strengths:

This work includes the largest SCD cohort analysed to date, enabling analysis with relatively strong statistical power. In addition to profiling the bacterial microbiome, the study also examines the gut proviral microbiome, thereby providing a more comprehensive investigation of the topic. The newly generated metagenomic dataset will also be valuable for further meta-analysis by the wider community. Overall, the authors have largely achieved their aims.

Weaknesses:

However, this study represents a single-centre cross-sectional investigation, and most findings remain correlative in nature. In particular, the claim that the study identifies targetable functional alterations in the gut microbiome for disease treatment may be somewhat overstated. Although the reported functional module changes in SCD patients

are intriguing, additional mechanistic and/or longitudinal evidence would be required before these features can realistically be considered targetable.

We thank the reviewer for their helpful comments and suggestions. We will note in the text that additional mechanistic and longitudinal studies are required before we can target the microbiome and virome in SCD and clarified that this is a single center, cross-sectional. We will make further modifications to the manuscript to clarify cohort features (specifically, age and race were matched, other baseline characteristics were balanced), to properly describe the Shannon diversity metric, and to fix several errors that this reviewer caught.

<https://doi.org/10.7554/eLife.111630.1.sa0>