ONLINE REPOSITORY TEXT

Intestinal microbiome and metabolome signatures in patients with chronic granulomatous disease.

Chandrasekaran, P., Han, Y., Zerbe, C.S., … Falcone, E.L.

METHODS

Fecal DNA extraction and amplicon sequencing

Fecal DNA was extracted using PowerSoil DNA Isolation Kit (MO BIO Laboratories), with Proteinase K addition (Qiagen) for 10 minutes at 65C° prior to bead-beating using TissueLyser II (Qiagen). The 16S rRNA gene (V4 region) sequencing was performed using primers (515F = GTGCCAGCAGCCGCGGTAA and 806R = GGACTACCAGGGTATCTAAT) that were modified to include a linker sequence, a 12 bp index sequence and a heterogeneity spacer [1]. PCR was set up with LA PCR™ Kit (Takara Bio USA). Libraries were cleaned up (Agencourt AMPure XP Kit; Beckman Coulter), quantified (Quant-IT dsDNA High-Sensitivity Assay Kit; Invitrogen) and sequenced using a 250-bp paired-end sequencing protocol on the MiSeq platform (Illumina Inc.) at the NIH Intramural Sequencing Center.

Fecal metabolomics

Liquid chromatography-mass spectrometry with the hybrid metabolomics method

The LC column was a Millipore ZIC-pHILIC (2.1 x150 mm, 5 μm) coupled to a Dionex Ultimate 3000 with 25^{o}C set for the gradient elution. MS analyses were carried out by coupling the LC system to a Thermo Q Exactive HF mass spectrometer operating in heated electrospray ionization mode. The parameters included 30 min duration, 3.5kV spray voltage, 320^{o}C capillary temperature, 35 sheath gas rate, 10 aux gas, 100 μA of max spray current, 120,000 resolution with an automatic gain control target of 3e6, a maximum IT of 100 ms, and the scan range from 67-1000 m/z. Tandem MS spectra used 15,000 resolution, AGC target of 1e5, maximum IT of 50 ms, isolation window of 0.4 m/z, isolation offset of 0.1 m/z, fixed first mass of 50 m/z, and 3-way multiplexed normalized collision energies (nCE) of 10, 35, 80. The minimum AGC target was 1e4 with an intensity threshold of 2e5. All data were acquired in profile mode.

Global metabolomics data processing and relative quantification of metabolites

Using an in-house python script, Thermo RAW files were converted to SQLite format to enable downstream peak detection and quantification, and MS/MS spectra were searched against the NIST17MS/MS [2, 3], METLIN [4] and respective Decoy spectral library databases [5, 6] applying a 100% FDR (False Discovery Rate). The peak heights for each sample and metabolite were extracted from the sqlite3 files based on the metabolite retention time ranges and accurate masses. Metabolite peaks were extracted based on the theoretical m/z of the expected ion type e.g., [M+H] +, with a 15-ppm tolerance ± 0.2 min peak apex retention time tolerance within an initial retention time search window of ± 0.25 min. The final peak detection was calculated based on a signal to noise ratio of 3X compared to blank controls with a floor of 10,000 (arbitrary units). To account for inter-batch variations, the median metabolite intensity of each sample was used to normalize the intensities in that sample. The data was processed with an in-house pipeline using Metabolyze version 1.0. and t-tests were performed using Python SciPy (1.1.0) [7, 8] library. The R-library DESeq2 (1.24.0) was used [9] to test for significant differences.

REFERENCES

1. Fadrosh DW, Ma B, Gajer P, Sengamalay N, Ott S, Brotman RM, et al. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. Microbiome 2014; 2:6.

2. Simon-Manso Y, Lowenthal MS, Kilpatrick LE, Sampson ML, Telu KH, Rudnick PA, et al. Metabolite profiling of a NIST Standard Reference Material for human plasma (SRM 1950): GC-MS, LC-MS, NMR, and clinical laboratory analyses, libraries, and web-based resources. Anal Chem 2013; 85:11725-31.

3. Simon-Manso Y, Marupaka R, Yan X, Liang Y, Telu KH, Mirokhin Y, et al. Mass Spectrometry Fingerprints of Small-Molecule Metabolites in Biofluids: Building a Spectral Library of Recurrent Spectra for Urine Analysis. Anal Chem 2019; 91:12021-9.

4. Smith CA, O'Maille G, Want EJ, Qin C, Trauger SA, Brandon TR, et al. METLIN: a metabolite mass spectral database. Ther Drug Monit 2005; 27:747-51.

5. Wang X, Jones DR, Shaw TI, Cho JH, Wang Y, Tan H, et al. Target-Decoy-Based False Discovery Rate Estimation for Large-Scale Metabolite Identification. J Proteome Res 2018; 17:2328-34.

6. Wang X, Cho JH, Poudel S, Li Y, Jones DR, Shaw TI, et al. JUMPm: A Tool for Large-Scale Identification of Metabolites in Untargeted Metabolomics. Metabolites 2020; 10.

7. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 2020; 17:261-72.

8. Jones EO, Travis; Peterson, Pearu. SciPy: Open Source Scientific Tools for Python. 2001.

9. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014; 15:550.

SUPPLEMENTAL FIGURE LEGENDS

Figure E1. Correlations between clinical parameters associated with inflammatory bowel disease in patients with chronic granulomatous disease. Correlogram for clinical parameters in patients with CGD. Square values =c oefficient of correlation (r value); Square size = strength of significance (red = positive correlation, blue = negative correlation, blank = no significant correlation). All presented r values had $p<0.05$.

Figure E2. Impact of CGD genotype on the composition of the intestinal microbiome. (A) Heatmap of the core microbiome at the genus level that are detected in high fraction in different CGD genotypes are shown (20% of the sample prevalence cut-off; blue = low prevalence, red = high prevalence). **(B)** Random forest analysis. Genera and species with the highest discriminatory power between different CGD genotypes are listed. The squares next to the plot indicate the abundance in each group; red = high abundance, blue = low abundance. **(C)** Heat tree depicting differential abundance of the genera in the different CGD genotypes compared to the Healthy group (red = high abundance, blue = low abundance; GP91 n= 49, P22 n=5, P47 n=23, P67 n=2).

Figure E3. Factor analysis with mixed data (FAMD) approach showing how much each metadata variable accounts for the variation captured by a given principal dimension (PD). (A) Top 20 metadata variables contributing to the variability in PD1 and PD2 in the NIH CC cohort and **(B)** top 20 amplicon sequencing variants contributing at the genus level to the variability observed in the dataset. **(C)** Linear discriminant analysis (LDA) score determined by the LEfSe analysis showing biomarkers at genus level for the factors that showed significant differences at the alpha diversity level in the CGD group.

Figure E4. Comparison of microbiome signatures between patients with CGD from NIH and PIDTC cohorts regardless of IBD status or antimicrobial use. (A) Alpha diversity analyses at genus level comparing CGD-only group of NIH CC cohort (n=79) to the PIDTC cohort (n=36). **(B)** PCoA plot of beta diversity based on weighted Unifrac distances for the two comparison groups with p values determined by PERMANOVA ($p<0.004$). **(C)** Linear discriminant analysis (LDA) score determined by the LEfSe analysis showing the most differential genera between the two cohorts. **(D)** Correlation at the genus level of the two cohorts is depicted as the network. Two taxa are connected by an edge when $p<0.05$ and the correlation threshold >0.3 (green = NIH CC cohort and red = PIDTC cohort).

Figure E5. Effect of Age on the microbiome signatures in subjects with no active IBD or a history of IBD. (A) Alpha diversity analyses comparing subjects with ≤12 years of age (n=26) to those with over 12 years of age (n=12). **(B)** PCoA plot of beta diversity based on weighted Unifrac distances for the two comparison groups with p values determined by Permanova. **(C)** Pie Charts showing the distribution of the phyla in both comparison groups. **(D)** Linear discriminant analysis (LDA) score determined by the LEfSe analysis showing the most differential genera between the two cohorts.

Figure E6. History of IBD and active IBD in patients with CGD have both shared and distinct microbiota. Linear discriminant analysis (LDA) score determined by the LEfSe analysis showing biomarkers at the genus level for **(A)** the CGD group with active IBD and **(B)** the CGD group with a history of IBD. **(C)** Relative abundance at the phylum level in the CGD group with vs. without a history of IBD, stratified by the presence or absence of active IBD. **(D)** Venn diagram generated post-EdgeR analysis identifying the differentially abundant microbiota. The Venn diagram shows the common and distinct genera and species between the comparison groups, with the names listed on the right side.

Table E1. Demographic and clinical characteristics of participants from the PIDTC cohort compared to the NIH CC cohort.

	PIDTC cohort CGD, n=36	**NIH CC cohort** CGD, n=79
Age, median (IQR), years	2.1 (5.5)	23 (20.5)
Female, n (%)	5 (13.9)	17 (21.5)
CGD genotype [affected protein], n (%) *CYBB*$^{0/-}$ [gp91$_{phox}$] *CYBA*$^{-/-}$ [p22$_{phox}$] *NCF1*$^{-/-}$ [p47$_{phox}$] *NCF2*$^{-/-}$ [p67$_{phox}$] *NCF4*$^{-/-}$ [p40$_{phox}$] Unknown	 27 (75) 1 (2.7) 3 (8.3) 3 (8.3) 0 (0) 2 (5.5)	 48 (60.8) 5 (6.3) 23 (29.1) 2 (2.5) 0 (0) 0 (0)
History of CGD-IBD, n (%)	7 (19.4)	54 (68.4)
Active GI symptoms at visit [a], n (%)	0 (0)	35 (44.3)
On prophylactic antibiotics, n (%) TMP-SMX Azole Unknown	 31 (86.1) 30 (83.3) 4 (11.1)	 64 (81) 71 (89.9) 8 (10)

On non-prophylactic antibiotics, n (%) Any Unknown	6 (16.7) 4 (11.1)	29 (36.7) 0 (0)
On steroids, n (%) Any Unknown On another immune modulator, n (%) Any Unknown	9 (25) 4 (11.1) 8 (22.2) 4 (11.1)	17 (21.5) 0 (0) 29 (36.7) 0 (0)

PIDTC: Primary Immune Deficiency Treatment Consortium, CGD: chronic granulomatous disease, NIH CC: National Institutes of Health Clinical Center, IQR: interquartile range, NA: not applicable, TMPSMX: trimethoprim-sulfamethoxazole.

[a] Active GI symptoms: watery bowel movements (BM) or more than 2 BMs per day or blood/mucus in stool or active fistulizing or perianal disease.

Table E2. Comparison of clinical parameters associated with a history of CGD-associated inflammatory bowel disease in participants from the NIH CC cohort.

	Normal range	**Healthy** [a] n=17	**CGD no IBD** n=25	**CGD-IBD** n=54	***P* value**
Total WBC count (10^9/L; mean (SD)) Out of normal range, n (%)	Adult: 4.23-9.07 Pediatric [f] 9-12: 4.5-13.5 Pediatric 13-18: 4.5-13.0	6.4 (2.1) 1 (5.9)	6.1 (1.8) 2 (8)	6.9 (4.2) 13 (24.1)	0.43 0.09
Hg (g/dL; mean (SD)) Out of normal range, n (%)	Adult men: 13.7-17.5 Adult women: 12.0-16.0 Pediatric 6-12: 11.5-15.5 Pediatric 13-18: 13.0-16.0	13.2 (1.2) 1 (5.9)	12.6 (2.3) 7 (28)	12.3 (1.9) 25 (46.3)	0.08 0.12
Hct (%; mean (SD)) Out of normal range, n (%)	Adult men: 41-50% Adult women: 36-44% Pediatric 6-12: 35-45 Pediatric 13-18: 36-51	40 (3) 1 (5.9)	38 (6) 13 (52)	37.7 (5.1) 18 (33.3)	0.07 0.11
Platelet count (10^9/L; mean (SD)) Out of normal range, n (%)	All ages: 161-347	255.5 (68.2) 1 (5.9)	237.8 (87.6) 3 (12)	285.9 (87) 1 (1.9)	0.07 0.06

CRP (mg/dL; mean (SD)) ESR (mm/hr; mean (SD)) Albumin (g/dL; mean (SD))	0-4.99 mg/dL 0-25 mm/hr 3.5-5.2 g/dL	NA NA NA	36.4 (56.5) 26 (28.2) 3.9 (0.6)	22.3 (31.1) 27.7 (21.5) 3.9 (0.6)	0.25 0.79 >0.99
Fecal occult blood positive, n (%) Fecal calprotectin >50 mcg/g, n (%)	NA <50 mcg/g	0 (0) 0 (0)	12 (48) 2 (8)	22 (40.7) 17 (31.5)	0.003 0.004
Clinical activity index (CAI) [b] >1, n (%)	NA	0 (0)	1 (4)	31 (57.4)	<0.0001
Numeric rating scale, mean (SD) ≥8, n (%) [c]	NA	8.9 (1.2) 13 (76.5)	7.9 (1.3) 14 (56)	7 (1.9) 17 (31.5)	<0.0001 0.003
SIBDQ, mean (SD) ≥56, n (%) [d]	NA	62.6 (5.9) 12 (70.6)	58 (9.4) 15 (60)	57.1 (9.4) 26 (48.1)	0.02 0.23
P-SCCAI, mean (SD) ≤3, n (%) [e]	NA	1.4 (1.2) 13 (76.5)	3 (2.9) 15 (60)	4.6 (3.5) 19 (35.2)	<0.0001 0.005

CGD: chronic granulomatous disease, IBD: inflammatory bowel disease, NIH CC: National Institutes of Health Clinical Center, SD: standard deviation, NA: not available, NS: not significant, WBC: white blood cell, Hg: hemoglobin, Hct: hematocrit, platelet, CRP: C-reactive protein, erythrocyte sedimentation rate, albumin, SIBDQ: Short Quality of Life Inflammatory Bowel Disease Questionnaire, P-SCCAI: Patientmodified Simple Clinical Colitis Activity Index.

[a] Excludes Healthy PPX patients (healthy participants on trimethoprim-sulfamethoxazole prophylaxis for recurrent urinary tract infection) because clinical laboratory data was not available for these patients.

[b] CAI (clinical activity index): CAI 1= 0-2 bowel movements (BM) per day, CAI 2= 2-4 BMs per day, CAI 3= >4 BMs per day and/or presence of blood or mucus in stool and/or fistulae and/or perianal disease

[c] Scores equal or greater than the indicated cutoff were associated good quality of life and/or inactive disease in Surti, B., et al., 2013.

[d] Scores equal or greater than the indicated cutoff were associated good quality of life and/or inactive disease in Irvine, E.J., et al., 1996.

[e] Scores equal or lesser than the indicated cutoff were associated good quality of life and/or inactive disease in Bennebroek, E., et al., 2013.

[f] Pediatric normal ranges apply to children aged 6-12 years and 13-18 years (for males) as the cohort only included 1 patient under 6 years of age, and the patients aged 13-18 years were all males.

Table E3. Significantly altered stool metabolites in patients with CGD compared to healthy individuals.

Metabolite	***p*-value**	**Log 2-fold change**
L-Lysine	<0.0001	5.71
Thiocyanic acid	<0.0001	7.90
Butylphosphonic acid	<0.0001	-4.40
Beta-chloro-D-alanine	<0.0001	-2.73
DL-Ornithine	<0.0001	2.37
Propionic acid	<0.0001	6.90

Table E4. Significantly altered stool metabolites in patients with CGD with active IBD vs. without IBD ever.

Metabolite	***p*-value**	**Log 2-fold change**
N-Acetyl-5-aminosalicylic acid	0.0002	2.44
Myclobutanil	0.0003	2.26
LY294002	0.0003	-1.49
Cytidine	0.0004	-2.65
Methyldopa	0.0005	2.01
Piscidic acid	0.0006	-1.71
N6-Carbamoyl-DL-lysine	0.0007	-1.66
N-Acetyl-D-galactosaminitol	0.0007	2.90
4-Pyrimidine methanamine	0.001	1.88
L-Histidine	0.001	4.54
D-Saccharic acid	0.001	4.81