Online Data Supplement

Title: Gut microbiota predict pulmonary infiltrates after allogeneic hematopoietic cell transplantation

Author list: Bianca Harris, Sejal M. Morjaria, Eric R. Littmann, Alexander I. Geyer, Diane E. Stover, Juliet N. Barker, Sergio A. Giralt, Ying Taur, and Eric G. Pamer

Methods

Defining the primary clinical endpoint

The primary endpoint, PC, employed a broad clinical-radiographic definition. It required a clinical indication for chest imaging (respiratory symptoms, e.g. cough,dyspnea, and/or abnormal vital signs, e.g. fever, hypoxia) and the development of a radiographic parenchymal abnormality that correlated clinically and influenced management. The operational procedure involved independent chart review by two study physicians (one pulmonary and one infectious diseases specialist) on each of the 94 subjects. First, this entailed manual extraction and review of radiographic findings, as described by the clinical radiologist in the electronic medical record (EMR), from all diagnostic chest images obtained from each subject between transplant hospitalization and study end, or death. When a study was identified as having any new parenchymal abnormality on CT or x-ray (only if CT was unavailable, as in 18/112 PC events), information regarding indication for imaging, radiographic patterns, subsequent diagnostic work-up, and therapeutic management were reviewed. In almost all cases, it was clear when a study was performed for fever or for respiratory complaints. Abnormal parenchymal findings,

as documented by the clinical radiologist in the EMR, were only deemed to be relevant to a PC if the transplant physician caring for the patient elicited a treatment plan based on these findings and the clinical picture.

Specific diagnoses were assigned to PC events only if supportive diagnostic results (Table 3) corroborated with the transplant provider's clinical impression, as documented in the EMR. PCs thought due to volume overload or cardiogenic pulmonary edema (as defined by radiographic appearance, supportive echocardiogram and/or BNP data, and stated as such in the EMR) were excluded when considered isolated explanations for the presenting respiratory signs or symptoms. Only a few cases were excluded on this basis. An impression of atelectasis by the clinical radiologist was also excluded if considered incidental (e.g. fever work-up but no respiratory complaints, with subsequent identification of an extra-pulmonary fever source); however, an infiltrate initially identified as atelectasis in a patient with hypoxia that clearly evolved into pneumonia on repeat imaging was considered a PC. This occurred in 3 cases, highlighting the importance of clinical correlation.

After PCs were identified, a detailed review of the clinical data and provider notes surrounding each PC event was repeated. Adjudication by two other study physicians (one pulmonary and one infectious disease specialist) was performed for complex cases.

Microbiota analysis

According to previously published methods by Taur *et al*, serial fecal specimens were collected from each patient during transplant hospitalization**,** beginning with pre-transplant conditioning (up to 15 days before stem cell infusion) until 35 days post-transplant or hospital discharge[6].

Following DNA extraction and purification from each specimen, the V1-V3 region of the 16s rRNA gene was amplified by polymerase chain reaction (PCR) using modified universal bacterial primers. Purified PCR products were sequenced on a 454 GS FLX Titanium platform. Sequences were filtered, aligned to the full-length 16s rRNA gene, and grouped into operational taxonomic units of 97% similarity according to our described methods[6,9].

Microbiota predictors were included in the time-dependent analysis from day 0 forward (not the start of pre-transplant conditioning, e.g. day -10, as in the prior analysis). A "baseline" microbiota specimen was defined as the fecal specimen collected just prior to day 0, and not the first specimen obtained upon hospitalization for transplant. Thus baseline microbiota characteristics, such as low diversity, do not necessarily reflect pre-hospitalization conditions in this study, since all patients would have received at least some antimicrobials for prophylaxis prior to HCT. However, as shown by Taur *et al*, fecal microbiota from this cohort had a largely diverse, non-dominant baseline at the time of transplant hospitalization, and gut microbial composition shifted rapidly upon exposure to transplant-related treatments soon after admission[6].

Results

Clinical observations regarding Enterococcus faecium as a potential respiratory pathogen

There were two interesting clinical scenarios pertaining to Enterococcus and PCs, as suggested in the footnote to Table 3. In both subjects, fecal domination by *E faecium* preceded the development of a PC, but in only one case was there contemporaneous identification of Vancomycin-resistant *Enterococcus faecium* (VRE) in blood to suggest translocation as a

possible mechanism. In this case, chest CT findings were suspicious for septic emboli in a subject with fever, dyspnea and blood culture positive for VRE. This was thought to explain his multi-focal pneumonia, supported further by improvement after targeted treatment of VRE.

The second subject underwent diagnostic bronchoscopy to evaluate diffuse infiltrates in the setting of fever and hypoxia. Examination of bronchoalveolar lavage fluid (BAL) was consistent with diffuse alveolar hemorrhage (DAH); however endobronchial biopsy was performed on an unexpected airway lesion, and tissue culture revealed VRE (BAL cultures were non-diagnostic). The leading differential diagnosis for the diffuse process in this case was inflammatory, e.g. IPS, but treatment was nonetheless directed at VRE given the degree of acuity and immunocompromise. It is unclear whether or not VRE was contributory.

Discussion

Pulmonary complications as a clinical endpoint

There is no formal validated or reproduced approach to defining combined clinical and radiographic endpoints for lung injury in these complex HCT patients. Presenting signs, symptoms and radiographic findings associated with pulmonary processes are highly variable, due in large part to heterogeneous inflammatory responses during pre-and peri-engraftment periods, as well as myriad co-morbidities. The HCT literature is generally comprised of small, retrospective cohort or case control studies limited to patient subgroups (e.g. umbilical cord blood recipients) or diagnostic categories (e.g. fungal pneumonia).

For this study, a broad clinical and radiographic definition for PCs was considered an acceptable approach given imprecise diagnostic and interpretive algorithms for defining PCs in

subjects with hematological malignancies, HCT or other immunocompromised states. For this study, case selection was not limited to a specific presentation (e.g. hypoxia) or radiographic pattern (e.g. nodular infiltrates), in order to provide an overview of clinical landscape.

Table E1. Gut microbiota states during transplant hospitalization by mortality status

Microbiota State	**Dead** n=48 (51.0%)	**Alive** n=46 (48.9%)
Taxon Domination*		
Gammaproteobacteria[ll]	9 (18.8%)	2 (4.3%)
Enterococcus	23 (47.9%)	15 (32.6%)
Streptococcus	17 (35.4%)	12 (26.1%)
Low Diversity[†]		
Baseline[‡]	7 (14.6%)	8 (17.4%)
Engraftment[§]	28 (58.3%)	15 (32.6%)

*Defined as relative abundance greater than 30% and the most dominant taxon in a stool specimen collected during initial transplant hospitalization (until day 35, or death).[6] 7/94 had a domination state at as the time of stem cell infusion (day 0).

[†] Low diversity as defined by Shannon Diversity Index <1.5[33, 34]

[‡]Baseline" defined as the pre-transplant specimen collected most proximal to day 0.

[§]2 individuals did not engraft

[ll] Borderline statistical significance (p=0.051) using two-sided P-values and based on Fisher's Exact Test, with a significance threshold of P<0.05

Table E2. Predictors of pre-engraftment PCs following allogeneic HCT

	Univariate		Multivariate	
Clinical Predictor	**Hazard Ratio**	**P-value**	**Hazard Ratio**	**P-value**
Age (≥60y)	0.90 (0.27 - 2.44)	0.843		
Sex (female)	2.18 (0.85 - 6.10)	0.104	2.18 (0.66 - 7.77)	0.199
Race (white vs. other)	0.91 (0.33 - 3.06)	0.867		
Disease (leukemia vs. other)	2.55 (0.98 - 7.64)	0.056	3.05 (0.80 - 12.96)	0.103
Donor stem cell type (matched related vs. other)	0.47 (0.16 - 1.29)	0.144	0.24 (0.03 - 2.85)	0.224
Double Umbilical Cord	2.66 (0.96 - 7.52)	0.061	0.54 (0.07 - 6.97)	0.590
T-cell depletion	0.46 (0.11 - 1.44)	0.188	0.44 (0.08 - 2.19)	0.312
Conditioning Intensity (myeloablative vs. other)	2.64 (0.95 - 8.01)	0.062	2.26 (0.65 - 8.31)	0.200
Busulfan	0.76 (0.19 - 2.34)	0.654		
HCT Co-morbidity Index[†] (>=4 vs. other)	1.50 (0.55 - 3.85)	0.411		

Disease risk[‡] (high vs. other)	1.03 (0.40 - 2.64)	0.950		
Abnormal baseline PFTs	0.95 (0.37 - 2.54)	0.916		
Bloodstream infection[*]	0.73 (0.14 - 2.41)	0.638		
Shannon Diversity Index[§]	**3.46 (1.12 - 9.56)**	**0.032**	**6.30 (1.42 - 31.80)**	**0.016**
<1.5 at baseline				
Gammaproteobacteria domination[*]	0.43 (0.00 - 3.28)	0.504		
Enterococcus domination[*]	**5.72 (2.12 - 15.45)**	**0.001**	1.38 (0.38 - 4.95)	0.625
Streptococcus domination[*]	**0.12 (0.00 - 0.95)**	**0.043**	0.29 (0.00 - 2.78)	0.342
Vancomycin	3.33 (0.44 - 425.96)	0.310		
Beta-Lactam	1.63 (0.53 - 6.56)	0.417		
Fluoroquinolone	**4.11 (1.59 - 11.14)**	**0.004**	**4.24 (1.15 - 17.00)**	**0.030**
Metronidazole	1.31 (0.40 - 3.58)	0.626		
Azithromycin	**9.42 (1.02 - 39.82)**	**0.049**	**26.77 (2.20 - 226.61)**	**0.015**

Statistically significant p value < 0.05

[*] Time-dependent variables (note: pre-engraftment model does not include GVHD)

[†] HCT Co-morbidity Index groups: low (0-1), intermediate (2-3) or high (≥ 4)[30]

‡ Disease Risk: ASBMT RFI Classification as low, intermediate or high[29]

§ Shannon Diversity Index takes into account number of species and relative abundance[33]

‖ Vancomycin includes intravenous mode of administration only

** Beta lactams include cephalosporins, beta-lactam-beta-lactamase combinations and carbapenems

Abbreviations: PFT= pulmonary function test

Table E3. Predictors of post-engraftment PCs following allogeneic HCT

	Univariate		Multivariate	
Clinical Predictor	**Hazard Ratio**	**P-value**	**Hazard Ratio**	**P-value**
Age (≥60y)	1.25 (0.70 - 2.17)	0.444		
Sex (female)	1.10 (0.64 - 1.88)	0.721		
Race (white vs. other)	1.11 (0.57 - 2.42)	0.773		
Disease (leukemia vs. other)	0.88 (0.51 - 1.50)	0.641		
Donor stem cell type (matched related vs. other)	0.93 (0.54 - 1.65)	0.799		
Double Umbilical Cord	0.85 (0.40 - 1.61)	0.632		
T-cell depletion	1.05 (0.62 - 1.79)	0.854		
Conditioning Intensity (myeloablative vs. other)	0.88 (0.52 - 1.50)	0.650		
Busulfan	1.57 (0.90 - 2.69)	0.111	1.45 (0.74 - 2.78)	0.274
HCT Co-morbidity Index[†] (>=4 vs. other)	**2.49 (1.40 - 4.30)**	**0.002**	**2.37 (1.25 - 4.42)**	**0.009**
Disease risk[‡] (high vs. other)	1.53 (0.90 - 2.64)	0.115	1.36 (0.71 - 2.60)	0.353

Abnormal baseline PFTs	**2.01 (1.16 - 3.62)**	**0.013**	1.49 (0.79 - 2.89)	0.219
Acute GVHD (Day 0 to 100)*	1.54 (0.79 - 2.89)	0.198	1.64 (0.78 - 3.39)	0.189
Bloodstream infection*	1.27 (0.73 - 2.18)	0.394		
Shannon Diversity Index§ <1.5 at baseline	0.77 (0.32 - 1.59)	0.507		
Gammaproteobacteria domination*	**3.33 (1.43 - 6.87)**	**0.007**	**3.68 (1.49 - 8.21)**	**0.006**
Enterococcus domination*	0.81 (0.45 - 1.41)	0.473		
Streptococcus domination*	1.25 (0.69 - 2.19)	0.447		
Vancomycin*,‖	0.53 (0.23 - 1.54)	0.218		
Beta-Lactam*,**	0.92 (0.43 - 2.42)	0.857		
Fluoroquinolone*	1.53 (0.89 - 2.66)	0.125	1.27 (0.67 - 2.44)	0.468
Metronidazole*	1.05 (0.60 - 1.80)	0.866		
Azithromycin*	1.83 (0.95 - 3.33)	0.071	1.85 (0.88 - 3.66)	0.103

Statistically significant p value < 0.05

* Time-dependent variables

† HCT Co-morbidity Index groups: low (0-1), intermediate (2-3) or high (≥ 4)[30]

‡ Disease Risk: ASBMT RFI Classification as low, intermediate or high[29]

§ Shannon Diversity Index takes into account number of species and relative abundance[33]

‖ Vancomycin includes intravenous mode of administration only

** Beta lactams include cephalosporins, beta-lactam-beta-lactamase combinations and carbapenems

Abbreviations: PFT= pulmonary function test; GVHD = graft-versus-host-disease

Figure Legend

Figure F1. Heat map of species belonging to the class Gammaproteobacteria identified in all specimens collected from subjects with and without PCs.

Figure F1 is a heat map showing the log-transformed maximum relative abundance of Gammaproteobacteria species identified in all 439 stool specimens collected from 94 subjects during transplant hospitalization. Each column represents one subject, stratified by PC status. Each row represents bacterial taxa resolved to the species level. The maximum relative abundance of each species depicted takes into account all specimens from each subject (range 3-8 per subject), and does not reflect one sample collected at a single time point. The most abundant Gammaproteobacteria species identified were *E coli* and *Klebsiella spp*, including common respiratory pathogens. *Abbreviation: PC= pulmonary complication*

Figure F2. Alpha diversity of fecal bacterial communities identified in all stool specimens from subjects with and without PCs.

Figure F2 shows alpha (within-sample) diversity, of bacterial communities identified in all stool specimens (n=439) from 94 subjects, as summarized by the Shannon Diversity Index (SDI; diversity increases from 0 along the Y-axis). The range of SDI values assigned to each specimen is wide, from close to 0 (not diverse) to greater than 3 (diverse) in subjects both with and without PCs. *Abbreviation: PC= pulmonary complication*

Figure F3. Beta diversity of fecal bacterial communities, grouped by taxon-domination state and by PC status.

Figure F3 shows beta (between-sample) diversity of bacterial communities identified in all stool specimens (n=439) collected from 94 subjects, as calculated by Bray-Curtis dissimilarity principal coordinates analysis. Principal coordinate 1 (PC1, x-axis) explains 54.1% of variance, and principal coordinate 2 (PC2, Y-axis) explains 19.5% of the variation in Bray-Curtis dissimilarity between samples grouped by PC status (shape: triangle = PC, circle = no PC) and by taxon domination state (color: red=Proteobacteria, green=Enterococcus, tan=Streptococcus and grey= other dominant taxa or specimens collected outside of the analysis period). Clustering of PCs by domination state is most apparent in the setting of Proteobacteria domination (red triangles) and in a subset of subjects with Enterococcus domination (green triangles). *Abbreviation: PC= pulmonary complication*


no PC
PC
Escherichia coli
Klebsiella variicola
Klebsiella pneumoniae
Leclercia adecarboxylata
Enterobacter asburiae
Citrobacter freundii
Klebsiella oxytoca
Haemophilus parainfluenzae
Pseudomonas aeruginosa
Hafnia alvei
Citrobacter braakii
Stenotrophomonas geniculata
Trabulsiella farmeri
Klebsiella sp. PB12
ambiguous Klebsiella
Aggregatibacter segnis
Shigella flexneri
Haemophilus sputorum
Rahnella aquatilis
Pantoea sp. 2344
Serratia marcescens
Erwinia billingiae
Enterobacter mori
Enterobacter aerogenes
unclassified
Cronobacter dublinensis
Shigella sonnei
Escherichia fergusonii ATCC 35469
Citrobacter werkmanii
ambiguous Legionella
Klebsiella sp. A712
Shigella dysenteriae
Proteus mirabilis
Pseudomonas migulae
Klebsiella sp. Y10
Shigella sp. CTB486
Klebsiella sp. NTUIOB KSA10
ambiguous Escherichia
Salmonella enterica subsp. enterica serovar Paratyphi A
Klebsiella sp. ICB527
Pantoea stewartii subsp. indologenes
Pectobacterium wasabiae
Pseudomonas sp. B3044
Citrobacter murliniae
Enterobacter oryzae
Klebsiella sp. Y1[2011]
Enterobacter hormaechei
Pseudomonas poae
Moraxella nonliquefaciens
Pseudomonas otitidis
Enterobacter ludwigii
Citrobacter koseri
Enterobacter sp. MPR11
Cronobacter malonaticus
Shigella dysenteriae Sd197
ambiguous Pantoea
Enterobacter gergoviae
Enterobacter sp. ICB479
Pantoea sp. NCCP−507
Pseudomonas sp. sw2
Dickeya sp. PC1
Pectobacterium atrosepticum SCRI1043
Haemophilus ducreyi ATCC 33921
Pseudomonas sp. enrichment culture clone 18[2012]
Escherichia coli O157:H7
Erwinia chrysanthemi
Enterobacter sp. NCCP−755
endosymbiont of Columbicola baculoides
Acinetobacter baumannii Naval−17
Pseudomonas sp. enrichment culture clone DTT.001.L
Haemophilus sputorum HK 2154
Enterobacter sp. MLB07
Legionella pneumophila
Moraxella osloensis
Ewingella americana
Pseudomonas gessardii
Haemophilus sp. Smarlab 3302188
Pseudomonas luteola
Pantoea sp. A2066
Citrobacter sp. enrichment culture clone BR40
Nevskia ramosa
Acinetobacter johnsonii
Enterobacter amnigenus
Pantoea sp. NCCP−532
Klebsiella sp. 50MoSuero
Kluyvera intermedia
Pseudomonas brassicacearum
Enterobacter sp. 2068
Klebsiella sp. SOR89
Kluyvera cryocrescens
Pantoea sp. PSI 001
Enterobacter cloacae
Stenotrophomonas sp. YRR09
Pasteurella pneumotropica
Trabulsiella sp. LB10
Salmonella enterica subsp. enterica serovar Typhi
Salmonella enterica subsp. enterica serovar Gallinarum
Salmonella enterica subsp. enterica serovar Enteritidis
Salmonella enterica subsp. enterica serovar Agona
Haemophilus sputorum CCUG 13788
Vibrio alginolyticus E0666
Coxiella cheraxi
Pseudomonas sp. L19
Pseudomonas sp. K1
Serratia nematodiphila
Pantoea dispersa
Stenotrophomonas chelatiphaga
Salmonella subterranea
Klebsiella sp. 29MoSuero
Pseudomonas sp. N1
Vibrio sp. PaD1.21b
Klebsiella sp. G1−6
Erwinia dispersa
Escherichia sp. A441
Citrobacter farmeri
Enterobacter sp. M34
marine bacterium pa−18
Klebsiella sp. TG−1
Pectobacterium carotovorum
Enterobacter soli
Shigella sp. XJ149−N3G1
Acinetobacter sp. 7m11
Citrobacter freundii ATCC 8090 = MTCC 1658
Salmonella enterica subsp. enterica serovar Typhimurium str. LT2
Shigella boydii
Pantoea wallisii
Gammaproteobacteria Species
Maximum Relative Abundance
0.1
0.01
0.001
0

Figure F1

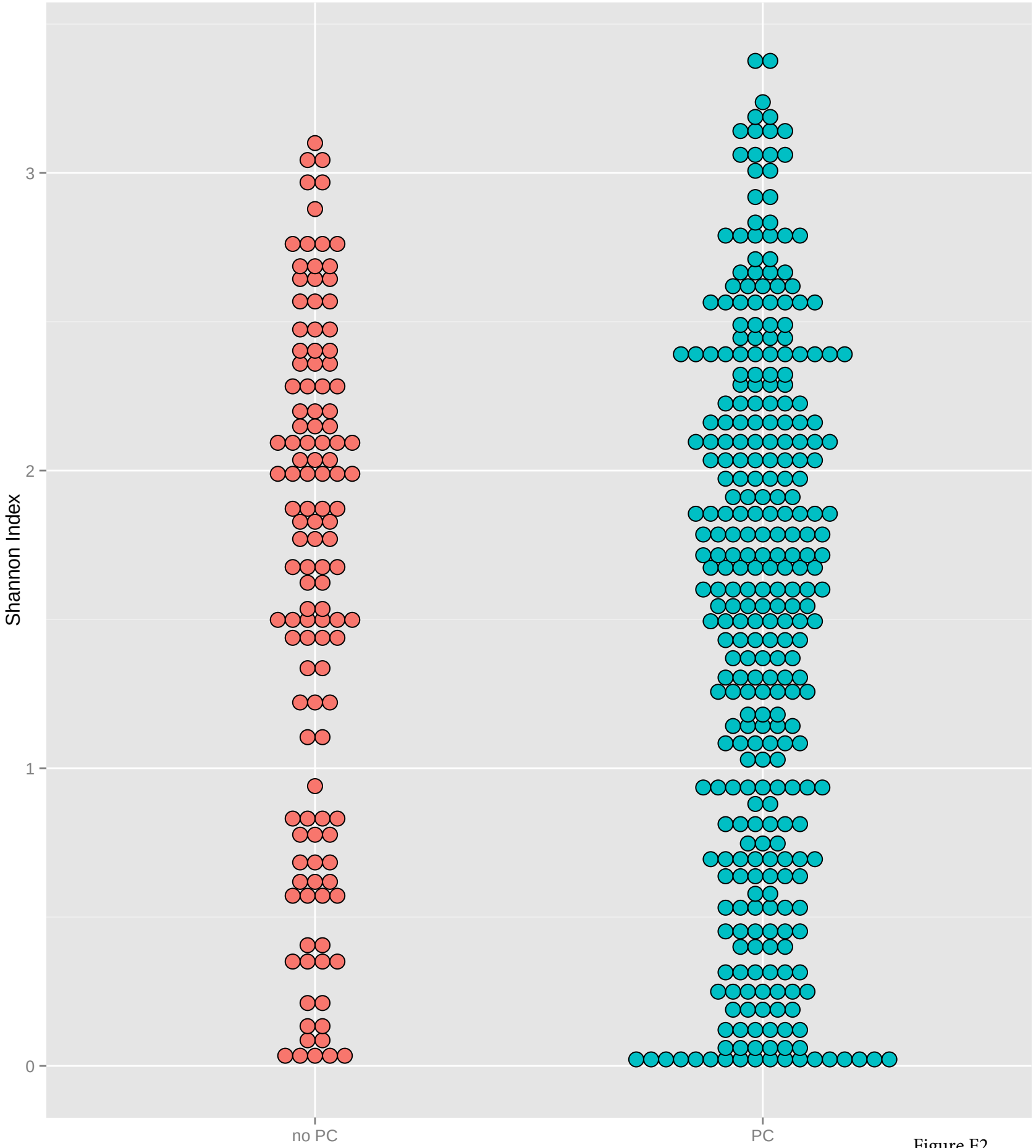

Figure F2

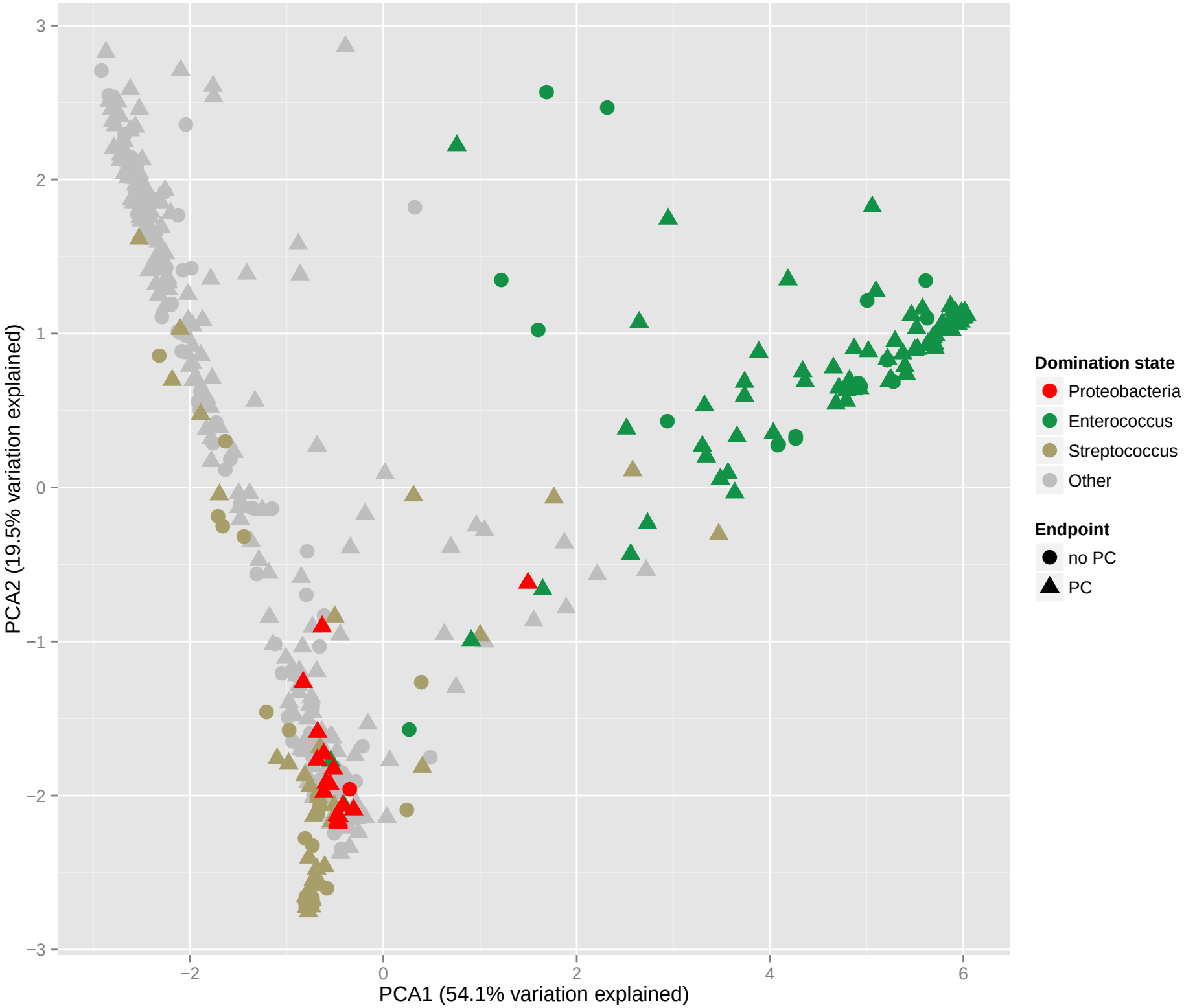
PCA2 (19.5% variation explained)
PCA1 (54.1% variation explained)
3
2
1
0
−1
−2
−3
−2
0
2
4
6
Domination state
Proteobacteria
Enterococcus
Streptococcus
Other
Endpoint
no PC
PC

Figure F3