Hyman *et al*.

Diversity of the vaginal microbiome correlates with preterm birth.

MS # RSCI-13-021

Supplementary data

Supplementary Tables

Table S1. Participant characterization by risk, ethnicity, pregnancy outcome, and vaginal swabs collected and analyzed.

ID	Risk	Ethnicity	Outcome	Swab 1	Swab 2	Swab 3
001	H	Other	Term		X	X
002	H	African-American	Term		X	X
003	H	Hispanic	Term	X		
004	H	Asian	Term		X	X
005	H	Caucasian	Term	X	X	X
006	H	African-American	PTB			X
007	H	Caucasian	PTB		X	
008	H	Caucasian	Term		X	
009	H	Asian	PTB			X
012	L	Asian	PTB		X	
015	L	Asian	Term		X	
016	L	Other	Term		X	
017	L	Other	Term		X	
018	L	Caucasian	Term		X	
019	L	Caucasian	Term		X	
021	L	Caucasian	Term		X	
026	L	Caucasian	Term	X	X	X
027	L	Other	Term	X	X	X
028	L	Caucasian	Term			X
029	L	Caucasian	Term	X	X	
030	L	Caucasian	Term	X	X	X
031	L	Caucasian	Term	X		
032	L	Caucasian	Term	X		
034	L	Caucasian	Term	X	X	
035	L	Hispanic	SAB	X		
036	L	Caucasian	Term	X		
038	L	Caucasian	Term	X	X	
039	L	Caucasian	Term			X
042	H	Caucasian	Term		X	X
044	L	Hispanic	PTB			X
045	L	Other	Term	X	X	
048	H	Asian	Term			X
049	H	Asian	Term	X	X	X
050	H	Caucasian	Term		X	
051	L	Other	Term	X	X	
052	L	Asian	Term		X	
053	L	Hispanic	Term	X		
055	H	Hispanic	Term			X
057	L	Caucasian	Term	X	X	

058	L	Caucasian	PTB*	X	X	X
059	H	Hispanic	Term			X
060	L	Caucasian	PTB	X	X	X
061	L	Other	Term		X	
062	L	Asian	Term	X		
063	H	Asian	Term			X
064	L	Caucasian	SAB	X		
065	L	Caucasian	Term	X	X	
067	L	Caucasian	Term	X		
069	L	Other	PTB			X
070	H	Asian	Term	X	X	X
072	H	Hispanic	Term		X	X
073	H	African-American	Term	X		
074	H	African-American	SAB	X		
075	H	Hispanic	Term			X
076	H	Asian	Term			X
077	H	Caucasian	Term			X
078	H	Hispanic	Term			X
079	H	Caucasian	Term	X	X	X
081	H	African-American	PTB			X
082	H	Other	Term			X
083	H	Caucasian	PTB	X	X	
084	H	Asian	Term			X
085	H	Caucasian	Term	X	X	X
086	H	Other	PTB	X	X	
087	H	Caucasian	PTB		X	
088	H	Asian	Term		X	
089	L	Hispanic	PTB			X
090	H	Caucasian	Term	X	X	X
091	H	Hispanic	Term	X	X	X
092	H	Caucasian	PTB		X	
093	H	Hispanic	Term		X	X
097	H	African-American	Term		X	X
098	H	Caucasian	SAB	X		
099	H	Asian	Term	X		X
100	L	Asian	PTB			X
101	H	Caucasian	Term	X	X	
102	H	Caucasian	Term	X		X
103	H	Asian	PTB	X	X	
104	L	Caucasian	Term	X		
105	L	Caucasian	SAB	X		
106	L	Caucasian	Term	X		
107	L	African-American	Term	X	X	X
108	L	Hispanic	Term	X	X	X
109	H	Caucasian	PTB		X	
110	H	Caucasian	Term		X	
111	L	Other	Term			X
112	L	Caucasian	Term	X	X	
113	H	African-	Term		X	X

		American				

Risk Factor for PTB: H = high-risk. L = low-risk. Outcomes: term; PTB, and SAB (miscarriage). Ethnicity: Caucasian, African-American, Hispanic, Asian, Other (mixed ethnicity, Pacific Islander, native Hawaiian, native American, and Alaska native). X = swabs for which there is sequence data. *Induced.

Table S2. Participant ethnicities and number of vaginal swabs.

Ethnicity	Number of participants	Percentage	Number of swabs	Percentage
Asian	16	18	23	17
African-American	8	9	13	9
Caucasian	40	45	66	48
Hispanic	13	15	19	14
Other	11	13	17	12
Total	88	100%	138	100%

Other= mixed ethnicity, Pacific Islander, native Hawaiian, Native American, and Alaska native.

Table S3. **A.** All of the bacterial identification data are available on the SGTC website. http://med.stanford.edu/sgtc/research/microbiome_preterm_birth.html . The percent of the sequence reads supporting the presence of any given bacterium was recorded when that percent was at least 1% (usually 3 sequence reads). Bacteria supported by <1% of the sequence reads are indicated by an asterix. **B.** The bacterial identification data for the 14 participants for whom we have a swab in all three trimesters. These data were abstracted from Table S3A.

Table S4. Little or no *Lactobacillus* (13 swabs).

Risk	**Outcome**	**ID/Swab**	***Lactobacillus***	**Percent**
Low	PTB	None		
	Term	107-1	Uncultured	7
		107-2	*L. iners*	8
		026-3		Zero
	SAB	None		
High	PTB	103-1		Zero
		103-2		Zero
	Term	001-2	*L. iners*	6
		059-3	*L. iners*	8
		070-1		Zero
		070-2		Zero
		070-3		Zero
		110-2	*L. acidophilus* + *L. gasseri*	4 + 9 = 13
		113-2	*L. iners*	6
	SAB	074-1	*L. iners* + *L. ultunensis*	6 + 17 = 23

Uncultured = uncultured *Lactobacillus*, as per the RDP.

Table S5. Putative Novel bacteria (> or = 1%). The closest named bacterium is given for 14 swabs.

Risk	**Outcome**	**ID/Swab**	**Closest named bacterium**	**Percent**
Low	PTB	None		
	Term	031-1	*Lactobacillus*	1

		053-1	*Bacteroidales*	1
			Parvimonas	1
			Prevotella	10
		104-1	*Mycoplasma*	6
			Prevotella	11
		107-2	*Parvimonas*	1
			Prevotella	1
		107-3	*Dialister*	1
	SAB	None		
High	PTB	None		
	Term	005-3	*Lactobacillus*	1
		070-3	*Lachnospiraceae*	4
		073-1	*Megasphaera*	1
		076-3	*Lactobacillus*	100
		079-1	*Lactobacillus*	3
		084-3	*Lactobacillus*	3
		090-3	*Lactobacillus*	1
		091-2	*Prevotella*	2
	SAB	074-1	*Prevotella*	33

Table S6. Percentage (> or =1%) of *Atopobium* content of 23 swabs.

Risk	**Outcome**	**ID/Swab**	***Atopobium* (%)**	**Uncultured *Atopobium* (%)**	**Total Atopobium (%)**
Low	PTB	None			
	Term	026-3	100		100
		057-2	4		4
		104-1	3	3	6
		107-1	4	5	9
		107-2	6	1	7
		107-3		1	1
		108-2		1	1
	SAB	None			
High	PTB	None			
	Term	001-2	2	3	5
		002-2	4	4	8
		002-3	4	29	33
		059-3	2	8	10
		070-1	5	19	24
		070-2	9		9
		070-3		11	11
		073-1	3	2	5
		075-3	3		3
		078-3		2	2
		079-3		6	6
		091-1	6	3	9
		091-2	2	2	4
		110-2	84		84
		113-2		12	12
	SAB	074-1	6	10	16

Uncultured = uncultured *Atopobium*, as per the RDP.

Table S7. Percentage (> or = 1%) of *Prevotella* content of 22 swabs.

Risk	**Outcome**	**ID/Swab**	***Prevotella* (%)**	**Uncultured *Prevotella* (%)**	**Novel *Prevotella* (%)**	**Total *Prevotella* (%)**
Low	PTB	089-3	6			6
	Term	028-3	4			4
		053-1	6		10	16
		104-1		3	11	14
		107-1	19	23		42
	SAB	None				
High	PTB	None				
	Term	001-2	8	3		11
		002-2	1	4		5
		002-3		29		29
		059-3		8		8
		070-1		19		19
		070-2		2		2
		070-3		11		11
		073-1	4	14		18
		078-3	16	2		18
		079-3		6		6
		090-2	1			1
		091-1	3	9		12
		091-2		4	2	6
		097-2	3			3
		113-2	2	12		14
	SAB	074-1		10	33	43

Uncultured = uncultured *Prevotella*, as per the RDP. Novel = matches < 97% to its closest match in the RDP. For this table, the closest match to a named bacterium is *Prevotella*.

Table S8. Assorted noxious bacteria at > or = 1% for 20 swabs.

Risk	Outcome	ID/Swab	Bacterium (%)	Uncultured (%)	Novel (%)	Total (%)
Low	PTB		*Ureaplasma*			
		069-3	1			1
		089-3	1			1
		100-3	73			73
	Term		*Finegoldia*			
		069-3	3			3
			Gardnerella			
		107-3	1			1
	SAB	None				
High	PTB		*Bifidobacterium*			
		009-3	3			3
		103-1	97			97
		103-2	97			97
	Term	004-3		1		1
		049-1		2		2
		049-2		3		3
			Enterococcus			
		049-1	14			14

			Finegoldia			
		005-1	2			2
			Gardnerella			
		002-2	4			4
		075-3	4			4
			Lachnospiraceae			
		059-3	61			61
		070-2		63		63
		070-3		59	4	63
		113-2		58		58
			Ureaplasma			
		003-1	2			
	SAB	None				

Uncultured = uncultured bacterium, as per the RDP; Novel = matches < 97% to any sequence in the RDP. The closest named bacterium is *Lachnospiraceae.*

Table S9. Shannon Diversity Index (SDI) for each swab.

Risk	**Outcome**	**ID**	**Swab 1**	**Swab 2**	**Swab 3**
Low	PTB	012		0.21	
		044			0.00
		058	0.22	0.00	0.34
		060	0.00	0.23	0.05
		069			2.25
		089			2.27
		100			1.40
	Term	015		0.37	
		016		0.40	
		017		0.25	
		018		0.99	
		019		0.13	
		021		0.00	
		026	0.03	0.03	0.06
		027	0.00	0.00	0.11
		028			1.92
		029	0.32	1.02	
		030	0.23	0.35	0.47
		031	1.18		
		032	0.28		
		034	1.23		
		036	1.33		
		038	0.68	0.17	
		039			1.19
		045	1.04	0.84	
		051	0.00	0.98	
		052		0.07	
		053	2.78		
		057	0.15	0.47	
		061		1.59	
		062	0.00		
		065	0.03	0.03	
		067	0.00		
		104	2.32		

		106	0.00		
		107	3.31	3.28	3.22
		108	0.11	2.17	1.70
		111			1.81
		112	0.00	0.43	
	SAB	035	0.21		
		064	0.00		
		105	0.03		
High	PTB	006			1.45
		007		0.11	
		009			0.63
		081			0.60
		083	0.13	0.26	
		086	0.60	1.08	
		087		0.03	
		092		0.00	0.77
		103	0.23	0.24	
		109		0.00	
	Term	001		2.19	0.12
		002		2.17	2.63
		003	1.25		
		004		0.46	0.66
		005	1.05	0.29	0.20
		008		0.13	
		042		0.03	0.05
		048			1.09
		049	1.95	1.42	1.02
		050		0.54	
		055			0.03
		059			1.95
		063			0.27
		070	2.05	1.59	1.84
		072		1.09	1.18
		073	3.05		
		075			0.57
		076			0.05
		077			0.13
		078			2.29
		079	1.88	1.02	0.65
		082			0.55
		084			0.35
		085	0.00	0.00	0.24
		088		0.16	
		090	0.05	0.32	0.15
		091	2.96	2.23	0.59
		093		1.00	
		097		1.71	0.15
		099	0.92		0.37
		101	1.55	0.03	
		102	0.00		0.42
		110		0.92	
		113		2.14	1.13
	SAB	074	3.03		

		098	0.28		
Number			43	51	44

SAB = miscarriage

Figures

Figure S1. SDI by race/ethnicity

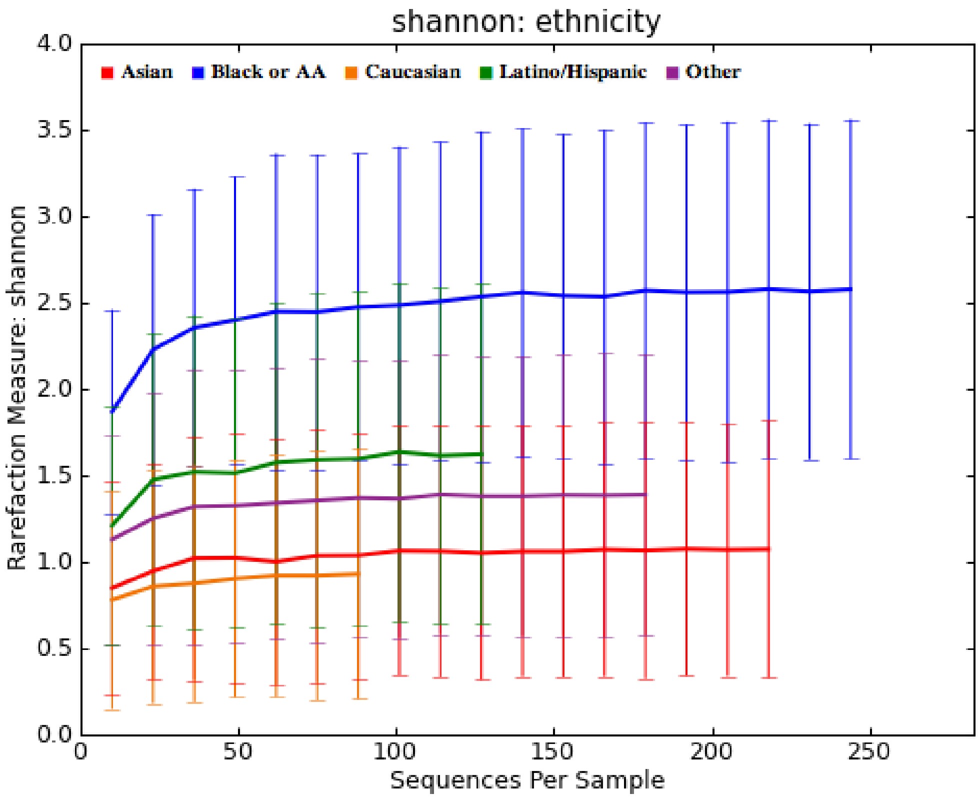
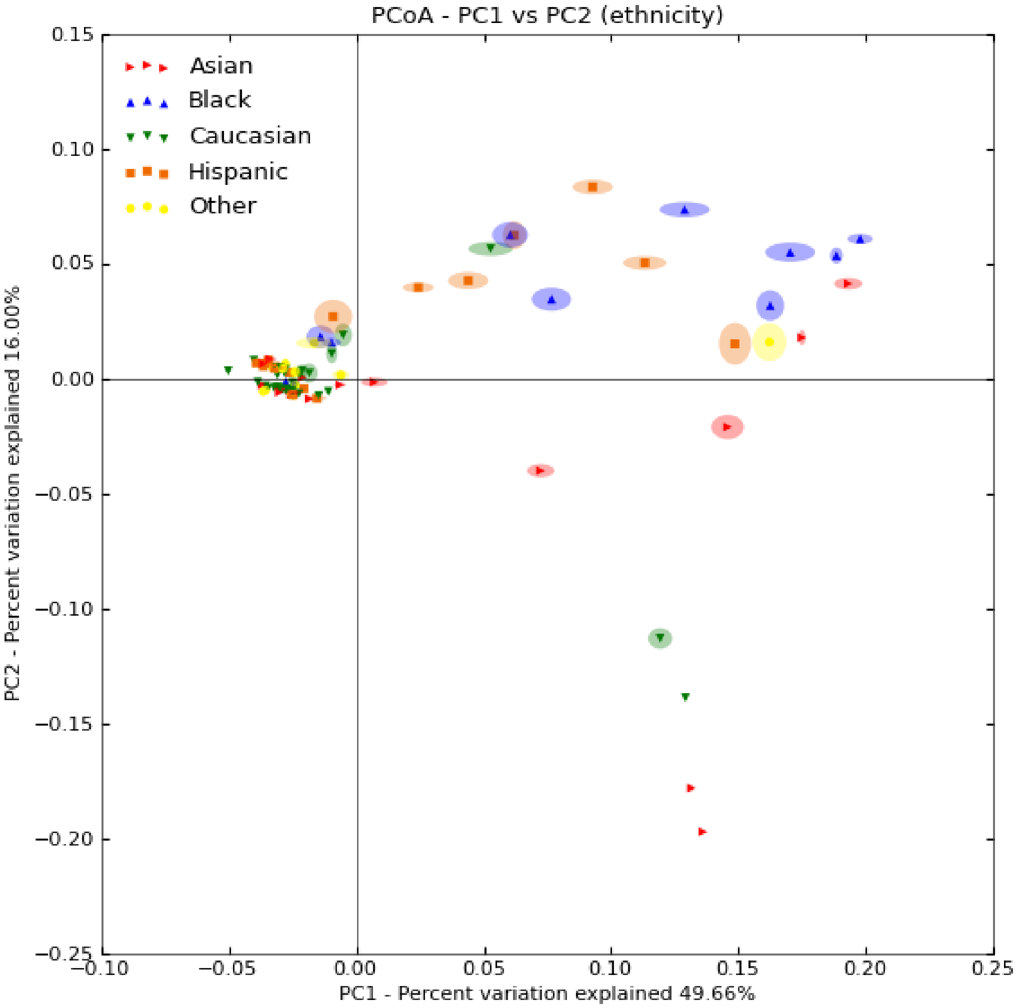
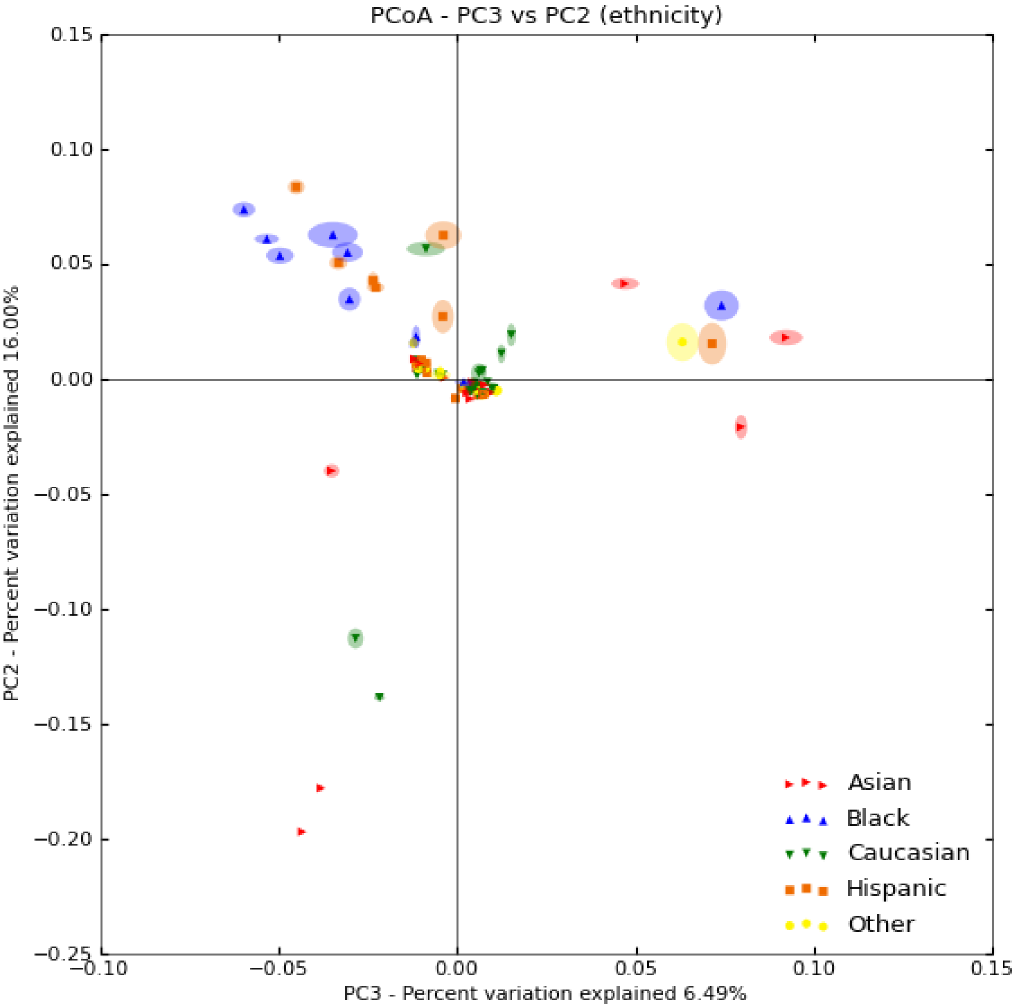


Figure S2. Principal Coordinate Analyses (PCoA) of ethnicity.




PCoA - PC3 vs PC2 (ethnicity)
PC2 - Percent variation explained 16.00%
PC3 - Percent variation explained 6.49%
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.15
−0.20
−0.25
−0.10
−0.05
0.00
0.05
0.10
0.15
Asian
Black
Caucasian
Hispanic
Other

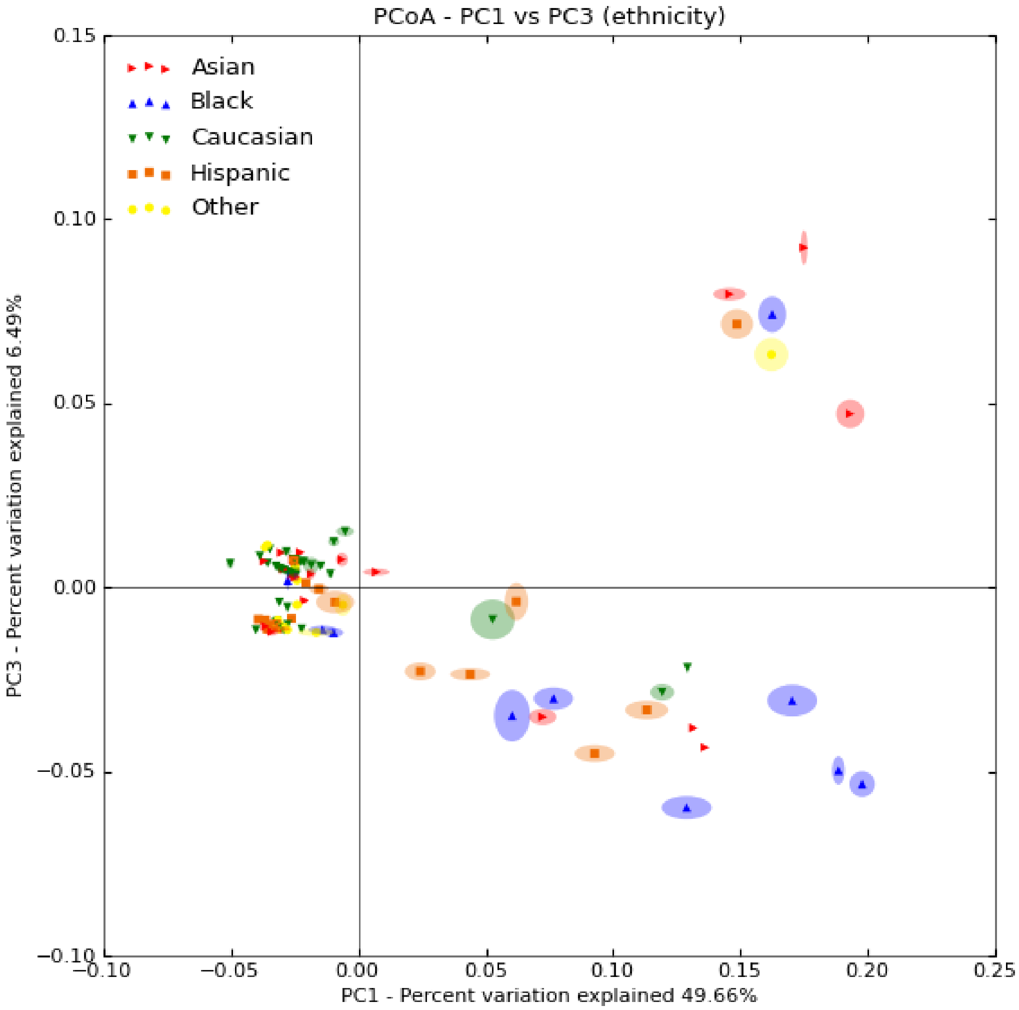
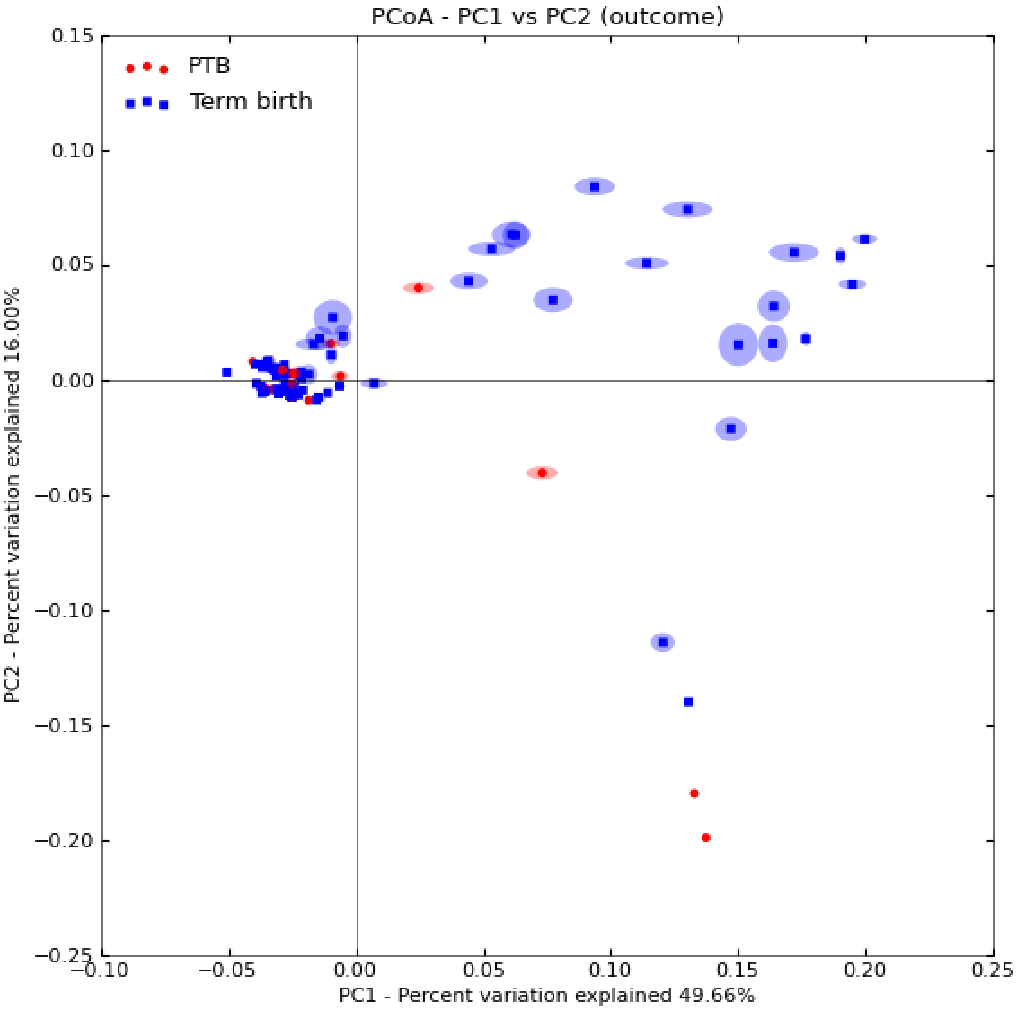
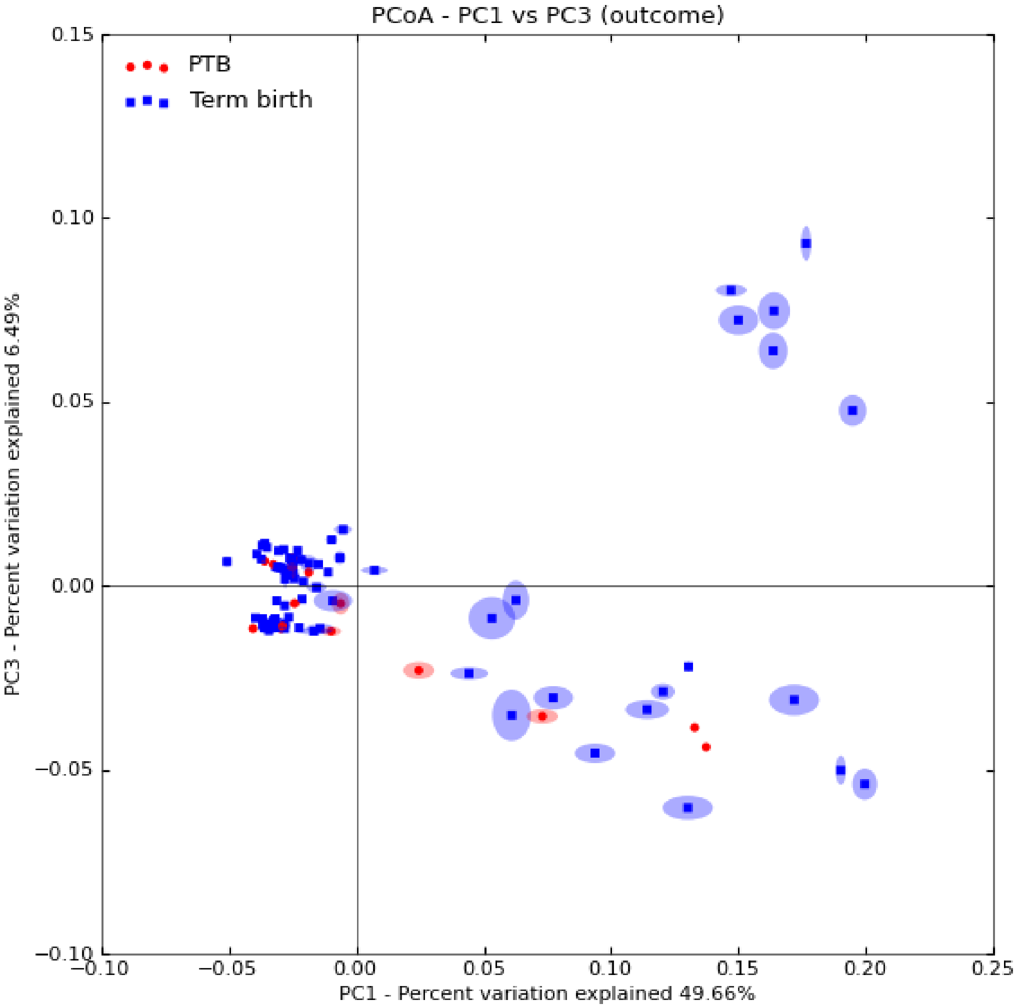
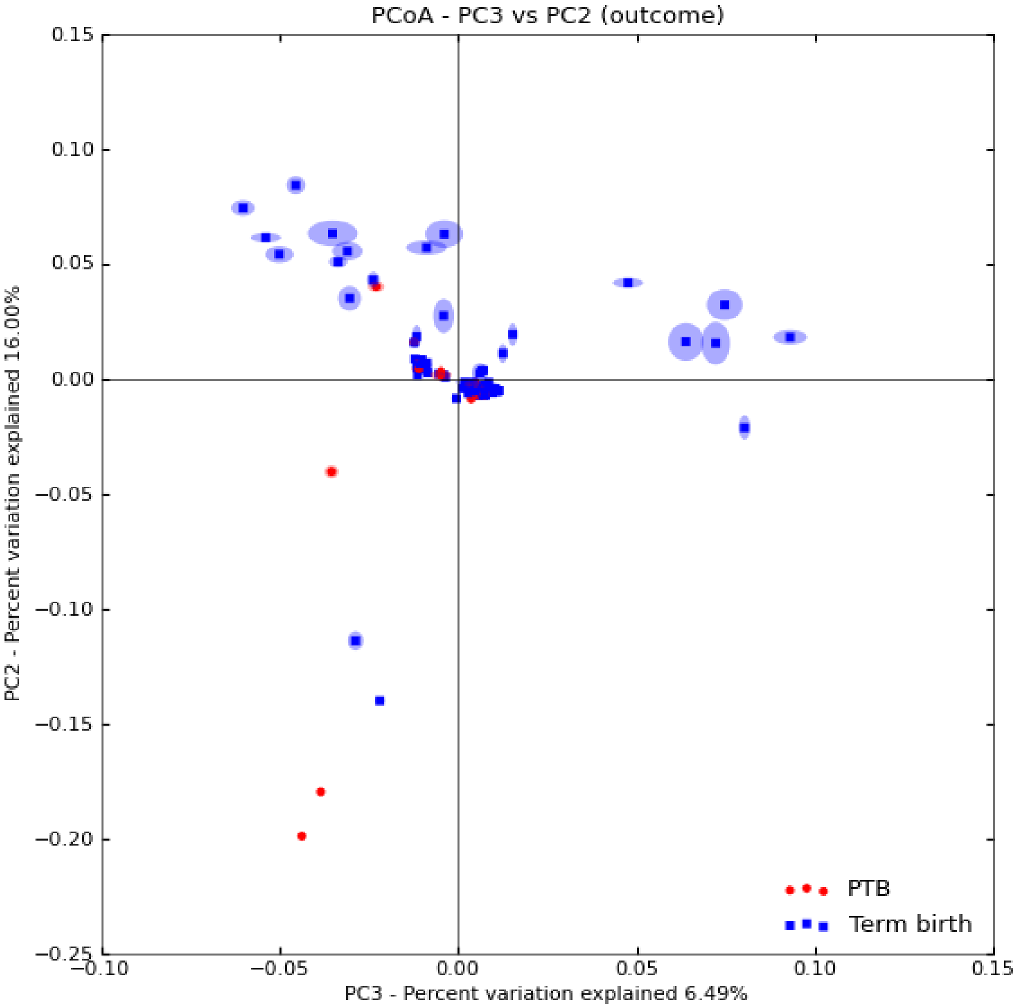

PCoA - PC1 vs PC3 (ethnicity)
Asian
Black
Caucasian
Hispanic
Other
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.10
−0.05
0.00
0.05
0.10
0.15
0.20
0.25
PC3 - Percent variation explained 6.49%
PC1 - Percent variation explained 49.66%

Figure S3. Principal Coordinate Analyses (PCoA) of outcome.




PCoA - PC1 vs PC3 (outcome)
PTB
Term birth
PC1 - Percent variation explained 49.66%
PC3 - Percent variation explained 6.49%
−0.10
−0.05
0.00
0.05
0.10
0.15
0.20
0.25
0.15
0.10
0.05
0.00
−0.05
−0.10


PCoA - PC3 vs PC2 (outcome)
PC2 - Percent variation explained 16.00%
PC3 - Percent variation explained 6.49%
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.15
−0.20
−0.25
−0.10
−0.05
0.00
0.05
0.10
0.15
PTB
Term birth