Prenatal Complications Are Associated With the Postnatal Airway Host Response and Microbiota in Pre-term Infants

ONLINE DATA SUPPLEMENT

Methods

Molecular Biology: DNA extractions were performed using the Qiagen EZ1 Advanced automated extraction platform (Qiagen Inc., Valencia, CA) with the bacterial card and tissue extraction kit. All sample manipulation was done in the BSL2 hood with appropriate laminar flow. Frozen samples were thawed at 4^0 C and vortexed to ensure mixing. An aliquot of 200 µl for extraction was transferred into the tube provided with the EZ1 kit. Remaining sample was placed in a clean 2 ml tube and stored at -70 C. Extraction reagent cartridges, elution tubes and tip holders were loaded into the EZ1 sample rack as instructed by the manufacturer. Elution volume of 100µl was selected and EZ1 DNA Tissue Kit program was run. Elution tubes with DNA extract were stored at -20^0C. DNA extraction reagents were confirmed free of bacterial DNA by performing control extractions utilizing buffer or PCR grade water.

Sequencing: Bacterial profiles were determined by broad-range amplification and sequence analysis of 16S rRNA genes following our previously described methods [1, 2]. Amplicons were generated using primers that target approximately 300 base pairs of the V1/V2 variable region of the 16S rRNA gene. PCR products were normalized using agarose gel densitometry, pooled, gel purified, and concentrated using DNA Clean and Concentrator Kit (Zymo, Irvine, CA). Pooled amplicons was quantified using Qubit Fluorometer 2.0 (Invitrogen, Carlsbad, CA). The pool was diluted to 4nM and denatured with 0.2 N NaOH at room temperature. The denatured DNA was diluted to 20pM and spiked with 10% of the Illumina PhiX control DNA prior to loading the sequencer. Illumina paired-end sequencing was performed on the Miseq platform using a 500 cycle version 2 reagent kit.

Analysis of Illumina Paired-end Reads. Illumina MiSeq paired-end sequences were sorted by sample via barcodes in the paired reads with a python script [1]. Sorted paired end sequence data

were deposited in the NCBI Short Read Archive under accession number SRP044029. The sorted paired reads were assembled using phrap [3, 4]. Pairs that did not assemble were discarded. Assembled sequence ends were trimmed over a moving window of 5 nucleotides until average quality met or exceeded 20. Trimmed sequences with more than 1 ambiguity or shorter than 200 nt were discarded. Potential chimeras identified with Uchime (usearch6.0.203_i86linux32) [5] using the Schloss Silva reference sequences were removed from subsequent analyses[6]. Assembled sequences were aligned and classified with SINA (1.2.11) [7]using the 629,125 bacterial sequences in Silva 111[8] as reference configured to yield the Silva taxonomy. Operational taxonomic units (OTUs) were produced by combining sequences with identical taxonomic assignments.

FIGURE LEGEND

Figure E1. Correlation coefficients between the inflammatory protein biomarkers and the prominent taxa revealed patterns of association between groups of proteins and dominance of bacteria within specimens for *Staphylococcus*, *Streptococcus*, and *Gemella.*

Figure E2: (A) Unsupervised clustering using the RandomForest package in R of the 12 protein analytes was performed and consisted of 5,000 trees (all other default parameters were used).[9] Random forests (RF) consisting of 5,000 classification trees were used to perform unsupervised clustering of subjects based on the 12 panel inflammatory protein profile. Partitioning around medoids clustering on the resulting proximity matrix was used to determine phenotypes and results in 3 clusters of subjects. (B) Individual proteins are displayed across each of the three clusters. Protein levels (pg/mL) were log transformed for display.

Figure E3: Subject level microbial communities are displayed across the three phenotype groups (columns) and the two different sites (rows). Bacteria are identified at the genus level.

Table E1. Protein levels across clusters

cluster	Protein	N	Median pg/mL	Minimum pg/mL	Maximum pg/mL
1	**VEGF**	24	455.89	96.66	1075.53
	GM-CSF	24	38.22	4.63	292.40
	TNF-a	24	111.40	9.22	1536.38
	IL-6	24	1324.52	40.65	4200.00
	IL-4	24	32.82	8.32	80.73
	IL-1B	24	1400.00	95.92	1500.00
	IL-8	24	3200.00	476.59	22829.15
	IL-5	24	0.82	0.45	3.19
	IL-2	24	9.02	1.38	32.88
	IL-10	24	5.50	1.30	43.29
	IFN-y	24	6.89	0.15	22.22
	IL-12	24	17.43	5.82	44.04
2	**VEGF**	23	353.52	18.57	1654.05
	GM-CSF	23	28.02	2.86	208.29
	TNF-a	23	215.55	14.02	1347.46
	IL-6	23	1551.82	323.90	15567.60
	IL-4	23	8.42	0.00	23.70
	IL-1B	23	1768.60	34.07	29977.28
	IL-8	23	24527.34	664.24	210490.60
	IL-5	23	0.37	0.17	2.51
	IL-2	23	1.37	0.31	8.47
	IL-10	23	0.23	0.23	2.11
	IFN-y	23	0.15	0.00	0.15
	IL-12	23	3.45	3.45	11.59
3	**VEGF**	24	105.46	0.00	765.10
	GM-CSF	24	8.90	0.45	256.55
	TNF-a	24	5.01	0.39	47.25
	IL-6	24	44.60	2.77	850.90
	IL-4	24	15.59	0.00	24.50
	IL-1B	23	34.25	3.31	438.47
	IL-8	24	1353.22	129.71	5520.33
	IL-5	24	0.17	0.00	0.61
	IL-2	24	0.15	0.00	2.32
	IL-10	24	0.23	0.00	3.37
	IFN-y	24	0.00	0.00	1.01
	IL-12	24	3.45	0.00	7.91

Table E2. Comparison of taxa across clusters

taxa	cluster	N	Mean	Std Dev	Median	Min	Max	p-value
Abiotrophia	**1**	24	0.00285	0.00840	0.00000	0.00	0.03119	0.99
	2	23	0.00081	0.00222	0.00000	0.00	0.00777	
	3	24	0.20323	0.99440	0.00000	0.00	4.87180	
Corynebacterium	**1**	24	0.10144	0.38993	0.00000	0.00	1.86788	0.22
	2	23	0.60080	2.76487	0.00311	0.00	13.28200	
	3	24	1.19675	5.83692	0.00080	0.00	28.60017	
Enterobacteriaceae	**1**	24	1.00695	3.93918	0.01590	0.00	19.29620	0.64
	2	23	5.73484	19.4992	0.00500	0.00	81.49576	
	3	24	0.96302	4.48167	0.01544	0.00	21.99669	
Enterococcus	**1**	24	0.00085	0.00137	0.00000	0.00	0.00471	0.41
	2	23	0.01136	0.02425	0.00000	0.00	0.08926	
	3	24	0.13404	0.41550	0.00000	0.00	1.56280	
Escherichia/ Shigella	**1**	24	4.13486	20.2086	0.00064	0.00	99.01137	0.49
	2	23	6.47075	22.6525	0.00406	0.00	99.10807	
	3	24	2.65612	12.8679	0.00563	0.00	63.06856	
Gemella	**1**	24	0.12046	0.56009	0.00000	0.00	2.74909	<0.01
	2	23	0.00038	0.00131	0.00000	0.00	0.00622	
	3	24	0.00053	0.00260	0.00000	0.00	0.01274	
Haemophilus	**1**	24	0.00901	0.02127	0.00154	0.00	0.10048	0.44
	2	23	0.45195	1.84714	0.00000	0.00	8.79103	
	3	24	0.00453	0.01118	0.00000	0.00	0.05291	

Klebsiella	**1**	24	4.14592	19.7626	0.00000	0.00	96.89771	0.20
	2	23	0.00069	0.00117	0.00000	0.00	0.00466	
	3	24	0.37079	1.81592	0.00000	0.00	8.89625	
Mycoplasma	**1**	24	0.00027	0.00090	0.00000	0.00	0.00329	0.31
	2	23	0.00007	0.00031	0.00000	0.00	0.00151	
	3	24	0.24640	1.19777	0.00000	0.00	5.86967	
Staphylococcus	**1**	24	66.7361	42.0971	94.7719	0.036	99.79775	0.03
	2	23	62.3131	43.0812	94.7001	0.039	99.95334	
	3	24	89.4290	25.7044	99.2279	14.30	99.95405	
Streptococcus	**1**	24	0.58363	1.98250	0.12582	0.001	9.83025	<0.01
	2	23	0.41666	1.85571	0.01299	0.00	8.92676	
	3	24	0.81536	3.64559	0.00807	0.00	17.90744	
Ureaplasma	**1**	24	22.7775	38.0536	0.03842	0.00	99.72568	0.43
	2	23	23.7972	37.5205	0.47130	0.00	99.83315	
	3	24	3.67098	12.0864	0.03309	0.002	46.52613	

References

[1] Markle JG, Frank DN, Mortin-Toth S, Robertson CE, Feazel LM, Rolle-Kampczyk U, et al. Sex Differences in the Gut Microbiome Drive Hormone-Dependent Regulation of Autoimmunity. Science. 2013.

[2] Hara N, Alkanani AK, Ir D, Robertson CE, Wagner BD, Frank DN, et al. Prevention of Virus-Induced Type 1 Diabetes with Antibiotic Therapy. J Immunol. 2012;189:3805-14.

[3] Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. 1998;8:175-85.

[4] Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. 1998;8:186-94.

[5] Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics. 2011;27:2194-200.

[6] Schloss PD, Westcott SL. Assessing and improving methods used in operational taxonomic unit-based approaches for 16S rRNA gene sequence analysis. Appl EnvironMicrobiol. 2011;77:3219-26.

[7] Pruesse E, Peplies J, Glockner FO. SINA: accurate high throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics. 2012.

[8] Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41:D590-D6.

[9] Breiman L. Random forests. Mach Learn. 2001;45:5-32.