Exploratory analysis of the plasma metagenome reveals no difference between individuals with Myalgic Encephalomyelitis/Chronic Fatigue Syndrome and healthy controls

Ruth R Miller1, Gardy, JL1;2; Parker, SM3; Steiner1; T; Montoya, V; Tang, P;1; Patrick, DM1; for the Complex Chronic Disease (CCD) Study Team
1. University of British Columbia; 2. British Columbia Centre for Disease Control; 3. Centre for Health Evaluation and Outcome Sciences, Providence Health Care

Introduction

• ME/CFS
• Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) is a debilitating disease causing profound fatigue in over 400,000 Canadians.
• The etiology of ME/CFS remains unknown.
• It has long been hypothesized that ME/CFS may have an infectious origin, however, no causal agent has been confirmed.
• Metagenomics
• Metagenomics is the analysis of all nucleic acid recovered directly from a clinical or environmental sample.
• Using next-generation sequencing, metagenomic analysis can be used to identify microbes present in a sample, and their relative abundance, from their DNA sequence.
• The CCD study
• We conducted a metagenomic, matched case-control study to compare the microbial content of plasma in individuals with ME/CFS to healthy controls, as well as a subset of controls with systemic lupus erythematosus (SLE).
• Such an approach may enable the association of novel microbes, not previously hypothesized to have a role in ME/CFS with the disease.

Methods

• Subjects were recruited as described [Patrick et al, 2015], checked for case definitions and excluding diagnoses, and matched across groups.
• This analysis reports on 14 with ME/CFS, 14 healthy controls, and 11 controls with SLE.
• Plasma samples were taken from each participant, and were sequenced at least once on the Illumina HiSeq along with 6 water controls, resulting in mean 13,000,000 read pairs per sample.
• After sequencing, rigorous filtering was performed to remove low quality and duplicate reads, host contamination, and rRNA contamination. A mean of 160,000 reads per sample remained (1%).
• Reads were assigned to bacterial genera using Kraken alignment to the NCBI nr database and viral genera using BLASTn alignment to the NCBI viral database, both followed by taxonomic assignment using MEGAN4.
• To ensure accuracy, BLASTn was used to validate a subset of the genera assignments made by Kraken.
• To aid comparison between samples with different numbers of reads, samples were subset at 61,829 reads, which was the number of reads in the smallest sample number after filtering.
• Multivariable analysis was performed using general linear models in R and P values were corrected for multiple testing using the Bonferroni correction.

Table 1: Genera that differed between study groups with P<0.05 in a multivariable model adjusting for DNA extraction batch and sequencing lane. Columns are % of filtered and subset reads assigned to each genera.

<table>
<thead>
<tr>
<th>Taxonomic Level</th>
<th>ME/CFS</th>
<th>Controls</th>
<th>P-value (Bon)</th>
<th>Adjusted Odds Ratio</th>
<th>95% CI Lower</th>
<th>95% CI Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genus</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Species</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Results

Taxonomic analysis revealed 291 bacterial and 18 viral genera present in at least one sample.

Figure 1: Twenty-five most abundant bacterial genera identified using Kraken and validated using BLASTn. Samples are ordered by hierarchical clustering of genera abundance.

Figure 2: Viral genera identified using BLASTn. Samples are ordered by hierarchical clustering of genera abundance.

As demonstrated in Figures 1 and 2, hierarchical clustering by taxonomic abundance did not cluster samples according to study group (i.e. ME/CFS, SLE or healthy control). However, a batch effect was observed, whereby taxonomic abundance was associated with DNA extraction batch and sequencing lane (Figure 3).

To account for batch-to-batch differences, multivariable regression was performed to investigate the effect of study group on the abundance of each genus, whilst adjusting for DNA extraction batch and sequencing lane. This approach identified three genera to be differentially abundant between study groups (Table 1).

Conclusion

• Our preliminary investigation has been unable to find any microbial associations with ME/CFS in plasma specimens.
• When performing metagenomic analyses, great care must be taken to control for contamination and batch-to-batch differences in both DNA extraction and library preparation, including use of negative controls.
• Our ongoing work will include analyses for further investigation of bacterial content, as well as RNAseq to explore host gene expression.
• Full exploration of ME/CFS may require study of other tissues and sample types, greater efforts to filter and concentrate virus from cellular samples, and exploration of the role of the gut microbiome.

References:
Patrick, DM; Miller, RR; Gardy, JL; et al. [2015] Lyme Disease Diagnosed by Alternative Methods: A Phenotype Similar to That of Chronic Fatigue Syndrome, CID, July 2015

Acknowledgements:
This work was funded by the BCCDC Foundation for Public Health