Visiting Fellowship Report

Visit to the Huttenhower Group, Harvard School of Public Health, 23/3/19-29/3/19

Dr Caroline Young, Wellcome Trust Research Training Fellow, University of Leeds

I am extremely grateful to the Pathological Society of Great Britain & Ireland for the award of a Visiting Fellowship to visit the Huttenhower Group at the Harvard School of Public Health for one week. Here I report on the progress made.

**Background**

I am in the final year of a Wellcome Trust sponsored PhD which is investigating the colorectal cancer (CRC)-associated microbiome. The dataset which we have created is large (1500 samples; hundreds of bacterial taxa). As the microbiome is an emerging field of research, methods of analysis are continually being developed and we lacked the expertise to analyse the data with confidence. I therefore requested to visit the Huttenhower Group, which is a world-leader in the bioinformatic and statistical analysis of microbiome data. Professor Huttenhower is also one of our collaborators on the OPTIMISTICC Grand Challenge. I visited the group with my supervisor, Dr Wood (a bioinformatician).

**Progress**

We presented our data to Professor Huttenhower who outlined a plan for data analysis: ordination, omnibus testing; use of MAASLIN2 analysis; Random Forest modelling and the generation of ROC curves. During the week we were guided by one of the group’s post-doctoral Research Fellows, Dr Thompson. We were provided with examples of existing code and links to relevant papers. We worked on the analysis of our data and met with Dr Thompson daily to work through our questions. We attended both the Huttenhower lab meeting and the Meyerson lab meeting (another world-leader in microbiome research and member of the OPTIMISTICC Grand Challenge) and met with Professor Meyerson separately to outline our data and discuss potential analysis and follow-on work.

**Outcome**

By the end of the week we had learnt how to perform the analysis outlined by Professor Huttenhower. We generated a preliminary Random Forest model to discriminate CRC versus non-CRC cases and were able to show that the area under the curve was improved using microbiome data compared with clinical data alone. We plan to continue this analysis, contacting Dr Thompson or Professor Huttenhower if we have follow-up questions. We have submitted abstracts to present at the Summer Meeting of the Pathological Society of Great Britain & Ireland and the BDIAP and plan to publish the data as soon as the analysis is complete. We will be able to apply these methods of analysis to all subsequent microbiome data which we generate.

The week was a fantastic opportunity to improve our understanding of how to analyse our complex microbiome data. We also forged important collaborations with world-leading researchers in the field of microbiome CRC research. I am extremely grateful to the Pathological Society of Great Britain & Ireland for this invaluable opportunity.