7

GENERAL SUMMARY

Genetically informative study designs have long been of great importance for helping to determine the role the environment and genetic susceptibility play in several human phenotypes. Of such designs, one of the most well-known is the classical twin design, including both monozygotic (MZ) and dizygotic (DZ) participants. In addition to the comparison of MZ and DZ twin pairs, studies focused on MZ twin pairs discordant for a specific phenotype allow for researchers to better understand how environmental influences impact a trait while controlling for the host genomic profile. Beyond the use of twin derived samples, with information on the genomic content of unrelated individuals, it is possible to create genetically informative designs aimed at understanding many biological phenomena.

In **Chapter 2** we study blood and buccal derived DNA samples collected simultaneously in order to investigate telomere repeat mass (TRM) in both tissue types. Performing the telomere measurement in a cohort containing MZ and DZ twin pairs allowed us to determine the contribution of genetic and environmental factors to each of the two blood and the buccal TRM phenotypes. Furthermore, the unique nature of the dataset allowed us to speak to the overall suitability of buccal derived DNA for TRM measurement. The two analyses in the blood samples showed a difference in estimated heritability, with genetic factors explaining 47.6% and 22.2% of total phenotypic variance in TRM in the first and second analysis of the same blood sample. Heritability of buccal TRM was 23.3%. The comparisons showed that, when performed by the same laboratory (AIHG), 11% to 12% of the variance in TRM in blood samples was predicted by TRM in buccal cells, i.e., the correlation was significant but modest ($r = 0.244 - 0.415$). The buccal associated DNA did display the same age and sex associated effects commonly documented in measurements based on blood-derived DNA, indicating that the TRM data is showing the expected male disadvantage and telomere attrition with age. **Chapter 2** provides support, albeit modest, for the use of buccal derived DNA in large scale biobank studies focused on the study of TRM, which will allow for less invasive longitudinal studies. It is also clear that repeated handling is ill-advised for TRM measurement for either buccal or blood samples.

Chapter 3 focuses on using multiple genetically informative designs to understand the relationship between the gut microbiota and host obesity-associated measures. These studies include the use of samples collected from body mass index (BMI) discordant MZ twin pairs and the four-corners study design, with unrelated individuals at high/low extremes for both genetic risk for obesity and actual BMI. **Chapter 3** provides support for a distinctly lower gut microbiota diversity in individuals with high BMI, particularly if they are low in the genetic susceptibility to obesity. Additionally, we identified a number of OTUs in the *Ruminococcaceae* and *Oxalobacteraceae* families that have a significant association with obesity-associated measures and their depletion may be a source of the lower microbiota diversity seen in more obese individuals.

Chapter 4 explored the relationship between variation in the host genome and the subsequent gut microbiome composition. As a consortium, we recently produced the largest GWA on human gut microbiome composition too date. It identified 30 loci affecting ten microbiome taxa at a genome-wide significant ($P<5x10^{-8}$) threshold, with one locus, the lactase (LCT) gene region, remaining study-wide significant after additional correcting for the number of taxa tested (GWAS signal $P= 8.6x10^{-21}$). **Chapter 4** describes results generated during the work needed to contribute to this GWA meta-analysis by the international consortium. While our main aim was to provide the NTR results for the meta-analysis, in parallel, we tested a few specific hypotheses using the genome-wide data in the NTR sample. The consortium identified the association between the *Bifidobacterium* genus and rs182549 as the most significant microbiome quantitative trait locus (mbQTL) ($p = 8.6332 \times 10^{-21}$) and the NTR-specific results for this mbQTL showed a p-value of 0.06, only just failing to achieve nominal significance. A more convincing resemblance between the NTR cohort and the consortium findings was found with regards to the link between heritability of a taxon and its ability to generate significant associations. Taxa defined at the family and genus levels that had the highest MZ twin correlations (>0.30) showed much lower p-values in their associations with genome-wide SNPs than less heritable taxa. Taken together, it seems appropriate to use the results of the MiBioGen consortium for extended analyses in smaller cohorts, such as performing MR with genetic instruments for the more heritable microbiota as predictors of inflammatory and cardiometabolic (including obesity) outcomes.

Chapter 5 was concerned with understanding how environmental factors, in the form of cohabitation, impact the gut microbiome composition. Samples collected from MZ twin pairs who cohabitate and those that do not, in addition to spouse pairs, allows a unique view into the role of cohabitation as an important environmental modulator of the gut microbiota. The work in **Chapter 5** identified two species-level OTUs (Otu0081 and Otu0190) that were significantly shared between cohabitating MZ twin pairs and spouse pairs but not between non-cohabitating MZ twins. The lack of sharing between the non-cohabitating MZ twins rules out genetic influences on these microbes. Microbes that are particularly influenced by cohabitation were further explored to determine relationships between host cardiometabolic and disease burden. Disease burden scores were regressed on the OTU counts of the two species-level OTUs shared due to cohabitation. One of these OTUs (OTU190) was significantly associated with lower inflammatory and cardiometabolic disease burden.

Considering the ultimate contributions of nature and nurture in dictating the gut microbiome composition, it is clear that both of these factors play a non-trivial role, but meaningful patterns are still largely yet to be deduced. During our discussion of these factors in **Chapter 6**, we aimed to reorient the microbiome field with regard to the discussion surrounding heritability of individual microbiome components. As

such, we define a microbiome constituent as heritable if the host provides, by means of its genetic architecture, conditions amenable to colonization by a particular microbe upon encounter. The key part of this definition relies upon environmental exposure to a microbe of interest in order for the heritable nature of the organism to be observed. What has only become clearer throughout the development of this thesis is that host genetics, host environmental factors, the gut microbiome composition and host health are deeply intertwined factors.