EXPLORING THE INTERACTION OF HOST GENETICS AND THE GUT MICROBIOME IN OBESITY IN AN AFRICAN POPULATION

Samantha Susan Schnell

A **research report (in the format of a "submissible" paper)** submitted to the Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, in partial fulfilment of the requirements for the degree of **Master of Science in Medicine in the field of Genomic Medicine**.

Johannesburg, 2022.

NATIONAL HEALTH LABORATORY SERVICE
School of Pathology, University of the Witwatersrand
DIVISION OF HUMAN GENETICS
Hospital Street, Johannesburg, 2001 I PO Box 1038, Johannesburg, 2000
[T] +27 11 489 9223 I [M] +27 78 080 8841 I [F] +27 11 489 9226 I [E] human.genetics@nhls.ac.za

22 November 2022

Dear Examiner,

RE: A RESEARCH REPORT SUBMITTED TO THE FACULTY OF HEALTH SCIENCES BY SAMANTHA SCHNELL IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN MEDICINE IN GENOMIC MEDICINE

Thank you for agreeing to examine the following research report.

This research report, entitled "Exploring the Interactions of Host Genetics and the Gut Microbiome in Obesity", is being submitted by the candidate, Samantha Schnell, in partial fulfilment of the degree, Master of Science in Medicine in Genomic Medicine. The research report is being submitted in the form of a "submissible" paper and contributes towards 30% of the degree.

The candidate, with approval from her supervisors, has selected the peer-reviewed journal, *Frontiers in Genetics*, for submission to the Faculty of Health Sciences. *Frontiers in Genetics* was selected as it is one of the most cited journals in the field of genetics and heredity and has a good impact factor, 4.274. The research report was written in accordance with the *Frontiers* Author Guidelines, which has been attached for ease of reference (see Appendix D). The *article* type selected was the format of original research. The *Frontiers* Author Guidelines state that figures must be individually submitted in the order in which they are referred to in the manuscript, and that tables should be inserted at the end of the manuscript. However, for ease of reference and for marking purposes these guidelines have been deviated from. The tables and figures will appear in text in the order in which they are referred to, as opposed to being submitted separately.

In Appendix A, the approved research protocol has been attached to serve as an extended literature review and to provide further context for the submitted research. The research protocol is not being submitted for examination as it was previously assessed and approved by the Faculty of Health Sciences.

Yours sincerely,

Samantha Schnell
Candidate

Dr Ovokeraye Oduaran
Supervisor

Scott Hazelhurst (Dec 12, 2022 13:41 GMT+2)

Prof Scott Hazelhurst
Co-Supervisor

Declaration

I, **Samantha Schnell**, declare that this Research Report is my own, unaided work. It is being submitted for the Degree of **Master of Science in Medicine in Genomic Medicine** at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.

Signature of Candidate

27th day of February 2023, in Johannesburg

Contribution of Candidate to the Paper

Declaration: Student's contribution to article(s) and agreement of co-author(s)

I, **Samantha Susan Schnell**, student number **1914985**, declare that this Research Report is my own work and that I contributed adequately towards research findings presented in the paper stated below, intended for publication.

Signature of Student: 

Date: **22/11/2022**

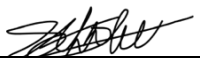
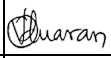
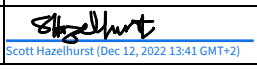
Name of Primary Supervisor: **Dr Ovokeraye Oduaran**

Signature of Primary Supervisor: 

Date: Dec 12, 2022

Agreement by co-authors: By signing this declaration, the co-authors listed below agree to the use of the article by the student as part of her Research Report.

Article Title: Exploring the Interaction of Host Genetics and the Gut Microbiome in Obesity

Authors	Name	Signature	Date
1st author	Samantha Schnell		22/11/2022
2nd author	Dr Ovokeraye Oduaran		Dec 12, 2022
3rd author	Prof Scott Hazelhurst	 Scott Hazelhurst (Dec 12, 2022 13:41 GMT+2)	Dec 12, 2022

This paper is dedicated to my parents, Sharon and Bruce Schnell, and to my sisters, Robyn and Maxine.

Abstract

Obesity is a highly prevalent health concern that is on the rise in Sub-Saharan Africa. Even though genetic variation and gut microbiota have been implicated in the development of obesity independently, the interactions between these factors have not been previously explored in a South African cohort. This study aimed to identify possible associations between host genomes, body mass index as a measure of obesity, and gut microbiota composition (in the form of V3-V4 16S rRNA sequencing) in a female African cohort. Polygenic risk scores predictive of body mass index in this cohort were generated to categorise the participants into high- and low-risk groups. Subsequently, several statistical analyses were performed comparing gut microbiota between these groups. High-risk participants with high body mass index had associations identified with increased abundances of *Prevotella_9* and *VadinBE97*. In contrast, the low polygenic risk and low body mass index sub-group was associated with greater *Bacteroides* levels. This study acknowledges the plausible interactions between these factors in an African cohort.

Acknowledgements

Above all, I would like to thank my supervisors, Dr Ovokeraye Oduaran and Prof Scott Hazelhurst. Thank you for the weekly check-ins and support over the last year. Thank you both for taking time out of your busy schedules to guide and support me, specifically through navigating the challenges of coding. Your patience with my use of the command line never went unnoticed. I appreciate it greatly.

Thank you to Dr Tinashe Chikowore for your invaluable help and guidance on the development and application of the polygenic risk scores.

Thank you to Dr Zané Lombard, Dr Robyn Kerr, and Monica Araujo for your words of encouragement and motivation throughout the year. Thank you for always having your doors open to me when I was seeking help or advice.

Thank you to the University of the Witwatersrand for funding my studies through a Postgraduate Merit Award.

Finally, a thank you to all the participants enrolled in the AWI-Gen and microbiome pilot studies. This research would not have been possible without your contributions.

Contents

List of Abbreviations and Symbols

α	alpha
APCDR	African Partnership for Chronic Disease Research
ASV	amplicon sequence variant
AWI-Gen	Africa Wits-INDEPTH Partnership for Genomic Research
BMI	body mass index
FTO	*fat mass and obesity-associated*
GWAS	genome-wide association study
H3Africa	Human Heredity and Health in Africa
HIV	human immunodeficiency virus
HR	hazard ratio
ID	identification
LDA	linear discriminant analysis
LEfSe	linear discriminant analysis effect size
LYPLAL1	*lysophospholipase-like 1*
MAMA	multi-ancestry meta-analysis
MC4R	*melanocortin 4 receptor*
OR	odds ratio
PAGE	Population Architecture using Genomics and Epidemiology
PCA	principal component analyses
PCoA	principal coordinates analysis
PERMANOVA	permutational multivariate analysis of variance
PGS	Polygenic Score
PLD1	*phospholipase D1*
QC	quality control
PRS	polygenic risk score
R^2	coefficient of determination
rRNA	ribosomal ribonucleic acid
SNP	single nucleotide polymorphism
UKBB	United Kingdom Biobank

List of Figures

List of Tables

A research report in the format of a “submissible” paper.

This work has been written up as a paper for submission to **Frontiers in Genetics**. The author guidelines adhered to can be found in **Appendix D**, which have been included for ease of reference.

Exploring the Interaction of Host Genetics and the Gut Microbiome in Obesity in an African Population

Samantha Schnell[1], Ovokeraye Oduaran[2], Scott Hazelhurst[2,3*]

[1]Division of Human Genetics, School of Pathology, Faculty of Health Science, University of the Witwatersrand and the National Health Laboratory Services, Johannesburg, South Africa
[2]Sydney Brenner Institute for Molecular Bioscience, University of the Witwatersrand, Johannesburg, South Africa
[3]School of Electrical and Information Engineering, University of the Witwatersrand, Johannesburg, South Africa

***Correspondence:**
Scott Hazelhurst
scott.hazelhurst@wits.ac.za

Number of words: 8,798 Number of figures: 5 Number of tables: 6

Abstract

Obesity is a highly prevalent health concern that is on the rise in Sub-Saharan Africa. Even though genetic variation and gut microbiota have been implicated in the development of obesity independently, the interactions between these factors have not been previously explored in a South African context. Therefore, this study aimed to identify possible associations between host genomes, body mass index as a measure of obesity, and gut microbiota composition in a female African cohort. Polygenic risk scores predictive of body mass index in this cohort were generated to categorise the participants into high- and low-risk groups. Subsequently, several statistical analyses were performed to compare gut microbiota and microbial abundances between the high- and low-risk groups. High-risk participants with high body mass index had a significant association identified with abundances of *Prevotella_9* and a moderate association with *VadinBE97*. Conversely, the low polygenic risk, low body mass index sub-group was moderately associated with greater levels of *Bacteroides*. This study acknowledges the plausible interactions between these factors in an African cohort.

Keywords: African, obesity, host genetics, polygenic risk score, gut microbiota

1 Introduction

Obesity is a highly prevalent public health concern and gives rise to a wide variety of non-communicable diseases. Even though the incidence of obesity appears to be declining in many regions worldwide, in low- and middle-income countries, such as South Africa, obesity continues to increase (Loos and Yeo, 2022). Consequently, this has led to a rise in mortality and morbidity. Although the aetiology of obesity is poorly understood, multiple driving factors that play pivotal roles in the development of this complex disease have been identified. Genetic and environmental influences are two such factors. However, these two factors do not entirely account for the obesity phenotype. In more recent studies, gut microbial composition has also been associated with the aetiology of obesity, specifically abundances of certain bacterial taxa and imbalances in the gut microbiome, "dysbiosis" (Cuevas-Sierra *et al.*, 2019). Even though these driving factors have been implicated in the development of obesity, significant challenges remain in deciphering the interactions between them. This includes interactions between host genetics and the gut microbiome. Previously, it has been illustrated that the gut microbiome tends to be influenced by host genetics in humans (Xu *et al.*, 2020). However, the interaction between host genetics and gut microbiota is a topic that has not been extensively explored in obesity. Furthermore, the composition of gut microbiota in those genetically predisposed to obesity is widely understudied and has not previously been investigated in African contexts.

Evidence has determined that obesity is partly heritable by an estimated 40% to 70% (Thaker, 2017; Loos and Yeo, 2022). However, genetic predispositions to obesity are often affected by lifestyle and behaviour, such as diet, sleep patterns, physical activity, etc. The common form of obesity is complex and polygenic in nature. Genome-wide association studies (GWASs) have demonstrated that there are hundreds of risk loci associated with obesity. Therefore, obesity may arise through an accumulation of many single nucleotide polymorphisms (SNPs), each possessing small, weighted effects. However, in obese individuals of African ancestry, significant associations have been identified in *fat mass and obesity-associated* (*FTO*) and *melanocortin 4 receptor* (*MC4R*) genes, as reviewed by Yako *et al.* (2015). Furthermore, the *FTO* gene remains the locus that possesses the strongest GWAS signal across multiple ancestries (Goodarzi, 2018). Genetic contributions to obesity have been widely investigated through GWASs of traits associated with body composition, the most common being body mass index (BMI). Additionally, studies have suggested that genetic variation accounts for 30% of the variance observed in BMI (Goodarzi, 2018). Body mass index is often used at a populational-level as an anthropometric measure for obesity, as it serves as a proxy and is easily

measurable. Due to the polygenic nature of obesity, genetic aetiologies are best assessed with polygenic risk scores (PRSs) using GWAS summary statistics for obesity-related traits, such as BMI. Polygenic risk scores allow for the integration of many SNPs into a single quantitative measure to account for the genetic liability of complex diseases (Khera *et al.*, 2019).

Even though the role of host genetics in shaping gut microbiota composition is not well known, it is generally hypothesised that factors of human physiology are genetics-dependent and may impact the gut microbiome (Dabrowska and Witkiewicz, 2016). These factors include, but are not limited to, metabolism, gut motility and energy regulation. Moreover, the gut microbiome acts as an environmental factor, forming a dynamic ecosystem within the host. Since bacteria maintain functions in energy harvesting and metabolism, gut microbiomes have been associated with regulating body weight and influencing the risk of developing obesity. Consequently, microbial communities may adversely affect human health (Dabrowska & Witkiewicz, 2016). Through advances in metagenomics and high-throughput next-generation sequencing, the phylogenetic composition of gut microbiota can be assessed to determine functional potential and diversity (Dugas *et al.*, 2016).

Recent evidence has identified associations between SNPs and gut microbiota in obesity (Cuevas-Sierra *et al.*, 2019). Through a GWAS, the rs4894707 variant near *phospholipase D1* (*PLD1*) has been associated with the *Akkermansia* genus *(Verrucomicrobiota* phylum) (Davenport *et al.*, 2015). Furthermore, the *PLD1* gene has been linked to BMI and *Akkermansia* was previously regarded as an obesity-related bacterial taxon. Similarly, abundances of the *Prevotella* genus (*Bacteroidota* phylum) have been significantly associated with rs878394 (Li *et al.*, 2018). This variant has also been associated with *lysophospholipase-like 1* (*LYPLAL1*), a gene related to insulin sensitivity and body fat distribution. Furthermore, associations have been identified between PRSs and gut microbiota. In a Spanish cohort, women with high genetic susceptibility to obesity and high BMI were seen to possess greater abundances of the *Prevotellaceae* family. Additionally, those with high polygenic risk but lower BMI had lower abundances of the same family (Cuevas-Sierra *et al.*, 2020). In contrast, women with low polygenic risk possessed increased levels of *Barnesiella* and *Alistipes* genera (*Bacteroidota* phylum) and *Peptostreptococcaceae* (*Firmicutes* phylum).

The Human Heredity and Health in Africa (H3Africa) initiative includes the Africa Wits-INDEPTH Partnership for Genomic Research (AWI-Gen) Collaborative Centre, in which this study is nested. This initiative is a combined effort to investigate genetic and environmental

aspects of human health in African individuals, including body composition and fat distribution (Ramsay *et al.*, 2016). This study cohort comprises a subset of the AWI-Gen phase I South African study population sub-sampling from the Agincourt (n =119) and Soweto (n = 50) cohorts. Research units of the University of the Witwatersrand manage both cohorts. Host genotype data was obtained from the 12,035 individuals genotyped in AWI-Gen phase I, with ~1.7 million SNPs genotyped using Illumina's Infinium™ H3Africa SNP Genotyping Array v1 (Mulder *et al*., 2018). Gut microbiome sequencing data (the V3-V4 hypervariable regions of the 16S ribosomal RNA (rRNA) gene) using Illumina's MiSeq® platform was performed on a smaller subset of the AWI-Gen phase I cohort. The samples from these participants (n = 169) were processed for both SNP array genotyping (host genetics) as well as 16S rRNA sequencing (microbiome composition) and made up the study cohort.

The present study aimed to identify possible associations between host genomes, using a PRS developed for BMI as a proxy for genetic association, BMI as an anthropometric measure for obesity, and gut microbiota composition in a female African cohort. Additionally, this study aimed to provide further insight into the association between factors influencing obesity in African populations. Notably, there is a paucity of genetic and gut microbiome studies in an African context.

2 Materials and Methods

2.1 Study Participants

The 169 study participants were recruited from two areas within South Africa. The majority were recruited from Agincourt, Bushbuckridge, in the Mpumalanga province. Bushbuckridge is a municipality with a mixture of rural and peri-urban settlements. Other participants were from an urban settlement in Soweto, Johannesburg, in the Gauteng province. Subsequently, participants were categorised based on their recorded BMI (kg/m^2) into lean ($< 25.0\ kg/m^2$), overweight ($25.0 - 29.9\ kg/m^2$) and obese ($\geq 30.0\ kg/m^2$) categories. To reduce confounding effects, only females who tested negative for the human immunodeficiency virus (HIV−) were included. This study was approved by the Human Research Ethics Committee (Medical) of the University of the Witwatersrand as an AWI-Gen sub-study (M220297), see Appendix B.

2.2 AWI-Gen Phase I Genotype Data

2.2.1 Polygenic Risk Score Generation

PRSice (v2.2.13) software (Choi and O'Reilly, 2019) was used to calculate PRS using BMI-

associated SNPs as a proxy for host genetic contribution to obesity. PRSice can calculate PRSs and apply, analyse and plot the calculated PRS values (Euesden, Lewis and O'Reilly, 2015). PRSice calculated the best-fit PRS by calculating *p*-values at several different thresholds. The subset of microbiome participants (n = 169) were extracted from the AWI-Gen target genotype file to prevent overfitting. The BMI (kg/m^2) phenotypes were included as a proxy for obesity and specified as a non-binary trait. For the PRS computation, PRSice included a full model at a threshold of $p = 1$. The PRS was calculated using a direct summation of effect sizes of SNPs and no clumping (removing weakly correlated SNPs based on linkage disequilibrium and *p*-values) was performed. In addition, models were adjusted for sex, age and the first five principal component analyses (PCAs) to account for population structure, all of which were included as covariate information.

The PRS calculations were computed using GWAS summary statistics (base datasets comprising SNPs, *p*-values and weightings) from three different studies (Table 1). PGS001228 and PGS001943 base datasets were accessed through an open database, the Polygenic Score (PGS) Catalog (Lambert *et al*., 2021), and the multi-ancestry meta-analysis (MAMA) dataset was provided by Chikowore, *et al.* (2022), received via private communication. The PRS most predictive of BMI in the cohort, MAMA, was selected for downstream analysis.

Table 1. Comparison of the three base datasets used for PRS calculations

	PGS001228	**PGS001943**	**MAMA**
Study	Tanigawa *et al*., 2022	Privé *et al*., 2022	Chikowore, *et al*., 2022
PGS Catalog Publication ID	PGP000244	PGP000263	N/A
PGS Catalog Name	GBE_INI21001	Portability-PLR_log_BMI	N/A
Discovery Dataset	268,856 European individuals	391,124 European individuals	678,661 multi-ancestry individuals
Target Dataset	12,035 AWI-Gen cohort	12,035 AWI-Gen cohort	12,035 AWI-Gen cohort
Replication Dataset	149 study participants	149 study participants	149 study participants
Number of SNPs used for PRS Calculations	27,126	110,153	322,639
Effect Weight Type	Log (OR) or log (HR)	Beta	Beta
***P*-value Threshold**	5×10^{-8}	5×10^{-8}	5×10^{-8}

*ID = identification; OR = odds ratio; HR = hazard ratio

2.2.2 Separation of Participants into Deciles to Form Cases and Controls

After PRS generation, 20 participants were excluded from the analysis due to insufficient covariate information, leaving a cohort of 149 individuals. Using the PRS most predictive of

BMI in the remaining study cohort, participants were separated into deciles (n = 15) to form groups of high and low polygenic risk for further downstream analysis. The first, second and third deciles (n = 45) formed the low polygenic control risk group. The eighth, ninth and tenth deciles (n = 44) were the high PRS case group with greater genetic predisposition to developing obesity. The remaining fourth, fifth, sixth and seventh deciles (categorised as other) were not included in the downstream analysis.

2.3 AWI-Gen Microbiome Pilot Study Data

2.3.1 Data Processing and Quality Control

For microbial data processing and quality control (QC), V3-V4 hypervariable sequences of the 16S rRNA gene were assessed using DADA2 v1.14 (Callahan *et al*., 2016). All analyses were performed using R statistical software (v3.6.0; R Core Team (2021)). The demultiplexed paired-end sequences for the 149 samples included in the PRS generation were imported into DADA2. Read quality profiles were inspected, and sequences were filtered and trimmed. Filtering included truncating read lengths of 240 and 280 nucleotide bases, maximum expected errors of 2 and 4 with minimum lengths of included sequences being 175. Post-filtering, identical reads were dereplicated and sample inferences were applied to denoised paired-end sequences, which were subsequently merged. Finally, an amplicon sequence variant (ASV) table containing sequence variants identified in individual samples and a non-chimeric sequence table were generated. The output data was exported for downstream analyses.

2.3.2 Taxonomic Assignment

The DADA2 package implemented the naïve Bayesian classifier method (Wang *et al.*, 2007) to sequence variants for the classification of taxonomies. The naïve Bayesian classifier method can accurately classify 16S rRNA sequences into higher-order taxonomic classes from domain to genus levels. For taxonomic classification, DADA2 utilised formatted training FASTA files derived from the SILVA rRNA data project (v138.1) and for species-level assignments, the SILVA species assignment (v138.1) was applied.

2.3.3 Rarefaction Analysis and Richness Estimates

The output of DADA2, taxonomic assignments and a metadata file containing the PRSs, and decile separation of participants was analysed using the phyloseq package (v1.30.0) in R (McMurdie and Holmes, 2013). Samples with a read depth < 34,311 were excluded for adequate capture of observed richness to be assessed with rarefaction. Therefore, from the 149 samples, three were removed for the alpha diversity analysis. Of the 146 samples remaining, samples

classified as other (neither cases nor controls) were excluded to allow for a comparison of alpha diversities between high and low polygenic risk groups. Eighty-eight samples remained, 44 high and 44 low PRS (one low PRS sample had a read depth < 34,311), and the rarefied dataset was utilised for the alpha diversity analysis.

2.3.4 Alpha Diversity Analysis

Alpha diversity metrics represent the diversity of microbial taxa within a sample, showing the richness (the number of taxa present) and evenness (the distribution of abundances). The observed, Chao1 (Chao, 1984) and Shannon (Shannon, 1948) alpha diversity measure estimates were calculated. Estimates were used to analyse qualitative measures of richness, estimations of richness and quantitative measures of richness, respectively. The pairwise Wilcoxon rank sum (Mann Whitney U) test was calculated to compare alpha diversity measures between high and low PRS groups using *p*-values to determine statistical significance. Subsequently, alpha diversity estimates, comparing the high and low polygenic risk groups, were visualised with boxplots generated in R (v3.6.0; R Core Team (2021)).

2.3.5 Beta Diversity Analysis

Beta diversity, a form of composition analysis, measures the community differences between samples. The Bray-Curtis dissimilarity distance matrix, a quantitative measure of dissimilarity, was the beta diversity measure estimate used. It was assessed through a principal coordinates analysis (PCoA) (Gower, 2005). The statistical significance between the high (n = 44) and low (n = 45) PRS groups was assessed using the permutational multivariate analysis of variance (PERMANOVA) package in R (Anderson, 2017).

2.3.6 Differential Analysis of Taxonomic Abundances

The phyloseq package in R (McMurdie and Holmes, 2013) was used for taxonomic analysis with taxa plots generated for phyla abundances and the top 15 most abundant genera between high and low PRS groups. To further explore relationships between PRS, BMI and the gut microbiome, the high and low polygenic risk groups were subdivided based on BMI. Those with lean BMI (< 25.0 kg/m^2) were classified as having low BMI, while overweight/obese participants (≥ 25.0 kg/m^2) were of high BMI. The 15 most abundant taxonomic genera were assessed between the four sub-groups: high PRS/high BMI (High/High), high PRS/low BMI (High/Low), low PRS/high BMI (Low/High) and low PRS/low BMI (Low/Low).

The DESeq2 package (v1.26.0) in R (Love, Huber and Anders, 2014) was utilised to further

evaluate differences in abundances of taxa between the high and low polygenic risk groups. The statistical significances were calculated using negative binomial generalised linear models with raw RNA sequencing counts. The Wald statistical test was then used to determine *p*-values from the regression models produced. Volcano plots were generated using DESeq2 at different alpha thresholds ($\alpha = 0.05$, $\alpha = 0.1$ and $\alpha = 1$).

The Huttenhower Lab's Galaxy application, linear discriminant analysis (LDA) effect size (Segata *et al*., 2011), was also used for differential analysis. Linear discriminant analysis effect size (LEfSe) assessed gut microbiota that differed between the high and low PRS groups (n = 89) and high and low BMI groups within high and low PRS groups. Linear discriminant analysis was calculated with factorial Kruskal-Wallis tests at $\alpha = 0.05$ and a logarithmic LDA score threshold of 2.0 to identify discriminative features between the groups.

2.3.7 Assessing the Interactions Between Polygenic Risk, Body Mass Index and Gut Microbiota Composition

Multiple linear regression models were generated with BMI as the dependent variable, and PRS and abundances of taxa of interest as independent variables (BMI ~ PRS + abundance of taxa of interest). Taxa of interest were identified through gut microbial analysis. Additionally, linear regression models were performed to determine how much of the variability in taxonomic abundance can be explained by PRS alone (taxa of interest ~ PRS). This was performed with taxa of interest as the dependent variable and PRS as the independent variable. Statistical analysis of the regression models was undertaken using dplyr v1.0.8 (Wickham, *et al*., 2022).

3 Results

3.1 Study Participants

Initially, the study cohort comprised 169 participants. However, during PRS calculations, 20 participants were excluded by PRSice software because of invalid phenotype information (insufficient covariant data). This exclusion left a study cohort of 149 for further downstream analyses. Of the remaining study participants (111 from Agincourt and 38 from Soweto), 24.83% were lean (including normal weight), 17.45% were overweight and 57.72% were obese. After PRS calculations, the cohort was separated into groups of high PRSs (cases) and low PRSs (controls) and other (intermediates not classified as case or control). Only high and low PRS groups (n = 89) were used for microbiome analyses. The distribution of the cohort and groups, separated based on PRS, are shown in Table 2, while BMI distributions are in Table 3.

Table 2. Distributions of study participants in the cohort and high PRS, low PRS and other groups

	Number of Participants (n)	Location		BMI Group		
		Soweto	Agincourt	Lean (< 25.0 kg/m^2)	Overweight (25.0 - 29.9 kg/m^2)	Obese (≥ 30.0 kg/m^2)
Cohort	149	38	111	37	26	86
High PRS	44	11	33	7	5	32
Low PRS	45	11	34	17	11	17
Other	60	16	44	13	10	37

Table 3. Body mass index distributions of the cohort and high PRS, low PRS and other groups

	BMI (kg/m^2)		
	Mean ± SD	Median	Range
Cohort	32.2 ± 8.2	31.8	19.5 – 64.4
High PRS	34.6 ± 8.2	34.7	20.6 – 52.7
Low PRS	29.3 ± 6.5	27.0	19.5 – 42.9
Other	32.8 ± 8.5	31.9	20.6 – 64.4

3.2 Polygenic Risk Score

GWAS summary statistics PGS001228, PGS01943 and MAMA were used to calculate PRSs for the study cohort. The resultant PRS parameters provided by PRSice are shown in Table 4. The R^2 values produced to determine the measure of predictivity were full R^2 (BMI ~ PRS + age + sex + PCAs), PRS R^2 (full R^2 – null R^2) and null R^2 (BMI ~ age + sex + PCAs). The PRS R^2 determined how much of the measured BMI phenotype can be explained by host genetics based on the SNPs included in each PRS calculation. MAMA produced the highest PRS R^2, R^2 = 0.0665. This means that approximately 6.65% of the variability in the BMI phenotype can be explained by host genetics, using the PRS as a proxy. PGS001228 produced a PRS R^2 = 0.0265 (2.65%) and PGS001943 produced a PRS R^2 = 0.0067 (0.67%). Since the PGS01943 PRS R^2 < 0.01, it did not explain the phenotypic variability in this cohort well and was a weak proxy for the genetic liability. PGS001228 and MAMA calculations produced results of statistical significance, p = 0.0475 and p = 0.0014, respectively. $p < 0.05$ was regarded as statistically significant. Therefore, PRSs using the summary statistics from Tanigawa *et al*. (2022) and Chikowore *et al*. reliably predicted BMI in the cohort. Conversely, PGS001943 summary statistics obtained from Privé *et al*., (2022) did not produce a statistically significant PRS (p = 0.3236) and was unreliable in predicting BMI. Taking both the PRS R^2 and the p-value into account (Table 4, Figure 1), MAMA was most predictive and selected for decile separation and high and low PRS assignment to be used for downstream microbiome analysis.

Table 4. Comparison of PRSice results for PGS001228, PGS001943 and MAMA calculations

	PGS001228	PGS001943	MAMA
Full R^2	0.0690	0.0491	0.1167
PRS R^2	0.0265	0.0067	0.0665
Null R^2	0.0425	0.0425	0.0502
Coefficient	1.1534	13.9984	0.1628
Standard Error	0.5772	14.1339	0.0501
***p*-value**	0.0475	0.3236	0.0014

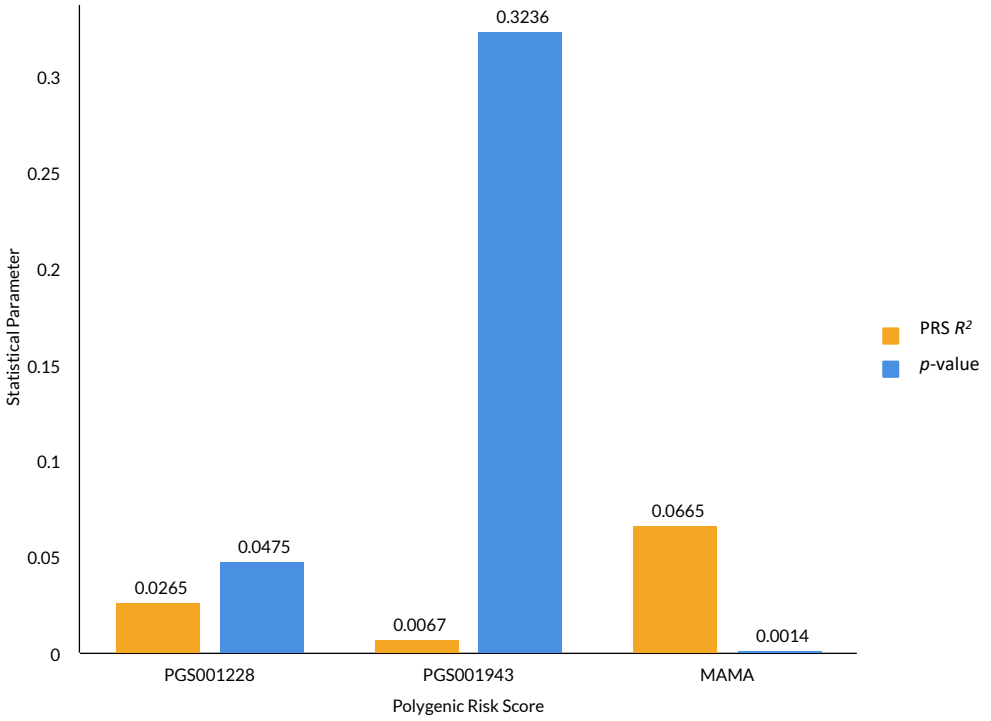


Figure 1. Comparison of PGS001228, PGS001943 and MAMA PRS R^2 coefficients and *p*-values

3.3 Taxonomic Analyses

Taxa were classified into seven taxonomic ranks (kingdom, phylum, class, order, family, genus and species) by ASVs and taxonomies of two kingdoms, *Archaea* and *Bacteria*, were identified. However, some taxa had not yet been classified at genus or species level. This is possibly due to African gut microbiomes being significantly underrepresented in microbiome studies, and consequently, in microbiome reference databases (Allali *et al.*, 2021).

3.4 Alpha and Beta Diversity Measure Estimates

A comparison of the alpha diversity measure estimates (Figure 2A) shows that differences in microbial richness were present between participants in the high and low polygenic risk groups. The alpha diversity measure estimates showed that the low polygenic risk group had greater alpha diversity. The greatest difference in alpha diversity measures between those with high

and low PRSs was seen with the Shannon diversity index (p = 0.13). Since Shannon diversity was higher in those with low polygenic risk, it suggests that there is greater species diversity within low polygenic risk participants. However, differences were not statistically significant (p > 0.05) when analysed with the non-parametric pairwise Wilcoxon rank sum test (Table 5). Therefore, there was a lack of association between alpha diversity measures (observed, Chao1 and Shannon) and PRSs.

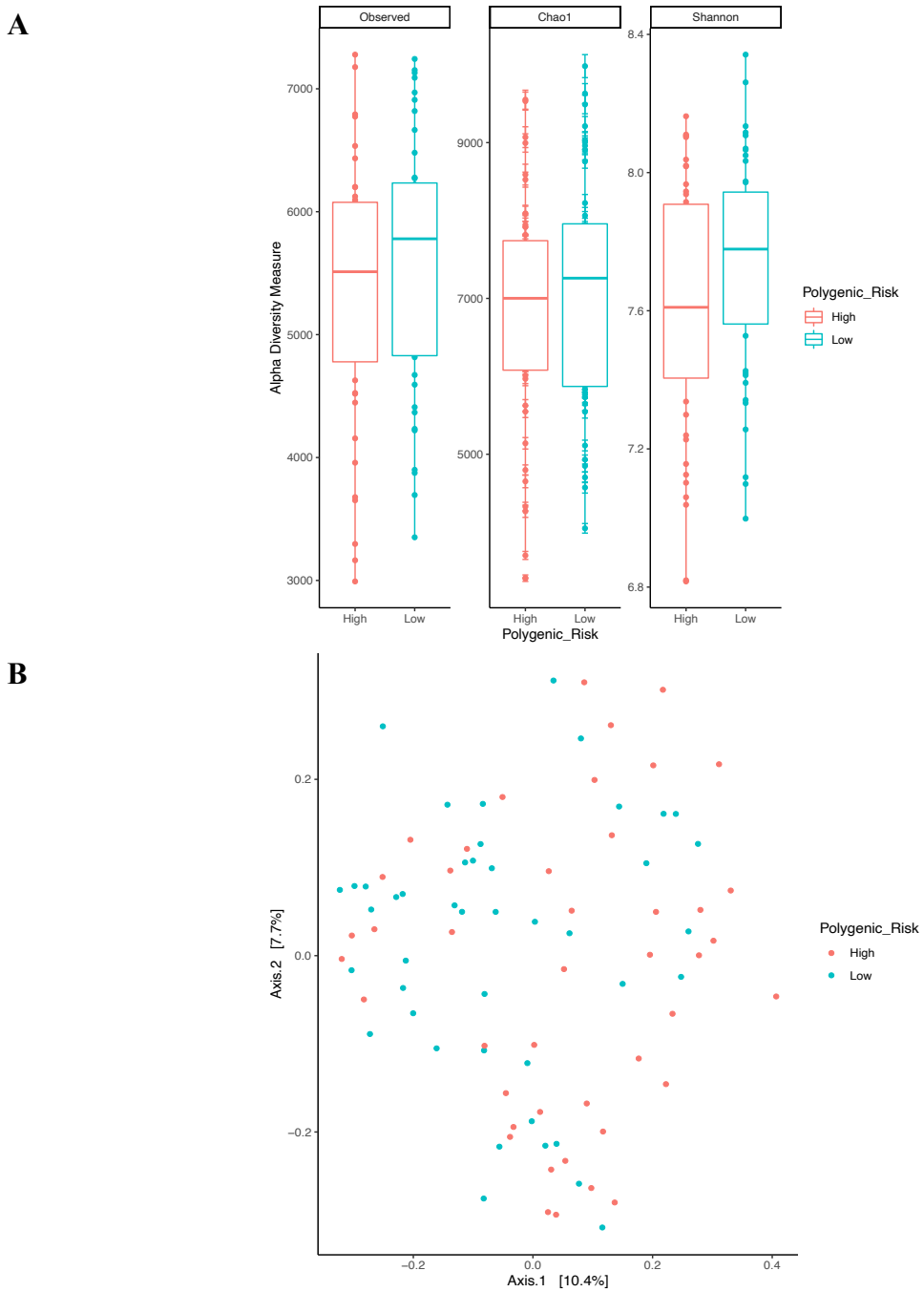


Figure 2. Alpha and beta diversity measure estimates between high and low PRS groups using (**A**) observed, Chao1 and Shannon alpha diversity measure estimates and (**B**) Bray Curtis dissimilarity distances for beta diversity

Table 5. Statistical significance (*p*-values) of alpha and beta diversity indicators in high versus low PRS groups

Comparison	Alpha Diversity *p*-values			Beta Diversity *p*-values
	Observed *p*-value	Chao1 *p*-value	Shannon *p*-value	Bray-Curtis *p*-value
High PRS versus Low PRS	0.29	0.57	0.13	0.079

Bray-Curtis dissimilarity was used to quantify the dissimilarity between species inhabiting the gut microbiomes of those of high and low polygenic risk. This was visualised using PCoA plots (illustrated in Figure 2B) and statistical significance was calculated with PERMANOVA (p = 0.079), see Table 5. The differences observed in the PCoA plots between the high and low PRS groups are suggestive of community differences between these groups. However, with the *p*-value not being statistically significant ($p > 0.05$), definitive conclusions cannot be drawn on associations between beta diversity and polygenic risk.

3.5 Differential Taxa Abundances between High and Low Polygenic Risk Groups

Overall, there were 17 unique phyla identified in the study cohort. The dominant phyla across the cohort were the *Firmicutes* phylum with the greatest abundance, followed by *Bacteroidota* (also referred to as *Bacteroidetes*) and *Proteobacteria*. The differences in abundances at the phylum level show participants with high polygenic risk to have greater abundances of *Proteobacteria* and lesser abundances of *Bacteroidota* compared to their low-risk counterparts (Figure 3). Other phyla with much lower abundances include *Cyanobacteria*, which had increased abundances in the high polygenic risk group, and *Actinobacteriota* with greater abundances in the low-risk participants.

Through the use of DESeq2, differences were seen in abundances of the *Bacteroidota*, *Cyanobacteria*, *Desulfobacterota*, *Firmicutes*, *Proteobacteria* and *Verrucomicrobiota* phyla between high and low polygenic risk groups (Supplementary Figure 1A). However, these differences were only observed at $\alpha = 1$ with a level of confidence of 0. At $\alpha = 0.1$, the only differences observed were in the abundances of the *Bacteroidota* and *Proteobacteria* phyla (Supplementary Figure 1B). This further suggests that there are differences in *Bacteroidota* and *Proteobacteria* abundances between PRS groups. However, at $\alpha = 0.05$, no statistically significant taxonomic differences were identified between high and low PRS groups. Therefore, no differences in taxonomic abundances were observed in the DESeq2 analysis at phyla level.

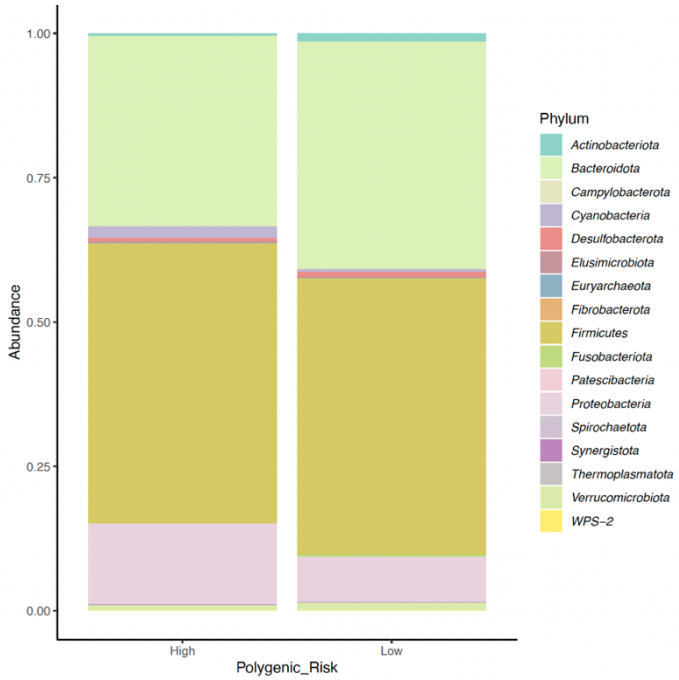


Figure 3. Relative taxonomic abundances at phyla level across the high and low PRS groups

At genus level, the 15 most abundant taxa in the high and low polygenic risk groups were also evaluated (Figure 4A, Supplementary Table 1). Seven genera were of the *Firmicutes* phylum, along with six belonging to *Bacteroidota* and two genera of the *Proteobacteria* phylum. The three most abundant genera in the high PRS group were *Prevotella_9* (*Bacteroidota* phylum) at a mean abundance of 19.07%, *Succinivibrio* (*Proteobacteria* phylum) at 11.40% and *Faecalibacterium* (*Firmicutes* phylum) at 9.28%. However, those with low polygenic risk possessed *Bacteroides* (*Bacteroidota* phylum) at a mean abundance of 15.54%, *Prevotella_9* at 13.17% and *Faecalibacterium* at 8.13%, as the three most abundant genera. This suggests that there are differences between the high and low polygenic risk groups at a genus level. In addition to there being greater abundances of *Bacteroides* in the low PRS group, the *Alistipes* genus was also observed at higher levels. In comparison, participants with high polygenic risk had greater abundances of the genera *Prevotella_9* and *Succinivibrio*.

The top 15 most abundant genera were further assessed by dividing PRS groups into high and low BMI sub-groups (Figure 4B, Supplementary Table 2). The participants with high risk and high BMI (High/High) had the greatest abundance of *Prevotella_9* (19.23%), compared to the low risk and low BMI (Low/Low) sub-group with the lowest *Prevotella_9* abundance at 8.46%. *Prevotella_9* was the predominant genus in High/High (n = 37), High/Low (n = 7) and Low/High (n = 28) sub-groups. On the contrary, the predominant genus in the Low/Low (n = 17) sub-group was *Bacteroides* (18.57%), found at higher abundances than in the other sub-

groups. In the High/High sub-group there was a mean relative abundance of *Bacteroides* of 7.18%. The abundances of the *Prevotella_9* and *Bacteroides* genera across sub-groups suggest associations between the high PRS and high BMI participants with *Prevotella_9*, and the low PRS, low BMI sub-group with *Bacteroides*. Thus, these two genera were considered as taxa of interest.

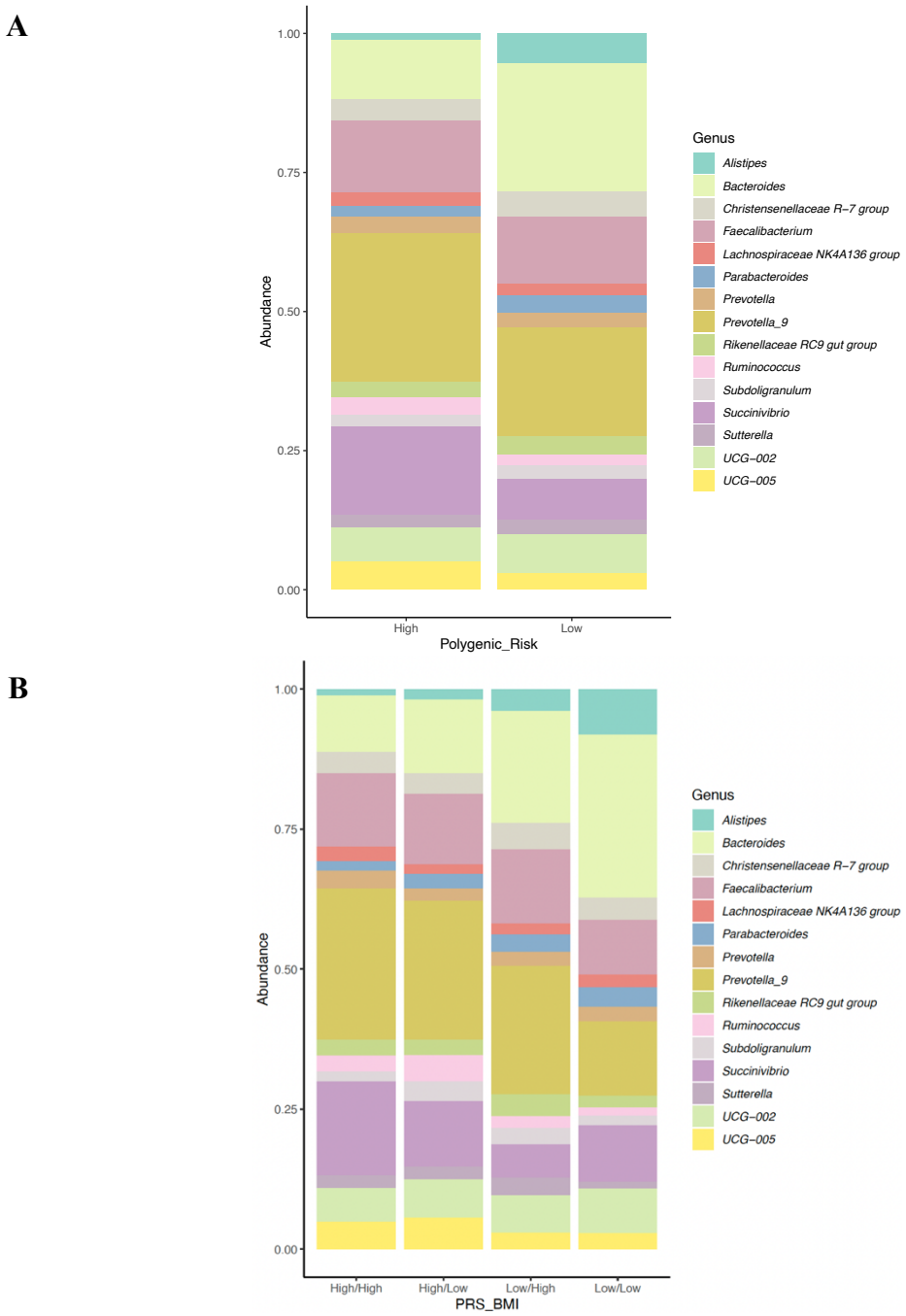


Figure 4. Taxonomic abundances comparing differential taxa abundances at a genus level for the 15 most abundant genera between (**A**) participants of high and low polygenic risk and (**B**) sub-groups of high and low BMI within the high and low PRS groups. High/High are participants with high polygenic risk and high BMI, High/Low are high risk and low BMI, Low/High are low risk and high BMI, and Low/Low are low risk and low BMI.

3.6 Identifying Discriminative Features of the Gut Microbiome

When BMI was used to evaluate differences that separate lean participants from the overweight/obese group within the high and low PRS groups (n = 89), ten discriminative features were found (Figure 5A). This was executed at the last known level of taxonomic rank, often at genus level. Of the 10 discriminative features found, one discriminative feature was detected in the lean group and nine were identified in the overweight/obese group. The *LachnospiraceaeAC2044group* genus (*Firmicutes* phylum) was the most predominant taxon in the lean BMI group. In contrast, in the overweight/obese group, the *Gammaproteobacteria* class (*Proteobacteria* phylum) was most dominant. Discriminative features identified in the overweight/obese group were largely of the *Enterobacterales* order (*Gammaproteobacteria* class) which included other discriminative features, the *Enterobacteriaceae* family and *Klebsiella* genus. The *Succinivibrio* genus, of the *Succinivibrionaceae* family, also within the *Enterobacterales* order, was also seen at its highest abundance in the High PRS/High BMI sub-group (11.93%) compared to the other sub-groups (Figure 4B, Supplementary Table 2).

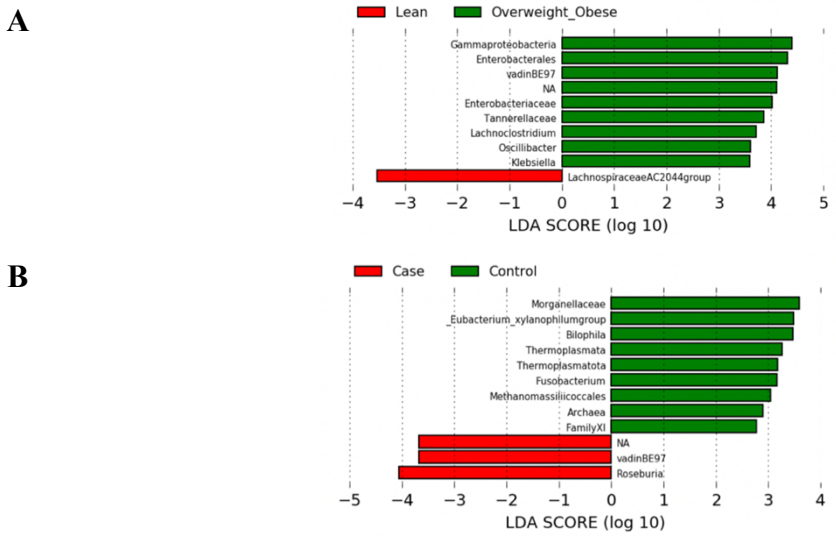


Figure 5. Linear discriminant analysis effect size at the level of last known taxonomic rank, between (**A**) lean and overweight/obese BMI groups, where red is taxa over-represented in the lean group, and green is taxa overrepresented in the overweight/obese group and (**B**) high (cases) and low (control) PRS groups, with red representing taxa that are overrepresented in cases and green being controls. NA is an unclassified genus within the *VadinBE97* family.

Additionally, a comparison was undertaken to identify discriminative features between high and low PRS groups. As a result, 12 significantly discriminative features were detected (Figure 5B). Nine features were overrepresented in the low PRS (control) group, compared to three features in the high PRS (case) group. Regarding *Bacteria*, the *Morganellaceae* family (*Proteobacteria* phylum, *Gammaproteobacteria* class) was the most predominant taxon in the

low PRS group, while the *Roseburia* genus (*Firmicutes* phylum) was most predominant in the high PRS participants. However, the low polygenic risk group was also observed to possess discriminative abundances of *Archaea*, specifically the *Thermoplasmatota* phylum (also referred to as *Euryarchaeota*). *Thermoplasmatota* encompasses other discriminative features that were identified, including the *Thermoplasmata* class and *Methanomassiliicoccales* order.

Through comparison of Figure 5A and 5B, it can be deduced that *VadinBE97* is overrepresented in both the overweight/obese and high polygenic risk groups. *VadinBE97* (including NA, an unclassified genus of *VadinBE97*) is a family in the *Verrucomicrobiota* phylum (*Lentisphaeria* class, *Victivallales* order). *VadinBE97* was included at a family level. This was the last known taxonomic rank and there has been no further classification at genus or species level. Since *VadinBE97* was a discriminative feature of both overweight/obese and high PRS groups, which is suggestive of possible associations, it was selected as a taxon of interest.

3.7 Analysis of Taxa of Interest with Polygenic Risk and Body Mass Index

Overall, three taxa of interest were identified and selected for multiple linear regression modelling (BMI ~ PRS + abundance of taxa of interest). The taxa of interest were the *Prevotella_9* genus, *Bacteroides* genus, and the *VadinBE97* family. No interactions were identified as statistically significant between polygenic risk and the taxa of interest alone (Table 6). However, a statistically significant interaction was observed for the *Prevotella_9* genus when including BMI in the regression model, $p = 0.0139$. Therefore, these results are suggestive of there being an interaction between BMI, PRS and *Prevotella_9*, as well as suggesting that factors, other than host genetics, influence BMI in this study cohort.

Table 6. Statistical parameters of linear regression models with BMI as the dependent variable and relative abundances of taxa of interest and PRS as independent variables

	Multiple R^2	Adjusted R^2	*p*-value
***Prevotella_9* Abundance ~ PRS**	$1.19e^{-05}$	-0.0116	0.9745
BMI ~ PRS + *Prevotella_9* Abundance	0.0957	0.0744	0.0139
***Bacteroides* Abundance ~ PRS**	0.0081	-0.0034	0.4040
BMI ~ PRS + *Bacteroides* Abundance	0.0974	0.0762	0.0697
***VadinBE97* Abundance ~ PRS**	0.0006	-0.0110	0.8240
BMI ~ PRS + *VadinBE97* Abundance	0.0173	0.0128	0.0911

4 Discussion

There is a paucity of studies exploring the interactions between host genetics and the gut

microbiome in obesity. This dearth of study data is especially true for African populations where this topic has not previously been investigated. Therefore, this study aimed to identify possible associations between host genomes, using a PRS predictive of BMI in this study cohort, and gut microbiota in a female African population. To achieve this, PRS generation and assessment was performed on the AWI-Gen phase 1 genotype data, and gut microbiome analysis was carried out on the corresponding AWI-Gen microbiome pilot study 16S rRNA sequence data.

The genetic contribution to obesity has been widely studied, including identifying SNPs associated with obesity and obesity-related traits (Cuevas-Sierra *et al.*, 2020). However, developing PRSs that are transferable across different ancestries (including those for obesity-related traits, such as BMI) need to include discovery datasets of diverse populations (Cavazos and Witte, 2021). Therefore, as expected, the PRS found to be the most predictive of BMI measures in the study cohort was the PRS generated with MAMA summary statistics, R^2 = 0.0665 and p = 0.0014 (Table 4, Figure 1). The MAMA PRS included a discovery dataset of participants from the United Kingdom Biobank (UKBB), the Japan Biobank, the African Partnership for Chronic Disease Research (APCDR) initiative and the Population Architecture using Genomics and Epidemiology (PAGE) Consortium. Thus, MAMA included cohorts of European, Asian, African and multi-ethnic descent, respectively.

PGS001228 (Tanigawa *et al.*, 2022) was also predictive of BMI, while PGS001943 (Privé *et al.*, 2022) had a poor predictive performance. Both the PGS001228 and PGS001943 PRS had discovery datasets of only European participants from the UKBB. Tanigawa *et al*. (2022) pointed out that PGS001228 suffered limited transferability in African cohorts for quantitative traits, such as BMI. Out of the four cohorts that transferability was assessed on, the African cohort had the lowest transferability of PGS001228, y = -0.145x (Tanigawa *et al.*, 2022). Similarly, Privé *et al*. (2022) assessed the transferability of PGS001943 on nine ancestral groups included in the UKBB, including an African ancestral group from Nigeria. It was found that the PRS was least predictive in the African cohort when compared to other ancestries, with a predictive performance of ~ 18% (Privé *et al.*, 2022). Factors that greatly influence the predictability and transferability of PRSs across ancestries include, but are not limited to, populational differences in linkage disequilibrium, SNP selection for the PRS calculations, populational allele frequencies, and different causal variants across populations. With regard to BMI, European-based PRSs with accuracies of 70-80% have been unable to predict BMI accurately in African cohorts (Cavazos and Witte, 2021), accounting for the lack of

transferability of PGS001943 in this study. These findings show that PRSs with discovery datasets from multiple ancestries are better predictors of BMI in an African cohort. Therefore, MAMA was more predictive and allowed for more accurate downstream analyses.

Unlike the well-documented evidence supporting genetic variation as a driver for obesity, many associations identified between gut microbiota and obesity remain inconclusive (Cuevas-Sierra *et al.*, 2020) and inconsistent. Within this study cohort, at phylum level, the participants with high PRS exhibited higher levels of *Proteobacteria* and decreased abundances of *Bacteroidota* compared to the low polygenic risk group (Figure 3). *Proteobacteria* is one of the four most predominant phyla in the gut microbiome and has been the phylum most consistently associated with obesity, as reviewed by Xu *et al*. (2022). Additionally, *Proteobacteria* has been proposed as a microbial signature of dysbiosis (Rizzatti *et al.*, 2017). This phylum is divided into six classes, one of which being *Gammaproteobacteria*, a discriminative feature identified in overweight/obese study participants (Figure 5A). One of the other four most populous phyla, *Bacteroidota*, has been largely associated with leanness (Xu *et al.*, 2022). However, this phylum encompasses genera of microbial taxa that have been both obesity- and lean-associated, including two taxa of interest, *Prevotella_9* and *Bacteroides*.

An inverse relationship in the abundances of *Prevotella* and *Bacteroides* was revealed when comparing phyla between high and low PRS groups (Figure 4A). Interestingly, *Prevotella* has been previously associated with diets high in carbohydrates and fiber, while *Bacteroides* have been reported to be predominant in individuals with a protein-rich Western diet (de Filippis *et al.*, 2016; Hjorth *et al.*, 2019). Study participants from Agincourt are more likely to have the former diet dominated by carbohydrates and fibers due to rural and peri-urban lifestyles. Conversely, Soweto participants are more likely to be influenced by Western diets. Therefore, it could be assumed that environmental differences, such as diet, are confounding factors influencing taxonomic abundances at a genus level. However, both the low and high PRS groups have an equal number of Soweto participants (n = 11) and approximately the same number of Agincourt recruits (n = 33 and 34, respectively). Therefore, it is suggested that *Prevotella* and *Bacteroides* abundances may be affected by factors other than diet.

The *Prevotella_9* genus was found to have significant associations with the High PRS/High BMI sub-group (n = 37). Additionally, *Prevotella_9* was the predominant genus in three sub-groups: High/High (with a mean abundance of 19.23%), High/Low (18.22%) and Low/High (15.87%), see Supplementary Table 2. The Low PRS/Low BMI sub-group had a significantly

lower mean abundance of *Prevotella_9* (8.46%) when compared to the other sub-groups. This evidence suggests a non-additive interaction between these factors, in addition to the significant association identified, $p = 0.0139$ (Table 6). However, host genetics does not seem to be driving the interaction demonstrated, as the association only becomes significant when adjusted for BMI. Additionally, high BMI seems to be associated with increased *Prevotella_9* abundance, as indicated by Oduaran *et al*. (2020). In line with these findings, the *Prevotellaceae* family (including the *Prevotella* genus) has previously been associated with women possessing high PRSs and high BMI (Cuevas-Sierra *et al.*, 2020). In the aforementioned study, it was demonstrated that the PRS proposed by the authors interacted with *Prevotellaceae*. Additionally, abundances of *Prevotellaceae* and *Prevotella* have also previously been associated with obesity and there have been reports of host genetics influencing abundances of *Prevotella_7* and *Prevotella_9* (Wells *et al*., 2020). Even though this relationship between host genetics and *Prevotella* has been observed previously, causality could not be ascertained in this study, which may be due to gut microbiome construction or small sample size.

On the other hand, the *Bacteroides* genus has been reported to have contradictory associations to lean and obese cohorts (Xu *et al*. (2022)). There was a mean *Bacteroides* abundance of 18.57% in the Low PRS/Low BMI sub-group, compared to 7.18% in High/High participants (Supplementary Table 2). Additionally, the High/Low and Low/High sub-groups possessed intermediate mean abundances of 9.66% and 13.82%, respectively. As the polygenic risk and the BMI of participants increases, there appears to be a steady decrease in the abundance of *Bacteroides*, possibly suggesting an additive effect. Additionally, the results of the multiple linear regression suggest that there is a moderate association between *Bacteroides* with low polygenic risk and low BMI participants (n = 17). However, this association is moderate and not of statistical significance, $p = 0.0697$ (Table 6), which may be due to small sample size and subsequent loss of power. Therefore, PRS and abundances of *Bacteroides* are likely independent factors with an additive effect on BMI, which is not suggestive of an interaction between these factors.

Linear discriminant analyses of the lean versus overweight/obese groups (Figure 5A) revealed that the *Gammaproteobacteria* class and *Enterobacterales* order were most predominant in the overweight/obese group. The *Enterobacteriaceae* family within the *Enterobacterales* order, including the *Klebsiella* genus, were also discriminant features of high BMI (overweight/obese) participants. Additionally, the *Succinivibrio* genus of the *Succinivibrionaceae* family, in the same order, was also seen at a greater abundance in participants with high PRS (Figure 4A).

Previous studies have supported that gram-negative bacteria, including *Enterobacterales* order, are associated with obesity (Narayanan *et al*., (2021)). Additionally, LEfSe results identified the *VadinBE97* family (*Verrucomicrobiota* phylum) as a discriminative feature of both the high PRS cases and the overweight/obese group (see Figure 5A and 5B). This resulted in *VadinBE97* being included as a taxon of interest as there were possible interactions between high PRS, high BMI and *VadinBE97*. This family has not previously been associated with host genetics or obesity; however, this study found moderate associations with high PRS and high BMI participants, $p = 0.0911$ (Table 6). Another taxon within the *Verrucomicrobiota* phylum, the *Akkermansia* genus, has previously been associated with obesity and linked to host genetics, specifically the rs4894707 variant (Davenport *et al*., (2015)). On the contrary, this association has been contradicted by findings that have categorised this taxon as lean-associated (Xu *et al.*, 2022), highlighting the need for more research investigating gut microbiota and obesity associations.

Overall, the findings of this study provide support that there are gut microbial differences between high and low polygenic risk groups in an obese female African cohort. Furthermore, this study supported the idea that African females with high BMI and high polygenic risk are significantly associated with *Prevotella_9*, which was only observed when incorporating BMI into the regression model. Additionally, high BMI and high-risk individuals were moderately associated with *VadinBE97*. By contrast, the low BMI and low polygenic risk sub-group was moderately associated with greater *Bacteroides* levels. However, these findings must be further explored in much larger sample sizes to form robust associations.

Potential Shortcomings and Study Limitations

The most pertinent limitation of this study is that of the sample size and the study, consequently, not being sufficiently powered. Due to the small sample size of 149 participants in the cohort and the 89 participants included in the combined high and low polygenic risk groups, the study was not adequately powered to pick up robust associations. Therefore, the interactions identified would need to be further explored with a larger sample size for novel associations to be discovered. Additionally, it is important to note that the four sub-groups did not have an equal distribution of study participants. The High PRS/High BMI subgroup was the largest sub-group, comprising 37 participants. The Low PRS/High BMI sub-group had 28 participants, the Low PRS/Low BMI sub-group had 17 and High PRS/Low BMI sub-group was underrepresented with seven participants, ranking in as the smallest sub-group. This may have influenced the findings of this study. Therefore, there should also be an attempt, when

increasing the cohort, to have equal or similar numbers of stratified participants, particularly for the microbiome analysis.

Although the PRS was predictive of BMI in the population, there are limits to the ability of a PRS to include all BMI-associated SNPs, specifically in populations of African ancestry that have been significantly understudied. Furthermore, African populations are underrepresented in human microbiome studies and much less is known about gut microbiota associations with obesity in individuals of African ancestry. The study participants came from two distinct types of settlements, the transitioning rural and peri-urban settlement of Agincourt and the urban settlement of Soweto. These environmental differences, as well as dietary intake, were not considered and left a potential for confounding differences between rural and urban gut microbiota. Obesity is also influenced by other environmental factors and can be measured by anthropometric measures other than BMI, which were not accounted for in this study. Finally, major time constraints hindered other analyses from being performed on the genotype and microbial data, including exploring other avenues of gut microbial analysis. This includes, but is not limited to, confirming findings with the AWI-Gen microbiome pilot study's metagenomic shotgun sequencing data, as shotgun sequencing data is more comprehensive than 16S rRNA sequencing. Furthermore, due to time constraints, other statistical analyses were not performed to further assess the interactions between PRS, BMI and taxa of interest.

Future Studies and Considerations

Since a larger sample size is required for robust associations to be demonstrated between host genetics and gut microbiota composition, this study could be extended to larger cohorts. This will allow for higher-powered studies and the identification of more robust associations. To possibly pick up stronger associations with host genetics, this study could be expanded to a cohort comprising children and adolescents, as the environmental impact would be less pronounced. This may be valuable in understanding the progression of obesity. Additionally, other anthropometric measures of obesity, other than BMI, and health metrics that influence obesity could be included for a more holistic analysis. Whole shotgun metagenomic sequencing could be used in conjunction with, or in place of, 16S rRNA sequencing for microbiome analysis. This would allow for a more comprehensive taxonomic characterisation (Brewster *et al.*, 2019), as shotgun sequencing can better identify less abundant taxa and enhance the prediction of taxonomic content. Furthermore, this study could be extended to assess previously identified associations between SNPs and abundances of taxa of interest, such as rs4894707 and *Akkermansia* or rs878394 and *Prevotella*, in the AWI-Gen microbiome pilot study cohort.

Both the rs4894707 and rs878394 variants were included on the Illumina's Infinium™ H3Africa SNP Genotyping Array v1 used for genotyping in AWI-Gen phase I. Additionally, these associations were identified in a German founder population of European ancestry and a cohort of Asian ancestry, respectively (Davenport *et al.*, 2015; Li *et al.*, 2018), and have not been investigated in individuals of African ancestry.

The downstream action of studies such as this, would primarily provide mechanistic insights into the knowledge of associations between host genetics and gut microbiota. This will deepen understandings of gut microbiota associated with obesity, specifically in an African context. Therefore, it could lead to a greater comprehension of which microbes to reduce or eliminate in disease states, such as obesity. Furthermore, data collected in this field could aid in alleviating the public health burden of obesity on the African continent and contribute to possible therapeutics and targeted healthcare intervention strategies.

5 Conclusion

This study informed the interaction between host genetics and gut microbiota, explored using SNP array genotyping and 16S rRNA microbiota data. *Firmicutes* was identified as the dominant phylum in both high and low polygenic risk groups. However, low PRS participants were found to possess greater abundances of the *Bacteroidota* phylum, while the high PRS group had greater levels of *Proteobacteria*. Although limitations in sample size hindered the identification of robust associations between these two driving factors of obesity, interactions were identified. These included a significant association and potential interaction between the high polygenic risk and high BMI participants with *Prevotella_9* and a moderate association between the low polygenic risk and low BMI sub-group and the *Bacteroides* genus. Additionally, *VadinBE97* was found to be a discriminative feature in overweight/obese and high polygenic risk groups. Overall, there were no associations detected between mean abundances of the taxa of interest, *Prevotella_9*, *Bacteroides* and *VadinBE97*, with PRS alone. However, due to the small sample size and subsequent lack of robust associations identified, this study should serve as a model for future research exploring relationships between host genetics and gut microbiota.

6 Author Contributions

All authors contributed. OO and SH conceived and formed the study design, based on a study by Cuevas-Sierra *et al*. (2020). SS performed the PRS generation and analysis, and with assistance from OO, the gut microbiota analysis. SS wrote the manuscript, reviewed and

approved by OO and SH.

7 Acknowledgements

The authors would like to thank Dr Tinashe Chikowore (Sydney Brenner Institute for Molecular Bioscience) for assisting in the development and application of PRS used in this study.

8 Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Allali, I. *et al*. (2021) Human microbiota research in Africa: a systematic review reveals gaps and priorities for future research, Microbiome, 9. DOI: https://doi.org/10.1186/s40168-021-01195-7.

Anderson, M. J. (2017). Permutational multivariate analysis of variance (PERMANOVA). Wiley StatsRef: Statistics Reference Online, pp. 1– 15.
DOI: https://doi.org/10.1002/9781118445112.stat07841

Brewster, R. *et al*. (2019) Surveying Gut Microbiome Research in Africans: Toward Improved Diversity and Representation, Trends in Microbiology, 21(10), pp. 824–835.
DOI: https://doi.org/10.1016/j.tim.2019.05.006.

Callahan, B.J. *et al*. (2016) DADA2: High-resolution sample inference from Illumina amplicon data, Nature Methods, 13(7), pp. 581–583. DOI: https://doi.org/10.1038/nmeth.3869.

Cavazos, T.B. and Witte, J.S. (2021) Inclusion of variants discovered from diverse populations improves polygenic risk score transferability, Human Genetics and Genomics Advances, 2(1).
DOI: https://doi.org/10.1016/j.xhgg.2020.100017.

Chao, A. (1984) Nonparametric Estimation of the Number of Classes in a Population, Scandinavian Journal of Statistics, 11(4), pp. 265-270. DOI: https://www.jstor.org/stable/4615964.

Choi, S.W. and O'Reilly, P.F. (2019) PRSice-2: Polygenic Risk Score software for biobank-scale data, GigaScience, 8(7). DOI: https://doi.org/10.1093/gigascience/giz082.

Cuevas-Sierra, A. *et al.* (2019) Diet, Gut Microbiota, and Obesity: Links with Host Genetics and Epigenetics and Potential Applications, Advances in Nutrition, 10, pp. S17–S30.
DOI: https://doi.org/10.1093/advances/nmy078

Cuevas-Sierra, A. *et al*. (2020) Sex-Specific Associations between Gut Prevotellaceae and Host Genetics on Adiposity, Microorganisms, 8(6), pp. 1–19.
DOI: https://doi.org/10.3390/microorganisms8060938.

Dabrowska, K. and Witkiewicz, W. (2016) Correlations of Host Genetics and Gut Microbiome Composition, Frontiers in Microbiology, 7. DOI: https://doi.org/10.3389/fmicb.2016.01357.

Davenport, E.R. *et al*. (2015) Genome-Wide Association Studies of the Human Gut Microbiota, PLoS

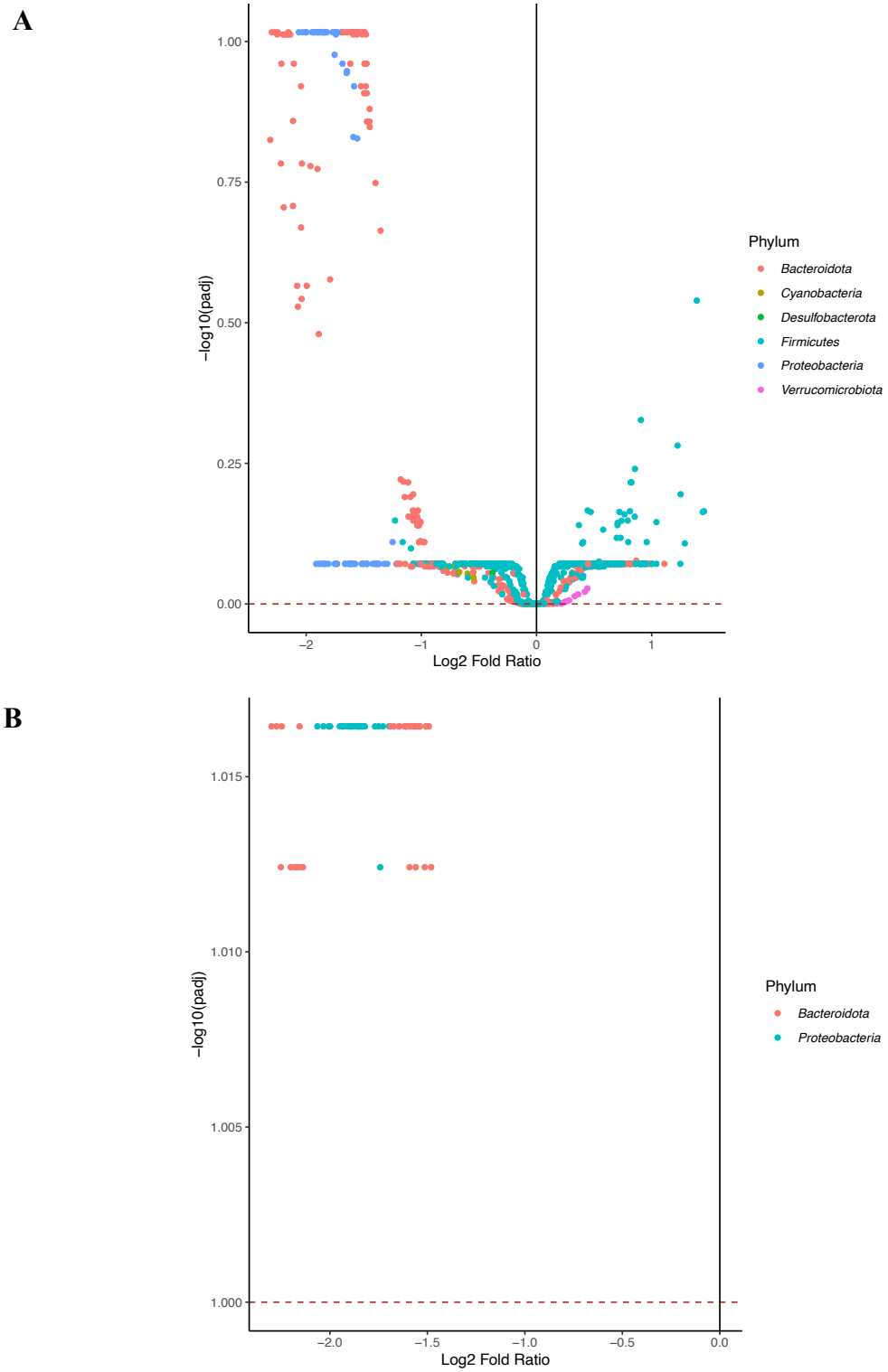
ONE, 10(11). DOI: https://doi.org/10.1371/journal.pone.0140301.
Dugas, L.R. *et al*. (2016) The obese gut microbiome across the epidemiologic transition, Emerging Themes in Epidemiology, 13(2). DOI: https://doi.org/10.1186/s12982-015-0044-5.
Euesden, J., Lewis, C.M. and O'Reilly, P.F. (2015) PRSice: Polygenic Risk Score software, Bioinformatics, 31(9), pp. 1466–1468. DOI: https://doi.org/10.1093/bioinformatics/btu848.
de Filippis, F. *et al.* (2016) Unusual sub-genus associations of faecal *Prevotella* and *Bacteroides* with specific dietary patterns, Microbiome, 4. DOI: https://doi.org/10.1186/s40168-016-0202-1.
Goodarzi, M.O. (2018) Genetics of obesity: what genetic association studies have taught us about the biology of obesity and its complications, The Lancet Diabetes and Endocrinology.
DOI: https://doi.org/10.1016/S2213-8587(17)30200-0.
Gower, J.C. (2005) Principal Coordinates Analysis, Encyclopedia of Biostatistics.
DOI: https://doi.org/10.1002/0470011815.b2a13070.
Hjorth, M.F. *et al.* (2019) *Prevotella*-to-*Bacteroides* ratio predicts body weight and fat loss success on 24-week diets varying in macronutrient composition and dietary fiber: results from a post-hoc analysis, International Journal of Obesity, 43(1), pp. 149–157.
DOI: https://doi.org/10.1038/s41366-018-0093-2.
Khera, A.V. *et al.* (2019) Polygenic prediction of weight and obesity trajectories from birth to adulthood, Cell, 177(3), pp. 587-596.e9. DOI: https://doi.org/10.1016/j.cell.2019.03.028.
Lambert, S.A. *et al* (2021) The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation, Nature Genetics, 53(4), pp. 420-425. DOI: 10.1038/s41588-021-00783-5
Li, J. *et al.* (2018) A metagenomic approach to dissect the genetic composition of enterotypes in Han Chinese and two Muslim groups, Systematic and Applied Microbiology, 41(1).
DOI: https://doi.org/10.1016/j.syapm.2017.09.006.
Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biology, 15(12).
DOI: https://doi.org/10.1186/s13059-014-0550-8.
Loos, R.J.F. and Yeo, G.S.H. (2022) The genetics of obesity: from discovery to biology, Nature Reviews Genetics (23), pp. 120–133. DOI: https://doi.org/10.1038/s41576-021-00414-z.
McMurdie, P.J. and Holmes, S. (2013) Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data, PLoS ONE, 8(4).
DOI: https://doi.org/10.1371/journal.pone.0061217.
Mulder N, *et al*. (2018). H3Africa: current perspectives. Pharmacogenomics and Personalized Medicine, 10(11), pp. 59-66. DOI: 10.2147/PGPM.S141546.
Narayanan, N. *et al*. (2021) Relationship Between Multidrug-Resistant Enterobacterales and Obesity in Older Adults, Infection and Drug Resistance, 14, pp. 2527–2532.
DOI: https://doi.org/10.2147/IDR.S317014.
Privé, F. *et al*. (2022) Portability of 245 polygenic scores when derived from the UK Biobank and applied to 9 ancestry groups from the same cohort, American Journal of Human Genetics, 109(1), pp.

12–23. DOI: https://doi.org/10.1016/j.ajhg.2021.11.008.
R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
Ramsay, M. *et al*. (2016) H3Africa AWI-Gen Collaborative Centre: a resource to study the interplay between genomic and environmental risk factors for cardiometabolic diseases in four sub-Saharan African countries, Global Health, Epidemiology and Genomics, 1.
DOI: https://doi.org/10.1017/gheg.2016.17.
Rizzatti, G. *et al.* (2017) Proteobacteria: A Common Factor in Human Diseases, BioMed Research International. DOI: https://doi.org/10.1155/2017/9351507.
Segata, N. *et al* (2011) Metagenomic biomarker discovery and explanation. Genome Biology, 12(6). DOI: 10.1186/gb-2011-12-6-r60.
Shannon, C.E. (1948) A Mathematical Theory of Communication, The Bell System Technical Journal, 27, pp. 379-423.
Tanigawa, Y. *et al*. (2022) Significant sparse polygenic risk scores across 813 traits in UK Biobank, PLoS Genetics, 18(3). DOI: https://doi.org/10.1371/journal.pgen.1010105.
Thaker, V.V. (2017) Genetic and Epigenetic Causes of Obesity, Adolescent Medicine: State of the Art Reviews, 28(2), pp. 379–405. DOI: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6226269/
Wang, Q. *et al.* (2007) Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy, Applied and Environmental Microbiology, 73(16), pp. 5261–5267.
DOI: https://doi.org/10.1128/AEM.00062-07.
Wells, P M *et al*. (2020) Associations between gut microbiota and genetic risk for rheumatoid arthritis in the absence of disease: a cross-sectional study, 2, pp. e418–27.
DOI: www.thelancet.com/rheumatology.
Wickham H, François R, Henry L, Müller K (2022). dplyr: A Grammar of Data Manipulation.
DOI: https://dplyr.tidyverse.org, https://github.com/tidyverse/dplyr.
Xu, Z. *et al.* (2022) Gut microbiota in patients with obesity and metabolic disorders — a systematic review, Genes and Nutrition, 17(2). DOI: https://doi.org/10.1186/s12263-021-00703-6.
Yako, Y.Y. *et al.* (2015) Genetic association studies of obesity in Africa: A systematic review, Obesity Reviews, 16, pp. 259–272. DOI: https://doi.org/10.1111/obr.12260.

Electronic References

https://www.pgscatalog.org/ [accessed August 2022]
https://huttenhower.sph.harvard.edu/galaxy/ [accessed November 2022]

Supplementary Material



Supplementary Figure 1. Volcano plots generated using DESeq2 to compare differential abundance between case and control sub-cohorts at (**A**) α = 1 and (**B**) α = 0.1

Supplementary Table 1. Taxonomic abundances (%) for the top 15 most abundant taxa between high and low polygenic risk groups

Genus	**Taxonomic Abundances of Top 15 Genera in PRS Groups (%)**	
	High PRS (n = 44)	Low PRS (n = 45)
Alistipes	0.91	3.63
Bacteroides	7.57	15.54
Christensenellaceae R-7 group	2.74	2.99
Faecalibacterium	9.28	8.13
Lachnospiraceae NK4A136 group	1.79	1.38
Parabacteroides	1.31	2.18
Prevotella	2.16	1.71
Prevotella_9	19.07	13.18
Rikenellaceae RC9 Gut Group	2.00	2.22
Ruminococcus	2.26	1.28
Subdoligranulum	1.51	1.68
Succinivibrio	11.40	4.97
Sutterella	1.61	1.69
UCG-002	4.43	4.75
UCG-005	3.57	1.96
Total % of Gut Microbiota Composition Accounted For	71.61	67.29

Supplementary Table 2. Taxonomic abundances (%) for the top 15 most abundant taxa between high and low body mass index participants within the high and low polygenic risk groups

Genus	**Taxonomic Abundances of Top 15 Genera in Sub-Groups (%)**			
	High PRS/High BMI (n = 37)	High PRS/Low BMI (n = 7)	Low PRS/High BMI (n = 28)	Low PRS/Low BMI (n = 17)
Alistipes	0.82	1.36	2.72	5.22
Bacteroides	7.18	9.66	13.82	18.57
Christensenellaceae R-7 group	2.74	2.73	3.27	2.51
Faecalibacterium	9.29	9.24	9.21	6.25
Lachnospiraceae NK4A136 group	1.90	1.22	1.33	1.45
Parabacteroides	1.20	1.90	2.14	2.25
Prevotella	2.26	1.64	1.75	1.65
Prevotella_9	19.23	18.22	15.87	8.46
Rikenellaceae RC9 Gut Group	2.00	1.99	2.71	1.36
Ruminococcus	2.03	3.48	1.48	0.92
Subdoligranulum	1.31	2.53	2.02	1.07
Succinivibrio	11.93	8.60	4.11	6.48
Sutterella	1.60	1.67	2.20	0.78
UCG-002	4.32	5.00	4.59	5.04
UCG-005	3.46	4.14	2.04	1.83
Total % of Gut Microbiota Composition Accounted For	71.27	73.38	69.26	63.84

Appendices

Appendix A: Approved Research Protocol

CANDIDATE'S SURNAME: **SCHNELL**	FIRST NAME/S: **SAMANTHA SUSAN**	STUDENT NUMBER: **1914985**	
CURRENT QUALIFICATIONS:	**BSc (Human Physiology, Genetics and Psychology)** **BHSc Hons (Human Genetics)**		
TEL: **082 331 1486**	CELL: **082 331 1486**	E-MAIL: **sammyschnell@gmail.com**	FAX: **N/A**
DEGREE FOR WHICH PROTOCOL IS BEING SUBMITTED: **MSc (Med) in Genomic Medicine**			
PART-TIME OR FULL-TIME: **Full-time**			
FIRST REGISTERED FOR THIS DEGREE:	TERM: **1**	YEAR: **2022**	
DEPARTMENT: **Division of Human Genetics, School of Pathology**			
TITLE OF PROPOSED RESEARCH: **Exploring the Interaction of Host Genetics and the Gut Microbiome in Obesity**			
CANDIDATE'S SIGNATURE:		DATE: **16/02/2022**	
SUPERVISOR 1 (NAME & SURNAME): **Dr Ovokeraye Oduaran**		% Supervision: **50%**	
SUPERVISOR'S QUALIFICATIONS: **MSc, PhD**			
SUPERVISOR'S DEPARTMENT: **The Sydney Brenner Institute for Molecular Biosciences**			
SUPERVISOR'S ADDRESS / TEL / E-MAIL: **ovokeraye.oduaran@wits.ac.za**			
SUPERVISOR 2 (NAME & SURNAME): **Prof Scott Hazelhurst**		% Supervision: **50%**	
SUPERVISOR'S QUALIFICATIONS: **MSc, PhD**			
SUPERVISOR'S DEPARTMENT: **School of Electrical & Information Engineering**			
SUPERVISOR'S ADDRESS / TEL / E-MAIL: **scott.hazelhurst@wits.ac.za**			
SYNOPSIS OF RESEARCH: **See next page.**			

WITS ETHICS NOT REQUIRED: Yes No WITS ETHICS PENDING: Yes No WITS ETHICS APPROVED: (Yes) No (Circle appropriate symbol) * ***Please note the final human ethics clearance certificate or animal ethics certificate must be available prior to starting research**	IF Y SUPPLY ETHICS CLEARANCE CERTIFICATE ASATTACHMENT AND INCLUDEETHICS NUMBER HERE: **M220297**
As supervisor/s, I/we confirm that I have read the protocol which has been submitted for assessment.	
SIGNATURE OF SUPERVISOR/S: Ovokeraye Oduaran	

SYNOPSIS OF RESEARCH CONTINUED:

Introduction:
The rapid increase in obesity worldwide has resulted in a need to better understand the aetiologies of this public health burden, including factors such as single nucleotide polymorphisms (SNPs) within host genetics, specifically those in *FTO*, *MC4R* and *SEC16B* genes, and gut microbial taxa like *Bacteroidetes* and *Firmicutes*. Associations between obesity- and body mass index (BMI)-related SNPs and the gut microbiome (GM) do exist and need to be evaluated further, specifically in African populations. This is because few associations have been found to be statistically significant at a genome-wide level and little is known about this interaction and the GM in an African context. Therefore, the aim of this study is to evaluate associations between BMI-related SNPs and gut microbial composition in an African population, with the research question of whether BMI-related SNPs show a pattern of association to gut microbial composition.

Justification for Study:
Although previous studies have been undertaken to determine the association between BMI-related SNPs and microbial taxa within the GM, many associations have not been found to be statistically significantat a genome-wide level. Furthermore, the interaction of host genetics and the GM is largely understudied in African populations. Research has independently implicated BMI-related SNPs, as well as abundances of certain bacterial taxa inhabiting the GM in the development of obesity. Therefore, the relationship between these two factors needs to be explored further. Identifying potential interactions will provide further insight into understanding the aetiology of obesity, which will improve precision medicine interventions.

Aim:
To evaluate the interaction between BMI-related SNPs within host genomes and gut microbial compositions in a female African population to identify possible associations.

Proposed Methodology:
1. Construct a genetic risk score using BMI as a continuous phenotypic trait based on an approach trained and validated on the AWI-Gen dataset, excluding the samples used in this study, and to apply it to the genotype data of the study participants.
2. Stratify the participants based on low- and high- genetic risk scores (GRS) and use a linear discriminant analysis effect size (LEfSe) to compare categories of microbial taxa present between the groups.
3. Use zero-inflated Gaussian analyses (metagenomeSeq) to assess microbial taxa that differ significantly in abundance between the low- and high-GRS samples.
4. Use general linear regression (GLR) to evaluate potential interactions between microbial taxa of interest and low- and high- GRS constructed using BMI as a continuous trait, as well as between microbial taxa of interest and BMI

Expected Outcome/s:
Through the data obtained from this study, knowledge of the association between host genetics and the GM will be improved. This, in turn, will provide a greater insight into the genetics of African populations. Furthermore, Africans will benefit from the study against the public health burden of obesity, as this study will contribute to more targeted healthcare intervention strategies in the future.

Research Proposal

Exploring the Interaction of Host Genetics and the Gut Microbiome in Obesity

Samantha Schnell

1914985

MSc (Med) in Genomic Medicine

Division of Human Genetics, School of Pathology, University of the Witwaterstrand and the National Health Laboratory Service

Supervised by:

Dr Ovokeraye Oduaran, MSc, PhD

Postdoctoral Researcher, The Sydney Brenner Institute for Molecular Biosciences

Prof Scott Hazelhurst, MSc, PhD

Professor, School of Electrical and Information Engineering, University of the Witwaterstrand and The Sydney Brenner Institute for Molecular Biosciences

PLAGIARISM DECLARATION TO BE SIGNED BY ALL HIGHER DEGREE STUDENTS

SENATE PLAGIARISM POLICY: APPENDIX ONE

I, **Samantha Schell** (Student Number: **1914985**), am a student registered for the degree of **MSc (Med) in Genomic Medicine** in the academic year **2022**.

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where I have explicitly indicated otherwise.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.
- I have included as an appendix a report from "Turnitin" (or other approved plagiarism detection) software indicating the level of plagiarism in my research document.

Signature: 

Date: 14/03/2022

1. Background Literature Analysis and Critique

1.1. Introduction

Obesity is a public health burden that is dramatically increasing worldwide. The incidence of obesity was believed to be primarily prevalent in developed countries; however, it has since been found to be a growing problem in the developing world. In Southern Africa, a region that is heavily affected by this epidemic, the rise in obesity has adversely influenced mortality and morbidity through the increased incidence of non-communicable diseases. This rapid increase in obesity has been reflected in the rise in body mass index (BMI) (Cois & Day, 2015), which serves as one of the many anthropometric measures used to assess obesity. The aetiologies influencing obesity are complex and not well understood, but multiple driving factors have been implicated, including environmental factors, such as unhealthy diet and lifestyle choices, as well as genetic factors (Cuevas-Sierra et al., 2019). Single nucleotide polymorphisms (SNPs) form part of the genetic aetiology, whereby predispositions to obesity and weight gain have been linked to the accumulation of variants in many different genes, including fat mass and obesity-associated (*FTO*) and melanocortin 4 receptor (*MC4R*) genes. Furthermore, the gut microbiome (GM) has been implicated in obesity, where dysbiosis, a disruption in the homeostasis of the microbiome, has been observed in affected individuals (Hall et al., 2017). Through genome-wide association studies (GWAS) and analyses of gut microbiota, it has been demonstrated that there are differences between gut microbial composition and genetic variation in overweight and obese individuals, compared to their underweight, lean, or normal weight counterparts. However, few of these associations between obesity and BMI-related SNPs and gut microbial taxa have been found to be statistically significant at a genome-wide level (Cuevas-Sierra et al., 2020). Subsequently, there is a large discrepancy in the number of studies, such as these, undertaken in African populations. Therefore, these associations need to be further evaluated, specifically in an African context.

1.2. Polymorphisms Present in Host Genetics Implicated in Obesity

Obesity is a complex and multifactorial disease that is influenced by environmental, genetic and epigenetics factors, as well as microbial determinants (Thaker, 2017). Large-scale studies, such as GWAS, have been used to identify host genetic variants that are common in cases of childhood and adult obesity. Although well-documented, the frequency of many of these variants are not sufficient in explaining the observed prevalence of obesity within the general population (Schlauch et al., 2019). Some SNPs are located near genes that have uncertain roles in obesity while others are associated with genes that influence biological pathways implicated

in metabolism and dietary behaviour. As such, it is known that the genetic contribution to the common obesity phenotype is highly polygenic and heritable in nature (Khera et al., 2019). Therefore, there is limited ability to predict obesity through analysing single variants, as the accumulation of multiple SNPs at different loci interact to produce an additive effect, where each SNP partially contributes to the total risk of developing obesity based on their effect size. This makes the calculation of a polygenic risk score (PRS) or genetic risk score (GRS) useful in determining genetic predispositions to obesity (Chalazan et al., 2021).

1.2.1. Obesity and BMI-Related SNPs found in Literature

Genome-wide association studies for anthropometric measures, such as BMI, adiposity, and hip-to-waist ratio have identified more than 300 SNPs to be associated with obesity, as reviewed by Goodarzi (2018) (see Figure A1 for the genes found to be associated with obesity-related traits). In research undertaken by Chalazan et al. (2021), 26 variants significantly associated with BMI and obesity in individuals of European ancestry were genotyped to identify associations in African Americans. In this study, it was found that a variant in the *FTO* gene, rs8050136, was associated with obesity in that population. However, once the *p*-values were corrected for multiple testing using the Bonferroni correction, the SNP was no longer significant. Additionally, in a study done by Cuevas-Sierra et al. (2020) on a Spanish cohort, 95 SNPs that related obesity and high BMI predisposition were selected and genotyped. These SNPs were identified through an exhaustive literature review and spanned over 54 genes, including variants from the *FTO, MCR4* and *SEC16B* genes. The SNPs found to most statistically associated with BMI in this Spanish cohort were then used to construct a GRS.

1.2.2. African-Specific Obesity/BMI-Related SNPs

As discussed in a review by Yako et al. (2015), more than 300 polymorphisms located in 42 different genes have been evaluated in African populations using candidate gene approaches. However, using GWAS, only SNPs detected in *FTO* and *MC4R* genes were identified as significantly associated with obesity in South African, Nigerian, and Ghanaian individuals of African ancestry (Yako et al., 2015). The *FTO* gene is one of the genes that has been most strongly associated with obesity in humans, specifically SNPs positioned within the first and second introns. The association between *FTO* variants and obesity has been confirmed in multiple studies, both in independent studies and large GWAS, as reviewed by Zarza-Rebollo et al. (2021). The SNP that has been studied in the greatest depth is rs9939609, an alternative *A* allele, which has been associated with weight gain and obesity. Variants that are present in the *MC4R* gene influence the functioning of the MC4R protein, which is a transmembrane

protein that is expressed in high levels in the hypothalamus, the region of the brain that controls appetite (Thaker, 2017). These *MC4R* variants have been found to influence early-onset obesity. In addition, associations between the anthropometric measure of body composition and the SEC16 homolog B, endoplasmic reticulum export factor (*SEC16B*) gene have been made, where the rs543872 variant has been linked to increased percentage body fat. In South African individuals of African ancestry, intergenic variants upstream of the *SEC16B* locus, such as the rs6664269 SNP, were found to have suggestive associations to body composition (Sahibdeen et al., 2018). Unfortunately, the studies that have been undertaken have proven to be insufficient in establishing the true nature of obesity-related SNPs in African populations, and more GWAS need to be performed (Yako et al., 2015).

1.2.3. Genetic Risk Scores Associated with BMI

Due to the multigenic nature of obesity, a polygenic predictor such as GRS, can be utilised to quantify disease susceptibilities as a single quantitative measure (Khera et al., 2019). Based on the study done by Cuevas-Sierra et al. (2020), a GRS was calculated using 10 SNPs determined to be strongly associated with obesity and BMI. SNPs were summed depending on which "risk" or "non-risk" alleles participants possessed using an unweighted approach (Cuevas-Sierra et al., 2020). However, other methods of GRS construction, such as using the aggregation of risk alleles from thousands of candidate SNPs with corresponding effect sizes calculated from GWAS, have great utility in quantifying susceptibility to obesity and BMI. A method of GRS calculation incorporating the latter approach will be used for the purpose of this study.

1.3. The Human Gut Microbiome and Obesity

The GM is composed of microbial bacteria, archaea, fungi, and viruses that reside mainly in the large intestine (including the colon, rectum, and anus). The microorganisms have many functional roles, which include energy extraction from carbohydrates, vitamin production, prevention of pathogen colonisation in the gut and immune system homeostasis (Hall et al., 2017). Alterations in the composition of gut microbial taxa and GM dysbiosis have been associated with pathogenesis in various organ systems, including cancers, metabolic diseases, and obesity. Since the GM has been proven to affect the progression of obesity, it has been suggested that it may be a vital component in weight control, balancing energy homeostasis and regulating host metabolism (Cuevas-Sierra et al., 2019; Kim et al., 2019). It has also been suggested that gut microbiota influence how nutrients are harvested and digested in the gut and affect inflammation and the functioning of intestinal barriers, which could contribute to the development of obesity.

1.3.1. The Biology of the Gut Microbiome

Gut microbiota have a commensal and mutually symbiotic relationship with human hosts, for whom homeostatic balance is highly beneficial (Cuevas-Sierra et al., 2019). Conversely, when there is dysbiosis in gut microbial composition, there is a drastic imbalance between the beneficial and harmful microbial taxa present in the gut. This may result in disease due to either loss of beneficial microbial taxa, excessive growth of potentially harmful bacteria or a decrease in microbial diversity. In healthy individuals with a normal GM, the primary microbial phyla that inhabit the gut are *Actinobacteria*, *Bacteroidetes* and *Firmicutes*, with phyla including *Cyanobacteria*, *Fusobacteria*, *Proteobacteria* and *Verrucomicrobia* being less abundant (Chakraborti, 2015; Kim et al., 2019).

1.3.2. Microbial Taxa Previously Identified in Obese Individuals

Gut microbiota in models of obesity tend to show decreased microbial diversity. In the GM of obese individuals, studies have shown a reduction in the abundance of *Bacteroidetes* and a proportional increase in *Firmicutes*, resulting in a greater *Firmicutes/Bacteroidetes* ratio. This promotes energy storage in adipose tissues within the host, because *Firmicutes* are actively involved in the metabolism of glucose and lipids (Kim et al., 2019; Stephens et al., 2018). Furthermore, an increased level of *Actinobacteria* and decreased levels of *Bifidobacterium* have been identified in obese individuals (Chakraborti, 2015; Kim et al., 2019). Cuevas-Sierra et al. (2020) discovered in the Spanish cohort that *Prevotellaceae*, including the *Prevotella* genus, was the most predominant bacterial taxon in individuals with high GRS. In contrast, individuals with a low GRS had greater abundances of members of the *Bacteroides* phylum.

1.3.3. Gut Microbiome Research in Africa

Research on the human microbiome, including the GM, have been covered extensively in populations in developed countries; however, there are large knowledge gaps regarding human microbiomes in African populations, as reviewed in Allali et al. (2021). Through studies being undertaken in collaborations such as the H3Africa consortium's AWI-Gen project, which includes this study, increased knowledge, and a greater understanding of gut microbial composition in Africans will be achieved. H3Africa includes a project, H3ABioNet, which has set in motion many initiatives allowing African scientists to conduct microbiome studies focusing on African populations (Allali et al., 2021).

In many understudied populations, with the AWI-Gen study population in South Africa serving as an example, there is an epidemiological transition from rural to peri-urban and urban, in which diet and geographical location is influencing the GM of the population. Oduaran et al.

(2020) analysed the same datasets as those that will be used in this study, using collected stool samples from 170 HIV-negative African women to generate 16S ribosomal RNA (rRNA) datasets to characterise and profile their gut microbial composition. The 16S rRNA gene is a highly conserved collection of 9 hypervariable regions that can be used to identify, and subsequently quantify different microbial taxa inhabiting the GM (Di Segni et al., 2018). In both rural and urban populations, *Prevotella* was found to be associated with obesity, as it had a higher relative abundance (Oduaran et al., 2020). Strains of *Prevotella* have been linked to several human diseases due to bacteria's abilities to promote chronic inflammation. Therefore, *Prevotella* could potentially be a pathobiont. However, it has been suggested that its effect is beneficial or detrimental, depending on diet and lifestyle (Oduaran et al., 2020).

1.4. Exploring Interactions Between Host Genetics and the Microbiome

Genome-wide association studies using microbiome variation as a complex trait have identified genes that are associated with microbe-polymorphism interactions (Luca et al., 2018). The traits of microbial taxa have been found to affect the host physiology and is often considered as a form of environmental exposure, as it can elicit regulatory responses within the host organism and alter host gene regulation. Subsequently, studies performed on mice have shown that the genotypes of hosts have important roles in the regulation of microbial composition (Abdul-Aziz et al., 2016). Expression quantitative trait loci (eQTL) have been connected to the microbiome as eQTLs may be implicated in gene regulation modified by host-microbiome relationships (Luca et al., 2018). Additionally, studies on the co-evolution of the microbiome and the human genome using ancient DNA have shown that the changes in interactions over generations can be analysed to determine the extent of host genetics in shaping the GM (Abdul-Aziz et al., 2016). An extensively documented example of an interaction between host genetics and the GM, that is unrelated to obesity, is the lactase (*LCT*) gene and *Bifidobacterium* (Cuevas-Sierra et al., 2020). *LCT* affects lactase persistence and enables lactose consumption, which has been associated with abundances of *Bifidobacterium*, an inheritable bacterial taxon that correlates to host *LCT* genotypes.

1.5. Previously Identified Interactions Between SNPs and the Gut Microbiome

Human beings are more than 99% similar in their genomic makeup, but their overall microbial composition varies by approximately 80-90% (Stephens et al., 2018). The host genetics of an individual affects the chemical and physical makeup of the GM; therefore, genetic variation partly accounts for the dramatic differences in the GM between individuals, as reviewed by Hall et al. (2017). Several studies support that host genetics are involved in shaping the GM as

it influences GM structure, diversity, and composition. Through a GWAS undertaken by Davenport et al. (2015), an association was identified between a variant (rs4894707) located near the phospholipase D1 (*PLD1*) gene and obesity-related microbial genus, *Akkermansia*. Furthermore, the rs4894707 variant has been associated with BMI. Similarly, an abundance in the *Prevotella* genus has been found to have significant association to the rs878394 variant in the lysophospholipase-like 1 (*LYPLAL1*) gene, which is involved in the distribution of body fat (Cuevas-Sierra et al., 2019). However, even though some SNPs, such as variants detected in the *FTO* gene, have been associated with an abundance of microbial taxa, few associations such as this, have been identified as significant at a genome-wide level (Cuevas-Sierra et al., 2020). Therefore, the aim of this study is to evaluate associations between obesity-related SNPs and gut microbial composition in an African population, with the research question of whether BMI-related SNPs show a pattern of association to gut microbial composition.

2. Aim and Objectives

2.1. Aim

To evaluate the interaction between BMI-related SNPs within host genomes and gut microbial compositions in a female African population to identify possible associations.

2.2. Objectives

1) To construct a GRS based on a method trained and validated on the AWI-Gen dataset, excluding the samples used in this study, and to apply it to the genotype data of the study participants.
2) To analyse the 16S rRNA sequence dataset to determine GM compositions in samples stratified as low- and high-GRS and based on BMI measurements to identify taxa of gut microbiota present.
3) To evaluate potential interactions between microbial taxa of interest and GRS, as well as microbial taxa of interest and BMI, using general linear regression (GLR) models.

3. Methods

3.1. Study Participants and Sites of Study

The study cohort comprises 170 African females that were previously recruited for the AWI-Gen I and microbiome pilot studies. A cohort subset of 119 individuals were recruited from Bushbuckridge located in Mpumalanga (24.8398°S, 31.0464°E), a community encompassing a collection of transitioning rural and some peri-urban settlements. This cohort is managed by the Rural Public Health and Health Transitions Research Unit of the Medical Research Council and the University of the Witwaterstrand (MCR/Wits Agincourt Unit). The other subset were

51 individuals from Soweto (26.2485°S, 27.8540°E), an urban settlement in Johannesburg. This cohort is managed in Gauteng by the SAMRC/Wits Developmental Pathways for Health Research Unit (DPHRU). The study participants had blood drawn for genotyping in the AWI-Gen I study, and their stool samples were collected and underwent gut microbial analysis during the microbiome pilot study. Sample collection was done by the Agincourt Health and Demographic Surveillance Site (HDSS) and the DPHRU.

3.1.1. Inclusion and Exclusion Criteria

Study participants within the Bushbuckridge and Soweto communities that were included in this study were recruited in the forementioned AWI-Gen studies. To minimise confounding effects, such as sex, all the samples used were collected from African females. Subsequently, samples obtained from individuals that tested positive for HIV and two samples with collection irregularities were excluded from this study. The cohort age (years) demographics and anthropometric indicator, BMI (kg/m^2), used for this study, are visualised in Table 1.

Table 1. The age (years) and BMI (kg/m^2) distributions of the cohorts (Oduaran et al., 2020).

	Cohort	Mean ± Standard Deviation	Median	Range
Age	**Bushbuckridge**	55.10 ± 7.77	55	43 - 72
	Soweto	54.10 ± 5.86	54	43 - 64
BMI	**Bushbuckridge**	32.60 ± 7.95	31.17	21.23 - 58.97
	Soweto	36.05 ± 9.23	36.52	20.43- 58.62

3.1.2. Control and Case Participants

The control participants are the study participants that have BMI in the underweight and lean categories (< 25.0 kg/m^2), which includes 30 individuals. Conversely, the case participants are those that are of overweight (25.0-29.9 kg/m^2) and obese (≥ 30.0 kg/m^2) BMI categories. 34 individuals are classified as overweight, and 106 individuals are classified as obese.

3.2. Design of the Study

The design of this study is a retrospective data analysis of the AWI-Gen pilot gut microbial dataset, obtained from stool samples, and the corresponding genotype data from the AWI-Gen I study, sequenced from DNA extracted from venous blood samples. The DNA of the study participants previously underwent genome-wide genotyping with the H3Arica SNP array and GM analysis using 16S rRNA sequencing was performed on the stool samples. The AWI-Gen research project has three broad objectives, one being to investigate interactions between the human genome with contributions to body composition and fat distribution, as they contribute to non-communicable diseases, such as obesity (Ramsay et al., 2016). Therefore, to investigate

the interactions between the human genome and gut microbial composition and their contributions to obesity development, the objectives of this study will be used to investigate interactions between host genetics and the GM in obese African females.

4. Data Analysis

Both the genotype dataset from the AWI-Gen I study and the microbial dataset from the AWI-Gen microbiome pilot study must be analysed to determine if there are any interactions occurring between identified microbial taxa of interest and host genetics.

4.1. Genotype Data Analysis: Construction of the Genetic Risk Score

The GRS of study participants will be calculated using PRSice software. Three published GRSs that have previously been trained and validated will be used as discovery datasets for the computing of GRS calculations specific to BMI (Tanigawa et al., Privé et al. and Chikowore et al.). The predictive performance of each GRS on BMI will then be evaluated in the AWI-Gen target dataset, which will then be replicated in the microbiome pilot study dataset. See Figure 1 for the schematic representing the forementioned approach.

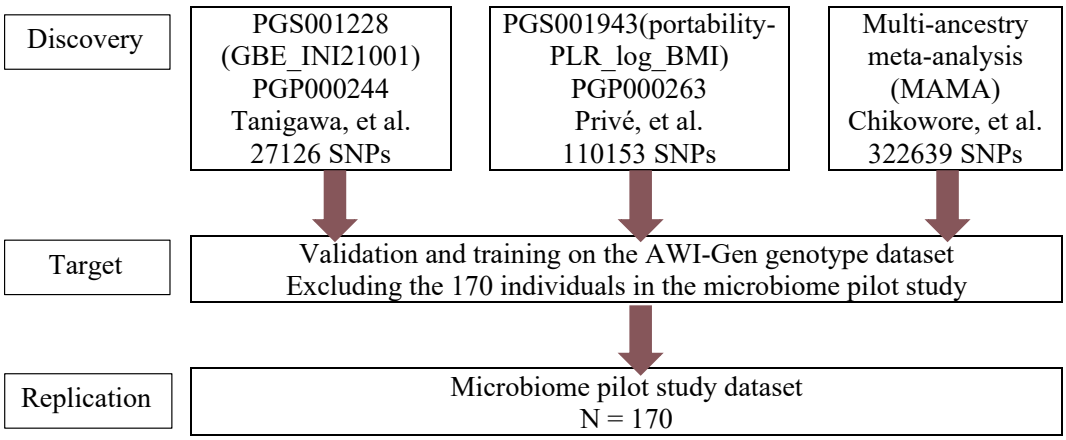


Figure 1. Schematic representing the approached used to develop the GRS for BMI, using the genotype data obtained from the microbiome pilot study participants.

Post-GRS generation and the retrieval of the most predictive GRS calculated in the microbiome pilot dataset, the mean BMI of the microbial pilot study participants will be plotted against GRS deciles (participants will be stratified into deciles of 17), to determine associations between GRS and BMI.

4.2. Microbial Data Analysis

The protocol visualised in Figure 2 represents the statistical analysis approaches that will be used to analyse the microbial datasets to assess potential interactions in association with the

GRSs generated. To identify which microbial taxa are associated with low- and high-GRS, study participants must be stratified according to GRS deciles. For the GLR models that will be used in the microbial analysis, a minimum of 30 samples is required, therefore, the top two deciles (N = 34) will form the high-GRS category, while the low-GRS category comprises the bottom two deciles (N = 34).

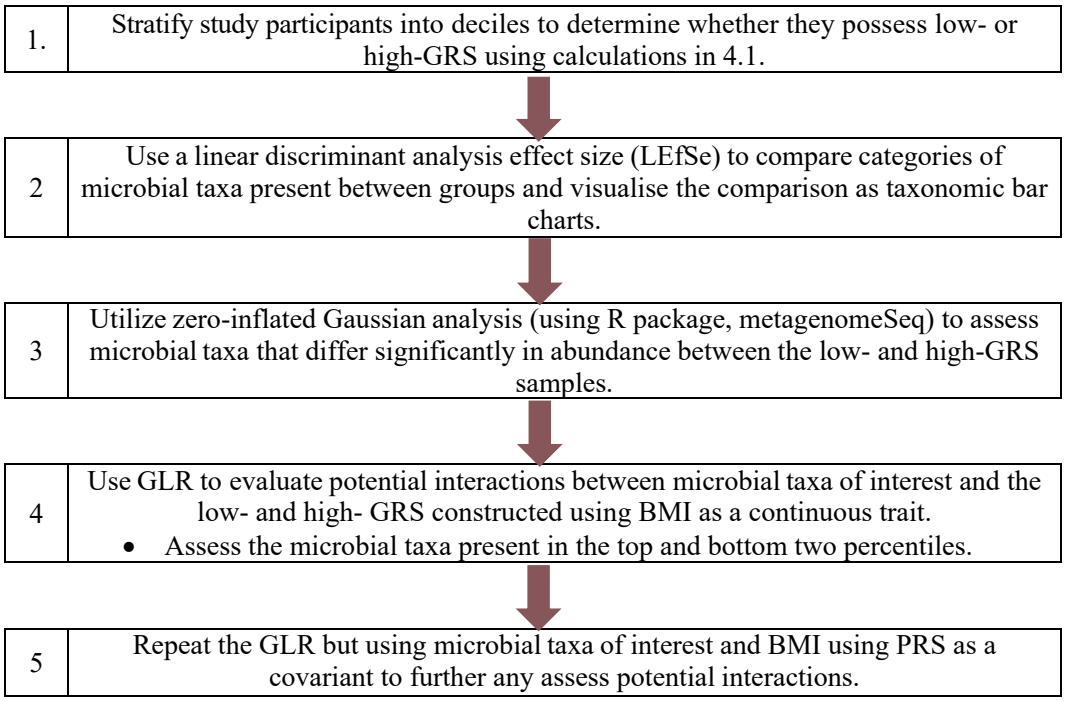


Figure 2. A flow diagram for the statistical methods used to analyse the microbial dataset.

5. Timing

This study will be undertaken over a ten-month period in 2022 (see Figure 3).

	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov
Literature Reviews										
Ethics Application										
Protocol Write-up										
Protocol Assessment										
Data Analysis										
Research Write-up										
Research Assessment										

Figure 3. A Gantt chart representing time allocations for the proposed study.

6. Funding

No funding will be required for this study as it is retrospective and will be using genotype and microbial datasets that have already been generated in previous AWI-Gen studies.

7. Ethics

AWI-Gen sub-study ethics clearance was submitted to the Human Research Ethics Committee (Medical) of the University of the Witwaterstrand on the 2nd of March 2022. Ethics clearance was granted conditionally on the 18th of March 2022 (ethics application reference number: M220297).

8. Possible Problems and Study Limitations

- To identify robust genetic associations between genetic variations and obesity and high BMI phenotypes, a larger sample size is required, as there are limitations regarding discovering novel associations with small or modest effect sizes (Ramsay et al., 2016). Furthermore, many associations that are identified between SNPs and microbial taxa are deemed to be no longer statistically significant once multiple testing is corrected for (Rothschild et al., 2018).
- GRSs are an effective method to quantify disease susceptibilities as a single measure, but it remains limited in its ability to capture the entire genetic loading of all variants affecting the susceptibility of complex multigenic disorders (Lewis & Vassos, 2020). Additionally, many other methods could be used to determine GRS, but due to the time constraints of this study, only one approach will be used.
- Based on the study participants included, all participants are female, which will limit confounding factors such as sex, but may result in generalized gender-biased findings. Subsequently, since the study participants from Bushbuckridge are from rural settlements, while those in Soweto are from peri-urban and urban settlements, there is potential for the confounding of rural versus urban microbiomes. The case and control study participants included also are not at an equal 1:1 ratio, which would be optimum. Instead, the ratio is approximately 5:1, based on which individuals were willing to participate in the microbiome pilot study.
- The 16S rRNA sequencing data generated in the microbiome pilot study will be analysed. It is cost-effective but can be adversely impacted by amplification bias and is not as comprehensive as metagenomic shotgun sequencing data, which will not be analysed due to time constraints (Brewster et al., 2019).
- Investigations like this that analyse interactions between host genetics and the GM are complicated. This is because the gut microbial composition is also influenced by environmental factors, such as lifestyle, exercise, diet as well as medication and antibiotic use, which are not accounted for in this study. Additionally, BMI will be the

only anthropometric measure that will incorporated in this study, as it is the only data that is available. Therefore, other possibly more informative anthropometric measures and their association with the human genome and gut microbiome will not be assessed.

References

Abdul-Aziz, M. A., Cooper, A., & Weyrich, L. S. (2016). Exploring relationships between host genome and microbiome: New insights from genome-wide association studies. *Frontiers in Microbiology*, *7*(Oct). Frontiers Media S.A. https://doi.org/10.3389/fmicb.2016.01611

Allali, I., Abotsi, R. E., Tow, L. A., Thabane, L., Zar, H. J., Mulder, N. M., et al. (2021). Human microbiota research in Africa: a systematic review reveals gaps and priorities for future research. *Microbiome*, *9*(1). https://doi.org/10.1186/s40168-021-01195-7

Brewster, R., Tamburini, F. B., Asiimwe, E., Oduaran, O., Hazelhurst, S., & Bhatt, A. S. (2019). Surveying Gut Microbiome Research in Africans: Toward Improved Diversity and Representation. *Trends in Microbiology*, *27*(10), 824–835. https://doi.org/10.1016/j.tim.2019.05.006

Chakraborti, C. K. (2015). New-found link between microbiota and obesity. *World Journal of Gastrointestinal Pathophysiology*, *6*(4), 110. https://doi.org/10.4291/wjgp.v6.i4.110

Chalazan, B., Palm, D., Sridhar, A., Lee, C., Argos, M., Daviglus, M., et al. (2021). Common genetic variants associated with obesity in an African-American and Hispanic/ Latino population. *PLoS ONE*, *16*(May). https://doi.org/10.1371/journal.pone.0250697

Cuevas-Sierra, A., Ramos-Lopez, O., Riezu-Boj, J. I., Milagro, F. I., & Martinez, J. A. (2019). Diet, Gut Microbiota, and Obesity: Links with Host Genetics and Epigenetics and Potential Applications. *Advances in Nutrition*, *10*, S17–S30. https://doi.org/10.1093/advances/nmy078

Cuevas-Sierra, A., Riezu-Boj, J. I., Guruceaga, E., Milagro, F. I., & Martínez, J. A. (2020). Sex-specific associations between gut prevotellaceae and host genetics on adiposity. *Microorganisms*, *8*(6), 1–19. https://doi.org/10.3390/microorganisms8060938

Davenport, E. R., Cusanovich, D. A., Michelini, K., Barreiro, L. B., Ober, C., et al. (2015). Genome-wide association studies of the human gut microbiota. *PLoS ONE*, *10*(11). https://doi.org/10.1371/journal.pone.0140301

Di Segni, A., Braun, T., Benshoshan, M., Barhom, S. F., Saar, E. G., Cesarkas, K., et al. (2018). Guided protocol for fecal microbial characterization by 16S rRNA-amplicon sequencing. *Journal of Visualized Experiments*, *2018*(133). https://doi.org/10.3791/56845

Goodarzi, M. O. (2018). Genetics of obesity: what genetic association studies have taught us about the biology of obesity and its complications. *The Lancet Diabetes and Endocrinology*, *6*(3), 223–236. https://doi.org/10.1016/S2213-8587(17)30200-0

Hall, A. B., Tolonen, A. C., & Xavier, R. J. (2017). Human genetic variation and the gut microbiome in disease. *Nature Reviews Genetics*, *18*(11), 690–699. https://doi.org/10.1038/nrg.2017.63

Khera, A. v., Chaffin, M., Wade, K. H., Zahid, S., Brancale, J., Xia, R., et al. (2019). Polygenic Prediction of Weight and Obesity Trajectories from Birth to Adulthood. *Cell*, *177*(3), 587-596.e9. https://doi.org/10.1016/j.cell.2019.03.028

Kim, B., Choi, H. N., & Yim, J. E. (2019). Effect of diet on the gut microbiota associated with obesity. *Journal of Obesity and Metabolic Syndrome*, *28*(4), 216–224. https://doi.org/10.7570/JOMES.2019.28.4.216

Lewis, C. M., & Vassos, E. (2020). Polygenic risk scores: From research tools to clinical instruments. *Genome Medicine*, *12*(1). https://doi.org/10.1186/s13073-020-00742-5

Luca, F., Kupfer, S. S., Knights, D., Khoruts, A., & Blekhman, R. (2018). Functional Genomics of Host–Microbiome Interactions in Humans. *Trends in Genetics*, 34(1), 30–40). https://doi.org/10.1016/j.tig.2017.10.001

Oduaran, O. H., Tamburini, F. B., Sahibdeen, V., Brewster, R., Gómez-Olivé, F. X., Kahn, K., et al. (2020). Gut microbiome profiling of a rural and urban South African cohort reveals biomarkers of a population in lifestyle transition. *BMC Microbiology*, *20*(1). https://doi.org/10.1186/s12866-020-02017-w

Ramsay, M., Crowther, N., Tambo, E., Agongo, G., Baloyi, V., Dikotope, S., et al. (2016). H3Africa

AWI-Gen Collaborative Centre: A resource to study the interplay between genomic and environmental risk factors for cardiometabolic diseases in four sub-Saharan African countries. *Global Health, Epidemiology and Genomics*, *1*. https://doi.org/10.1017/gheg.2016.17

Rothschild, D., Weissbrod, O., Barkan, E., Kurilshikov, A., Korem, T., Zeevi, D., et al. (2018). Environment dominates over host genetics in shaping human gut microbiota. *Nature*, *555*(7695), 210–215. https://doi.org/10.1038/nature25973

Sahibdeen, V., Crowther, N. J., Soodyall, H., Hendry, L. M., Munthali, R. J., Hazelhurst, S., et al. (2018). Genetic variants in SEC16B are associated with body composition in black South Africans. *Nutrition and Diabetes*, *8*(1). https://doi.org/10.1038/s41387-018-0050-0

Schlauch, K. A., Kulick, D., Subramanian, K., de Meirleir, K. L., Palotás, A., & Lombardi, V. C. (2019). Single-nucleotide polymorphisms in a cohort of significantly obese women without cardiometabolic diseases. *International Journal of Obesity*, *43*(2), 253–262. https://doi.org/10.1038/s41366-018-0181-3

Stephens, R. W., Arhire, L., & Covasa, M. (2018). Gut Microbiota: From Microorganisms to Metabolic Organ Influencing Obesity. *Obesity, 26*(5), 801–809). https://doi.org/10.1002/oby.22179

Thaker, V. V. (2017). Genetic and Epigenetic Causes of Obesity. *Adolescent Medicine: State of the Art Reviews*, *28*(2), 379–405.

Yako, Y. Y., Echouffo-Tcheugui, J. B., Balti, E. V., Matsha, T. E., Sobngwi, E., Erasmus, R. T., et al. (2015). Genetic association studies of obesity in Africa: A systematic review. *Obesity Reviews*, *16*(3), 259–272. https://doi.org/10.1111/obr.12260

Zarza-Rebollo, J. A., Molina, E., & Rivera, M. (2021). The role of the FTO gene in the relationship between depression and obesity. A systematic review. *Neuroscience and Biobehavioral Reviews*, *127*, 630–637. https://doi.org/10.1016/j.neubiorev.2021.05.013

Appendix One

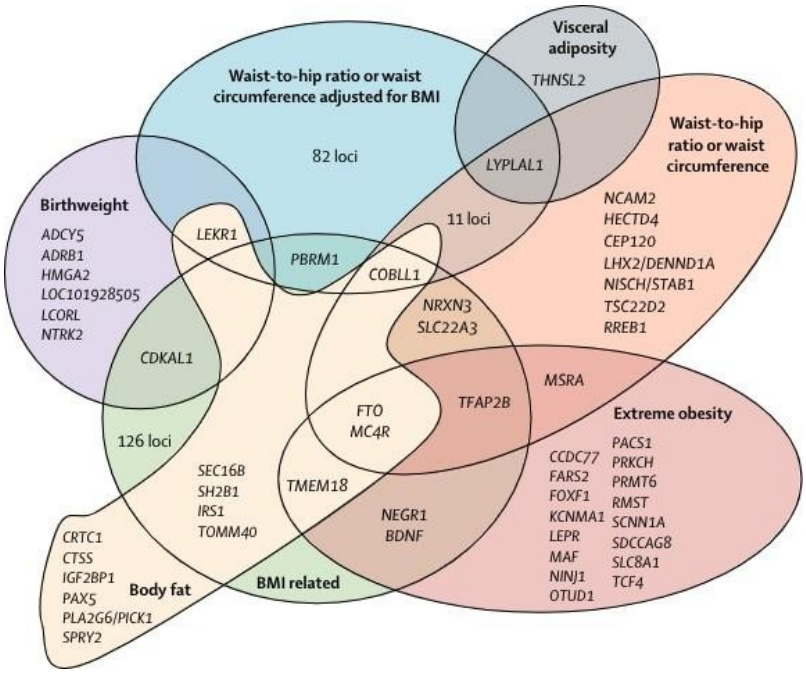


Figure A1. Genes identified to be associated with obesity-related traits through GWAS (Goodarzi, 2018).

A

Appendix Two: Ethics Clearance

UNIVERSITY OF THE WITWATERSRAND JOHANNESBURG	HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

Office of the Deputy Vice-Chancellor (Research and Innovation)

TO:	Ms S Schnell School of Pathology Division of Human Genetics Medical School University E-mail: sammyschnell@gmail.com
CC:	Supervisor: Dr O Oduaran and Professor S Hazelhurst <Ovokeraye.Oduaran@wits.ac.za> and <HREC-Medical Research Office@wits.ac.za>
FROM:	Mr Iain Burns Human Research Ethics Committee (Medical) Tel: 011 717 1252 E-mail: Iain.Burns@wits.ac.za
DATE:	2022/03/18
REF:	R14/49
PROTOCOL NO:	**M220297** (This is your ethics application reference number. Please quote it in all enquiries, oral or written, relating to this study.)
PROJECT TITLE:	*Exploring the interaction of host genetics and gut microbiome in obesity*

Please find attached the Clearance Certificate for the above project. I hope it goes well and that an article in a recognized publication comes out of it. This will reflect well on your professional standing and contribute to Government funding of the University.

MSWorks2000/Iain0007/Clearscan.wps

B

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

R49 Ms S Schnell

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)
CLEARANCE CERTIFICATE NO. M220297

NAME: (Principal Investigator)	Ms S Schnell
DEPARTMENT:	School of Pathology Division of Human Genetics Medical School University
PROJECT TITLE:	*Exploring the interaction of host genetics and gut microbiome in obesity*
DATE CONSIDERED:	Ad hoc
DECISION:	Approved conditionally
CONDITIONS:	Sub-study under M160121, M170880 and M180535 The HREC requires evidence of Faculty protocol approval
NOTE:	If contact information regarding student study participants is required, please contact the Registrar's office - <Nicoleen.Potgieter@wits.ac.za>
SUPERVISOR:	Dr O Oduaran and Professor S Hazelhurst
APPROVED BY:	Dr CB Penny, Chairperson, HREC (Medical)
DATE OF APPROVAL:	2022/03/18

This Clearance Certificate is valid for 5 years from the date of approval. An extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office secretariat on the 3rd floor, Phillip Tobias Building, Parktown, University of the Witwatersrand, Johannesburg.

I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated from the research protocol as approved, I/we undertake to submit details to the Committee. **I agree to submit a yearly progress report**. When a funder requires annual re-certification, the application date will be one year after the date when the study was initially reviewed. In this case, the study was initially reviewed in **February** and therefore reports and re-certification will be due in the month of **February** each year. Unreported changes to the study may invalidate the clearance given by the HREC (Medical).

Signature of Principal Investigator

20/03/2022
Date

C

Appendix Three: Turnitin Originality Report for Original Literature Review

Samantha Schnell_Literature Review.pdf

ORIGINALITY REPORT

2%	2%	1%	0%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Fumitaka Inoue, Walter L. Eckalbar, Yi Wang, Karl K. Murphy, Navneet Matharu, Christian Vaisse, Nadav Ahituv. "Genomic and epigenomic mapping of leptin-responsive neuronal populations involved in body weight regulation", Cold Spring Harbor Laboratory, 2019 Publication	1%
2	aging-us.com Internet Source	1%
3	lib.bioinfo.pl Internet Source	<1%
4	www.medrxiv.org Internet Source	<1%
5	www.advances.umed.wroc.pl Internet Source	<1%

Exclude quotes	On	Exclude matches	< 8 words
Exclude bibliography	On		

D

Appendix B: Ethics Clearance

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

R49 Ms S Schnell

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)
CLEARANCE CERTIFICATE NO. M220297

NAME: **(Principal Investigator)**	Ms S Schnell
DEPARTMENT:	School of Pathology Division of Human Genetics Medical School University
PROJECT TITLE:	*Exploring the interaction of host genetics and gut microbiome in obesity*
DATE CONSIDERED:	Ad hoc
DECISION:	Approved unconditionally
CONDITIONS:	Sub-study under M160121, M170880 and M180535 Previous condition (Faculty approval) satisfied on 2022/07/06
NOTE:	If contact information regarding student study participants is required, please contact the Registrar's office - <Nicoleen.Potgieter@wits.ac.za>
SUPERVISOR:	Dr O Oduaran and Professor S Hazelhurst
APPROVED BY:	CB Penny Dr CB Penny, Chairperson, HREC (Medical)
DATE OF APPROVAL:	2022/03/18

This Clearance Certificate is valid for 5 years from the date of approval. An extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office secretariat on the 3rd floor, Phillip Tobias Building, Parktown, University of the Witwatersrand, Johannesburg.

I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated from the research protocol as approved, I/we undertake to submit details to the Committee. **I agree to submit a yearly progress report**. When a funder requires annual re-certification, the application date will be one year after the date when the study was initially reviewed. In this case, the study was initially reviewed in **February** and therefore reports and re-certification will be due in the month of **February** each year. Unreported changes to the study may invalidate the clearance given by the HREC (Medical).

Signature of Principal Investigator

29/09/2022
Date

Appendix C: Turnitin Originality Report

Samantha Schnell_Research Report-3.pdf

ORIGINALITY REPORT

8%	5%	7%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.researchgate.net Internet Source	1%
2	www.science.gov Internet Source	<1%
3	Premathilake, Hasitha Upendra. "Elucidation of Structural and Functional Characteristics of the Gut Microbiome of Beef Cattle under Water Stress", Oklahoma State University, 2022 Publication	<1%
4	www.ncbi.nlm.nih.gov Internet Source	<1%
5	samj.org.za Internet Source	<1%
6	bmcmicrobiol.biomedcentral.com Internet Source	<1%
7	www.frontiersin.org Internet Source	<1%

Oduaran

Dr Ovokeraye Oduaran
Supervisor
Date: Dec 12, 2022

Scott Hazelhurst (Dec 12, 2022 13:49 GMT+2)

Prof Scott Hazelhurst
Co-Supervisor
Date: Dec 12, 2022

Appendix D: Frontiers Author Guidelines

Article Types

Original Research
Original Research articles report on primary and unpublished studies. Original Research may also encompass confirming studies and disconfirming results which allow hypothesis elimination, reformulation and/or report on the non-reproducibility of previously published results. Original Research articles are peer-reviewed, have a maximum word count of 12,000 and may contain no more than 15 Figures/Tables. Authors are required to pay a fee (A-type article) to publish an Original Research article. Original Research articles should have the following format: 1) Abstract, 2) Introduction, 3) Materials and Methods, 4) Results, 5) Discussion.

Author Guidelines

General standards

Article type

Frontiers requires authors to select the appropriate article type for their manuscript and to comply with the article type descriptions defined in the journal's 'Article types' page, which can be seen from the 'For authors' menu on every Frontiers journal page. Please pay close attention to the word count limits.

Templates

If working with Word please use our Word templates. If you wish to submit your article as LaTeX, we recommend our LaTeX templates.
For LaTeX files, please ensure all relevant manuscript files are uploaded: .tex file, PDF, and .bib file (if the bibliography is not already included in the .tex file).
During the interactive review, authors are encouraged to upload versions using track changes. Editors and reviewers can only download the PDF file of the submitted manuscript.

Manuscript length

Frontiers encourages the authors to closely follow the article word count lengths given in the 'Article types' page of the journals. The manuscript length includes only the main body of the text, footnotes, and all citations within it, and excludes the abstract, section titles, figure and table captions, funding statement, acknowledgments, and references in the bibliography. Please indicate the number of words and the number of figures and tables included in your manuscript on the first page.

Language editing

Frontiers requires manuscripts submitted to meet international English language standards to be considered for publication.
For authors who would like their manuscript to receive language editing or proofreading to improve the clarity of the manuscript and help highlight their research, Frontiers recommends the language-editing services provided by the following external partners.

Note that sending your manuscript for language editing does not imply or guarantee that it will be accepted for publication by a Frontiers journal. Editorial decisions on the scientific content of a manuscript are independent of whether it has received language editing or proofreading by these partner services or other services.

Editage
Frontiers recommends the language-editing service provided by our external partner Editage. These services may be particularly useful for researchers for whom English is not the primary language. They can help to improve the grammar, syntax, and flow of your manuscript prior to submission. Frontiers authors will receive a 10% discount by visiting the following link: editage.com/frontiers.

The Charlesworth Group

Frontiers recommends the Charlesworth Group's author services, who has a long-standing track record in language editing and proofreading. This is a third-party service for which Frontiers authors will receive a 10% discount by visiting the following link: www.cwauthors.com/frontiers.

Frontiers推荐您使用在英语语言编辑和校对领域具有您久历史和良好口碑的查尔斯沃思作者服务。此项服务由第三方为您提供，Frontiers中国作者通过此链接提交稿件时可获得

10%的特别优惠: www.cwauthors.com.cn/frontiers.

Language style

The default language style at Frontiers is American English. If you prefer your article to be formatted in British English, please specify this on the first page of your manuscript. For any questions regarding style, Frontiers recommends authors to consult the Chicago Manual of Style.

Search engine optimization (SEO)

There are a few simple ways to maximize your article's discoverability and search results.

- Include a few of your article's keywords in the title of the article
- Do not use long article titles
- Pick 5-8 keywords using a mix of generic and more specific terms on the article subject(s)
- Use the maximum amount of keywords in the first two sentences of the abstract
- Use some of the keywords in level 1 headings

CrossMark policy

Title

The title should be concise, omitting terms that are implicit and, where possible, be a statement of the main result or conclusion presented in the manuscript. Abbreviations should be avoided within the title.

Witty or creative titles are welcome, but only if relevant and within measure. Consider if a title meant to be thought-provoking might be misinterpreted as offensive or alarming. In extreme cases, the editorial office may veto a title and propose an alternative.
Authors should avoid:

- titles that are a mere question without giving the answer
- unambitious titles, for example starting with 'Towards,' 'A description of,' 'A characterization of' or 'Preliminary study on'
- vague titles, for example starting with 'Role of', 'Link between', or 'Effect of' that do not specify the role, link, or effect
- including terms that are out of place, for example the taxonomic affiliation apart from species name.

For Corrigenda, General Commentaries, and Editorials, the title of your manuscript should have the following format.

- 'Corrigendum: [Title of original article]'
- General Commentaries:
  'Commentary: [Title of original article]'
  'Response: Commentary: [Title of original article]'
- 'Editorial: [Title of Research Topic]'

The running title should be a maximum of five words in length.

Authors and affiliations

All names are listed together and separated by commas. Provide exact and correct author names as these will be indexed in official archives. Affiliations should be keyed to the author's name with superscript numbers and be listed as follows:

- Laboratory, Institute, Department, Organization, City, State abbreviation (only for United States, Canada, and Australia), and Country (without detailed address information such as city zip codes or street names).

Example: Max Maximus1
1 Department of Excellence, International University of Science, New York, NY, United States.

Correspondence

The corresponding author(s) should be marked with an asterisk in the author list. Provide the exact contact email address of the corresponding author(s) in a separate section.
Example: Max Maximus*
maximus@iuscience.edu
If any authors wish to include a change of address, list the present address(es) below the correspondence details using a unique superscript symbol keyed to the author(s) in the author list.

Equal contributions

The authors who have contributed equally should be marked with a symbol (†) in the author list of the doc/latex and pdf files of the manuscript uploaded at submission.
Please use the appropriate standard statement(s) to indicate equal contributions:

- **Equal contribution**: These authors contributed equally to this work
- **First authorship**: These authors share first authorship

- **Senior authorship**: These authors share senior authorship
- **Last authorship**: These authors share last authorship
- **Equal contribution and first authorship**: These authors contributed equally to this work and share first authorship
- **Equal contribution and senior authorship**: These authors contributed equally to this work and share senior authorship
- **Equal contribution and last authorship**: These authors contributed equally to this work and share last authorship

Example: Max Maximus 1†, John Smith2† and Barbara Smith1
†These authors contributed equally to this work and share first authorship

Consortium/group and collaborative authors

Consortium/group authorship should be listed in the manuscript with the other author(s).

In cases where authorship is retained by the consortium/group, the consortium/group should be listed as an author separated by a comma or 'and'. The consortium/group name will appear in the author list, in the citation, and in the copyright. If provided, the consortium/group members will be listed in a separate section at the end of the article.

For the collaborators of the consortium/group to be indexed in PubMed, they do not have to be inserted in the Frontiers submission system individually. However, in the manuscript itself, provide a section with the name of the consortium/group as the heading followed by the list of collaborators, so they can be tagged accordingly and indexed properly.

Example: John Smith, Barbara Smith and The Collaborative Working Group.
In cases where work is presented by the author(s) on behalf of a consortium/group, it should be included in the author list separated with the wording 'for' or 'on behalf of.' The consortium/group will not retain authorship and will only appear in the author list.
Example: John Smith and Barbara Smith on behalf of The Collaborative Working Group.

Abstract

As a primary goal, the abstract should render the general significance and conceptual advance of the work clearly accessible to a broad readership. In the abstract, minimize the use of abbreviations and do not cite references, figures or tables.

For Clinical Trial articles, please include the Unique Identifier and the URL of the publicly accessible website on which the trial is registered.

Keywords
All article types require a minimum of five and a maximum of eight keywords.

Text

The entire document should be single-spaced and must contain page and line numbers in order to facilitate the review process. The manuscript should be written using either Word or LaTeX. See above for templates.

Nomenclature

The use of abbreviations should be kept to a minimum. Non-standard abbreviations should be avoided unless they appear at least four times, and must be defined upon first use in the main text. Consider also giving a list of non-standard abbreviations at the end, immediately before the acknowledgments.

Equations should be inserted in editable format from the equation editor.
Italicize gene symbols and use the approved gene nomenclature where it is available. For human genes, please refer to the HUGO Gene Nomenclature Committee (HGNC). New symbols for human genes should be submitted to the HGNC here. Common alternative gene aliases may also be reported, but should not be used alone in place of the HGNC symbol. Nomenclature committees for other species are listed here. Protein products are not italicized.

We encourage the use of Standard International Units in all manuscripts.
Chemical compounds and biomolecules should be referred to using systematic nomenclature, preferably using the recommendations by the International Union of Pure and Applied Chemistry (IUPAC).

Astronomical objects should be referred to using the nomenclature given by the International Astronomical Union (IAU) provided here.

Life Science Identifiers (LSIDs) for ZOOBANK registered names or nomenclatural acts should be listed in the manuscript before the keywords. An LSID is represented as a uniform resource name (URN) with the following format:
urn:lsid:<Authority>:<Namespace>:<ObjectID>[:<Version>]
For more information on LSIDs please see the 'Code' section of our polices and publication ethics.

Sections

The manuscript is organized by headings and subheadings. The section headings should be those appropriate for your field and the research itself. You may insert up to 5 heading levels into your manuscript (i.e.,: 3.2.2.1.2 Heading Title).
For Original Research articles, it is recommended to organize your manuscript in the following sections or their equivalents for your field.

Introduction

Succinct, with no subheadings.

Materials and methods

This section may be divided by subheadings and should contain sufficient detail so that when read in conjunction with cited references, all procedures can be repeated. For experiments reporting results on animal or human subject research, an ethics approval statement should be included in this section (for further information, see the 'Bioethics' section of our polices and publication ethics.)

Results

This section may be divided by subheadings. Footnotes should not be used and must be transferred to the main text.

Discussion

This section may be divided by subheadings. Discussions should cover the key findings of the study: discuss any prior research related to the subject to place the novelty of the discovery in the appropriate context, discuss the potential shortcomings and limitations on their interpretations, discuss their integration into the current understanding of the problem and how this advances the current views, speculate on the future direction of the research, and freely postulate theories that could be tested in the future.

For further information, please check the descriptions defined in the journal's 'Article types' page, in the 'For authors' menu on every journal page.

Acknowledgements

This is a short text to acknowledge the contributions of specific colleagues, institutions, or agencies that aided the efforts of the authors. Should the content of the manuscript have previously appeared online, such as in a thesis or preprint, this should be mentioned here, in addition to listing the source within the reference list.

Contribution to the field statement

When you submit your manuscript, you will be required to briefly summarize in 200 words your manuscript's contribution to, and position in, the existing literature in your field. This should be written avoiding any technical language or non-standard acronyms. The aim should be to convey the meaning and importance of this research to a non-expert.

While Frontiers evaluates articles using objective criteria, rather than impact or novelty, your statement should frame the question(s) you have addressed in your work in the context of the current body of knowledge, providing evidence that the findings – whether positive or negative – contribute to progress in your research discipline. This will help the chief editors to determine whether your manuscript fits within the scope of a specialty as defined in its mission statement; a detailed statement will also facilitate the identification of the editors and reviewers most appropriate to evaluate your work, ultimately expediting your manuscript's initial consideration.

Example statement on: Markram K and Markram H (2010) The Intense World Theory – a unifying theory of the neurobiology of autism. Front. Hum. Neurosci. 4:224. doi: 10.3389/fnhum.2010.00224

Autism spectrum disorders are a group of neurodevelopmental disorders that affect up to 1 in 100 individuals. People with autism display an array of symptoms encompassing emotional processing, sociability, perception and memory, and present as uniquely as the individual. No theory has suggested a single underlying neuropathology to account for these diverse symptoms. The Intense World Theory, proposed here, describes a unifying pathology producing the wide spectrum of manifestations observed in autists. This theory focuses on the neocortex, fundamental for higher cognitive functions, and the limbic system, key for

processing emotions and social signals. Drawing on discoveries in animal models and neuroimaging studies in individuals with autism, we propose how a combination of genetics, toxin exposure and/or environmental stress could produce hyper-reactivity and hyper-plasticity in the microcircuits involved with perception, attention, memory and emotionality. These hyper-functioning circuits will eventually come to dominate their neighbors, leading to hyper-sensitivity to incoming stimuli, over-specialization in tasks and a hyper-preference syndrome. We make the case that this theory of enhanced brain function in autism explains many of the varied past results and resolves conflicting findings and views and makes some testable experimental predictions.

Figure and table guidelines

CC-BY license

All figures, tables, and images will be published under a Creative Commons CC-BY license, and permission must be obtained for use of copyrighted material from other sources (including re-published/adapted/modified/partial figures and images from the internet). It is the responsibility of the authors to acquire the licenses, follow any citation instructions requested by third-party rights holders, and cover any supplementary charges.

For additional information, please see the 'Image manipulation' section of our polices and publication ethics.

Figure requirements and style guidelines

Frontiers requires figures to be submitted individually, in the same order as they are referred to in the manuscript; the figures will then be automatically embedded at the end of the submitted manuscript. Kindly ensure that each figure is mentioned in the text and in numerical order.

For figures with more than one panel, panels should be clearly indicated using labels (A), (B), (C), (D), etc. However, do not embed the part labels over any part of the image, these labels will be replaced during typesetting according to Frontiers' journal style. For graphs, there must be a self-explanatory label (including units) along each axis.

For LaTeX files, figures should be included in the provided PDF. In case of acceptance, our production office might require high-resolution files of the figures included in the manuscript in EPS, JPEG or TIF/TIFF format.

To upload more than one figure at a time, save the figures (labeled in order of appearance in the manuscript) in a zip file and upload them as 'Supplementary Material Presentation.'

Please note that figures not in accordance with the guidelines will cause substantial delay during the production process.

Captions

Captions should be preceded by the appropriate label, for example 'Figure 1.' Figure captions should be placed at the end of the manuscript. Figure panels are referred to by bold capital letters in brackets: (A), (B), (C), (D), etc.

Image size and resolution requirements
Figures should be prepared with the PDF layout in mind. Individual figures should not be longer than one page and with a width that corresponds to 1 column (85 mm) or 2 columns (180 mm).

All images must have a resolution of 300 dpi at final size. Check the resolution of your figure by enlarging it to 150%. If the image appears blurry, jagged, or has a stair-stepped effect, the resolution is too low.

The text should be legible and of high quality. The smallest visible text should be no less than eight points in height when viewed at actual size.

Solid lines should not be broken up. Any lines in the graphic should be no smaller than two points wide.

Please note that saving a figure directly as an image file (JPEG, TIF) can greatly affect the resolution of your image. To avoid this, one option is to export the file as PDF, then convert into TIFF or EPS using a graphics software.

Format and color image mode

The following formats are accepted: TIF/TIFF (.tif/.tiff), JPEG (.jpg), and EPS (.eps) (upon acceptance). Images must be submitted in the color mode RGB.

Chemical structures

Chemical structures should be prepared using ChemDraw or a similar program. If working with ChemDraw please use our ChemDraw template. If working with another program please follow the guidelines below.

- Drawing settings: chain angle, 120° bond spacing, 18% width; fixed length, 14.4 pt; bold width, 2.0 pt; line width, 0.6 pt; margin width, 1.6 pt; hash spacing, 2.5 pt. Scale 100% Atom Label settings: font, Arial; size, 8 pt
- Assign all chemical compounds a bold, Arabic numeral in the order in which the compounds are presented in the manuscript text.

Table requirements and style guidelines

Tables should be inserted at the end of the manuscript in an editable format. If you use a word processor, build your table in Word. If you use a LaTeX processor, build your table in LaTeX. An empty line should be left before and after the table.

Table captions must be placed immediately before the table. Captions should be preceded by the appropriate label, for example 'Table 1.' Please use only a single paragraph for the caption.

Ensure that each table is mentioned in the text and in numerical order.

Large tables covering several pages cannot be included in the final PDF for formatting reasons. These tables will be published as supplementary material.

Tables which are not according to the above guidelines will cause substantial delay during the production process.

Accessibility

We encourage authors to make the figures and visual elements of their articles accessible for the visually impaired. An effective use of color can help people with low visual acuity, or color blindness, understand all the content of an article.

These guidelines are easy to implement and are in accordance with the W3C Web Content Accessibility Guidelines (WCAG 2.1), the standard for web accessibility best practices.

Ensure sufficient contrast between text and its background

People who have low visual acuity or color blindness could find it difficult to read text with low contrast background color. Try using colors that provide maximum contrast.

WC3 recommends the following contrast ratio levels:

- Level AA, contrast ratio of at least 4.5:1
- Level AAA, contrast ratio of at least 7:1

You can verify the contrast ratio of your palette with these online ratio checkers:

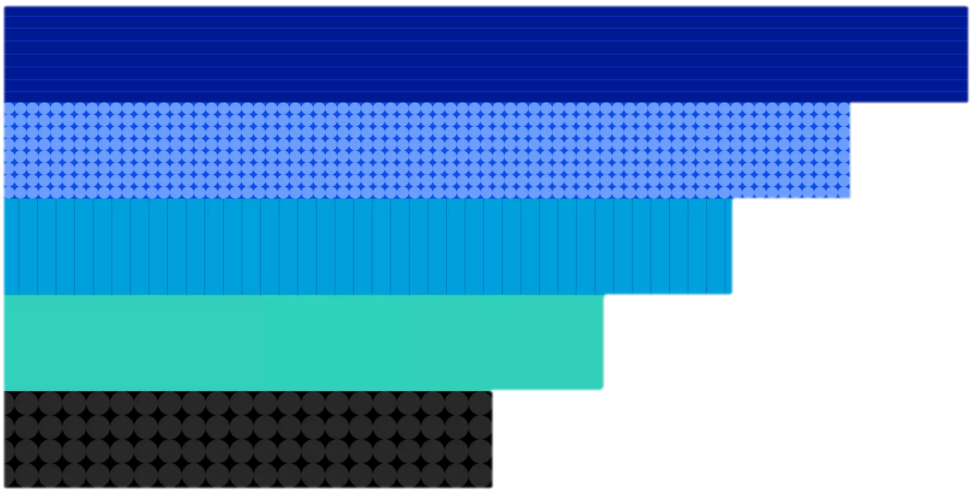
- WebAIM
- Color Safe

Avoid using red or green indicators

More than 99% of color-blind people have a red-green color vision deficiency.

Avoid using only color to communicate information

Elements with complex information like charts and graphs can be hard to read when only color is used to distinguish the data. Try to use other visual aspects to communicate information, such as shape, labels, and size. Incorporating patterns into the shape fills also make differences clearer; for an example please see below:



Supplementary material

Data that are not of primary importance to the text, or which cannot be included in the article because they are too large or the current format does not permit it (such as videos, raw data traces, and PowerPoint presentations), can be uploaded as supplementary material during the submission procedure and will be displayed along with the published article. All supplementary files are deposited to figshare for permanent storage and receive a DOI.

Supplementary material is not typeset, so please ensure that all information is clearly presented without tracked changes/highlighted text/line numbers, and the appropriate caption is included in the file. To avoid discrepancies between the published article and the supplementary material, please do not add the title, author list, affiliations or correspondence in the supplementary files.
The supplementary material can be uploaded as:

- data sheet (Word, Excel, CSV, CDX, FASTA, PDF or Zip files)
- presentation (PowerPoint, PDF or Zip files)
- image (CDX, EPS, JPEG, PDF, PNG or TIF/TIFF),
- table (Word, Excel, CSV or PDF)
- audio (MP3, WAV or WMA)
- video (AVI, DIVX, FLV, MOV, MP4, MPEG, MPG or WMV).

Technical requirements for supplementary images:

- 300 DPIs
- RGB color mode.

For supplementary material templates (LaTeX and Word), see our supplementary material templates.

References

Frontiers' journals use one of two reference styles, either Harvard (author-date) or Vancouver (numbered). Please check our help center to find the correct style for the journal to which you are submitting.

- All citations in the text, figures, or tables must be in the reference list and vice-versa
- The names of the first six authors followed by et al. and the DOI (when available) should be provided
- Given names of authors should be abbreviated to initials (e.g., Smith, J., Lewis, C.S., etc.)
- The reference list should only include articles that are published or accepted
- Unpublished data, submitted manuscripts, or personal communications should be cited within the text only, for article types that allow such inclusions
- For accepted but unpublished works use 'in press' instead of page numbers
- Data sets that have been deposited to an online repository should be included in the reference list. Include the version and unique identifier when available
- Personal communications should be documented by a letter of permission
- Website URLs should be included as footnotes
- Any inclusion of verbatim text must be contained in quotation marks and clearly reference the original source
- Preprints can be cited as long as a DOI or archive URL is available, and the citation clearly mentions that the contribution is a preprint. If a peer-reviewed journal publication for the same preprint exists, the official journal publication is the preferred source. See the preprints section for each reference style below for more information.

Harvard reference style (author-date)

Many Frontiers journals use the Harvard referencing system; to find the correct reference style and resources for the journal you are submitting to, please visit our help center. Reference examples are found below, for more examples of citing other documents and general questions regarding the Harvard reference style, please refer to the Chicago Manual of Style.

In-text citations

- For works by a single author, include the surname, followed by the year
- For works by two authors, include both surnames, followed by the year
- For works by more than two authors, include only the surname of the first author followed by et al., followed by the year
- For humanities and social sciences articles, include the page numbers.

Reference list examples

Article in a print journal
Sondheimer, N., and Lindquist, S. (2000). Rnq1: an epigenetic modifier of protein function in yeast. Mol. Cell. 5, 163-172.

Article in an online journal

Tahimic, C.G.T., Wang, Y., Bikle, D.D. (2013). Anabolic effects of IGF-1 signaling on the skeleton. Front. Endocrinol. 4:6. doi: 10.3389/fendo.2013.00006

Article or chapter in a book
Sorenson, P. W., and Caprio, J. C. (1998). "Chemoreception," in The Physiology of Fishes, ed. D. H. Evans (Boca Raton, FL: CRC Press), 375-405.

Book
Cowan, W. M., Jessell, T. M., and Zipursky, S. L. (1997). Molecular and Cellular Approaches to Neural Development. New York: Oxford University Press.

Abstract
Hendricks, J., Applebaum, R., and Kunkel, S. (2010). A world apart? Bridging the gap between theory and applied social gerontology. Gerontologist 50, 284-293. Abstract retrieved from Abstracts in Social Gerontology database. (Accession No. 50360869)

Website
World Health Organization. (2018). E. coli. https://www.who.int/news-room/fact-sheets/detail/e-coli [Accessed March 15, 2018].

Patent
Marshall, S. P. (2000). Method and apparatus for eye tracking and monitoring pupil dilation to evaluate cognitive activity. U.S. Patent No 6,090,051. Washington, DC: U.S. Patent and Trademark Office.

Data
Perdiguero P, Venturas M, Cervera MT, Gil L, Collada C. Data from: Massive sequencing of Ulms minor's transcriptome provides new molecular tools for a genus under the constant threat of Dutch elm disease. Dryad Digital Repository. (2015) http://dx.doi.org/10.5061/dryad.ps837

Theses and dissertations
Smith, J. (2008) Post-structuralist discourse relative to phenomological pursuits in the deconstructivist arena. [dissertation/master's thesis]. [Chicago (IL)]: University of Chicago

Preprint
Smith, J. (2008). Title of the document. Preprint repository name [Preprint]. Available at: https://persistent-url (Accessed March 15, 2018).

Vancouver reference style (numbered)

Many Frontiers journals use the numbered referencing system; to find the correct reference style and resources for the journal you are submitting to, please visit our help center.

Reference examples are found below, for more examples of citing other documents and general questions regarding the Vancouver reference style, please refer to Citing Medicine.

In-text citations

- Please apply the Vancouver system for in-text citations

- In-text citations should be numbered consecutively in order of appearance in the text – identified by Arabic numerals in the parenthesis (use square brackets for physics and mathematics articles).

Reference list examples

Article in a print journal
Sondheimer N, Lindquist S. Rnq1: an epigenetic modifier of protein function in yeast. Mol Cell (2000) 5:163-72.

Article in an online journal
Tahimic CGT, Wang Y, Bikle DD. Anabolic effects of IGF-1 signaling on the skeleton. Front Endocrinol (2013) 4:6. doi: 10.3389/fendo.2013.00006

Article or chapter in a book
Sorenson PW, Caprio JC. "Chemoreception". In: Evans DH, editor. The Physiology of Fishes. Boca Raton, FL: CRC Press (1998). p. 375-405.

Book
Cowan WM, Jessell TM, Zipursky SL. Molecular and Cellular Approaches to Neural Development. New York: Oxford University Press (1997). 345 p.

Abstract
Christensen S, Oppacher F. An analysis of Koza's computational effort statistic for genetic programming. In: Foster JA, editor. Genetic Programming. EuroGP 2002: Proceedings of the 5th European Conference on Genetic Programming; 2002 Apr 3–5; Kinsdale, Ireland. Berlin: Springer (2002). p. 182–91.

Website
World Health Organization. E. coli (2018). https://www.who.int/news-room/fact-sheets/detail/e-coli [Accessed March 15, 2018].

Patent
Pagedas AC, inventor; Ancel Surgical R&D Inc., assignee. Flexible Endoscopic Grasping and Cutting Device and Positioning Tool Assembly. United States patent US 20020103498 (2002).

Data
Perdiguero P, Venturas M, Cervera MT, Gil L, Collada C. Data from: Massive sequencing of Ulms minor's transcriptome provides new molecular tools for a genus under the constant threat of Dutch elm disease. Dryad Digital Repository. (2015) http://dx.doi.org/10.5061/dryad.ps837

Theses and dissertations
Smith, J. (2008) Post-structuralist discourse relative to phenomological pursuits in the deconstructivist arena. [dissertation/master's thesis]. [Chicago (IL)]: University of Chicago

Preprint
Smith, J. Title of the document. Preprint repository name [Preprint] (2008). Available at: https://persistent-url (Accessed March 15, 2018)