UNIVERSITY OF TARTU

Faculty of Science and Technology

Institute of Bioengineering

Stepan Strelchenko

Genetic and environmental interactions in human obesity

Master's Thesis (30 ECTS)

Curriculum Bioengineering

Supervisor:

Associate Professor, PhD Jon Anders Eriksson

Tartu 2025

Title: Genetic and environmental interactions in human obesity

Abstract

Obesity is a rising global problem, clinically defined as having a body mass index (BMI) > 30 or a waist-to-hip ratio >0.9 for males and >0.85 for females. It is a complex disease with both genetic and environmental factors affecting its development and progression. However, although BMI is highly heritable (60-90% according to twin studies), only ~6% of the variance is explained by additive genetic variation discovered in large genome-wide association studies. Possible sources of unexplained BMI variance are gene-gene (GxG) and gene-environment (GxE) interactions, but to date, only a few studies have addressed this question, and mostly focused on the UK biobank and data from smaller longitudinal studies

This study used data from the Estonian biobank to investigate the potential of using the Brown-Forsythe (BF) test to detect heteroscedasticity in BMI based on the genotype and phenotype groups. Moreover, in this work, the BF test was used to select genotypes and environmental variables that were significantly associated with BMI and WHR variance. The study demonstrated sex and age-dependent differences in heteroscedasticity, with higher levels of heteroscedasticity in females compared to males and decreasing heteroscedasticity with increasing age for both sexes.

Next, the gene-environment interactions for the outcome variables BMI and WHR were modeled and studied for two genes (FTO and MC4R) and environmental factors, such as physical activity, smoking, and education level. The study found evidence for significant interaction of FTO with smoking for the age group of 31-50 years for males, along with significant interaction of FTO with physical activity for Females (18-30 age group) and males (31-50 age group) in the case of BMI as the outcome variable. Moreover, the evidence for significant interaction of FTO and physical activity was found for females (age group 18-30) for the WHR as the outcome variable. Furthermore, the significant interaction was observed in the case of MC4R and physical activity for BMI in males (18-30 year old group). Additionally, it was shown that the significant interaction between MC4R and education level for males (31-50 age group) was present for BMI.

Finally, the study outlined the ground for future research work in the area of gene-environment interactions and statistical modeling, indicating potential environmental factors that can be studied for possible gene-environment interactions.

Keywords: Biosciences and Environment, Genetics

CERCS: B220

Institute name: Institute of Genomics

Research group: cGEM

Inimese rasvumise geneetilised ja keskkonnamõjud

Lühikokkuvõte

Rasvumine on kasvav ülemaailmne probleem, mis on kliiniliselt määratletud kui kehamassiindeks (KMI) > 30 või kui meeste puhul on vöökoha ja puusa suhe > 0,9 ja naiste puhul > 0,85. Tegemist on keerulise haigusega, mille arengut ja kulgu mõjutavad nii geneetilised kui ka keskkonnategurid. Kuigi kehamassiindeks on väga hästi pärilduva iseloomuga (60-90% vastavalt kaksikuuringutele), on vaid ~6% variatsioonist seletatav additiivse geneetilise variatsiooniga, mis on avastatud suurte genoomi hõlmavate assotsiatsioonuuringute käigus. Selgitamata kehamassiindeksi varieeruvuse võimalikud allikad on geeni ja geeni ning geeni ja keskkonna (GxE) koostoimed, kuid siiani on seda küsimust käsitletud vaid vähestes uuringutes, mis on peamiselt keskendunud Ühendkuningriigi biopangale ja väiksemate pikisuunaliste uuringute andmetele.

Käesolevas uuringus kasutati Eesti biopanga andmeid, et uurida Brown-Forsythe'i (BF) testi kasutamise võimalusi kehamassiindeksi heteroskedastilisuse avastamiseks genotüübi ja fenotüübi rühmade alusel. Lisaks kasutati BF-testi selles töös genotüüpide ja keskkonnamuutujate valimiseks, mis olid oluliselt seotud BMI ja WHR varieeruvusega. Uuring näitas heteroskedastilisuse soost ja vanusest sõltuvaid erinevusi, kusjuures heteroskedastilisus oli naistel suurem kui meestel ja heteroskedastilisus vähenes vanuse kasvades mõlemal sugupoolel.

Seejärel modelleeriti ja uuriti geenide ja keskkonna vastastikmõju tulemusmuutujate BMI ja WHR puhul kahe geeni (FTO ja MC4R) ja keskkonnategurite, nagu füüsiline aktiivsus, suitsetamine ja haridustase, puhul. Uuringus leiti tõendeid FTO ja suitsetamise olulise koostoime kohta 31-50-aastaste meeste vanuserühmas ning FTO ja füüsilise aktiivsuse olulise koostoime kohta naiste (18-30-aastaste vanuserühm) ja meeste (31-50-aastaste vanuserühm) puhul, kui tulemusmuutujaks oli kehamassiindeks. Lisaks leiti tõendeid FTO ja füüsilise

aktiivsuse olulise koostoime kohta naiste (vanuserühm 18-30) puhul WHR kui tulemusmuutuja puhul. Lisaks täheldati MC4R ja füüsilise aktiivsuse olulist koostoimet BMI puhul meestel (18-30-aastaste rühm). Lisaks ilmnes, et MC4R ja haridustaseme vahel oli oluline koostoime meeste (31-50-aastaste vanuserühm) puhul kehamassiindeksi puhul.

Lõpuks visandati uuringus pinnas edasiseks uurimistööks geenide ja keskkonna koostoimete ja statistilise modelleerimise valdkonnas, näidates ära võimalikud keskkonnategurid, mida saab uurida võimalike geenide ja keskkonna koostoimete puhul.

Võtmesõnad:

Bioteadused ja keskkond, geneetika

CERCS: B220

TABLE OF CONTENTS

TERMS, ABBREVIATIONS, AND NOTATIONS **6**
INTRODUCTION **8**
1 LITERATURE REVIEW **10**
1.1 Biological background of obesity 10
1.2 Obesity treatment 12
1.2.1 Lifestyle changes 13
1.2.2 Pharmacological treatment of obesity 15
1.3 Genetics of obesity 17
1.3.1 Common genes related to BMI variation and obesity 18
1.4 Gene-environment interactions 19
1.5 Methods to detect variance differences across different study groups 23
2 THE AIMS OF THE THESIS **26**
3 EXPERIMENTAL PART **28**
3.1 MATERIALS AND METHODS 28
3.1.1 Data 28
3.1.1.1 Kaggle dataset 28
3.1.1.2 Dataset from the Estonian Biobank 28
3.1.2 Methods 29
3.1.2.1 Software and data processing 29
3.1.2.2 Variance analysis with the Brown-Forsythe test 30
3.1.2.3 Brown-Forsythe test validation 31
3.1.2.4 Selection of phenotypes and genotypes based on variance heterogeneity 32
3.1.2.5 Linear model to capture GxE 32
3.2 RESULTS 34
3.2.1 Validation of the BF test on the training dataset from Kaggle 34
3.2.2 Validation of the BF test on the Estonian Biobank data 35
3.2.3 EstBB dataset summary statistics 35
3.2.4 Genotype selection based on the BF test 36
3.2.5 Phenotypic traits' variance difference between defined groups 37
3.2.6 GxE of the strongest candidates from the phenotypic variables 40
3.2.6.1 FTO and physical activity 40
3.2.6.2 MC4R vs Physical activity 42
3.2.6.3 FTO and Smoking 44
3.2.6.4 MC4R and Smoking 46
3.2.6.5 FTO and Education level 47
3.2.6.6 MC4R and Education Level 49
3.3 DISCUSSION 52
3.3.1 The Brown-Forsythe test proved to be robust against skewed data 52
3.3.2 Difference in response to the environment between males and females across different age groups 52
3.3.3 Possible biological explanations of heteroscedasticity 52
3.3.4 Limitations of the study 53
3.3.5 Future perspectives 54

SUMMARY 55
REFERENCES 56
APPENDIX 63
NON-EXCLUSIVE LICENCE TO REPRODUCE THESIS AND MAKE THESIS PUBLIC 64

TERMS, ABBREVIATIONS, AND NOTATIONS

ACTH – adrenocorticotropic hormone

AgRP – agouti-related peptide

Alpha-MSH (α-MSH) – α-Melanocyte-stimulating hormone

ARC – arcuate nucleus

BBS – Bardet-Biedl syndrome

BF test – Brown-Forsythe test

BMI – body mass index

CCK – Cholecystokinin

CVD – cardiovascular disease

DA – dopamine

EMA – European Medicines Agency

EstBB – Estonian Biobank

EWL – excess weight loss

FDA – Federal Drug Administration

FGF-21 – Fibroblast growth factor 21

FTO – Fat mass and obesity-associated gene

GABA – gamma-aminobutyric acid

GxE – gene-environment interactions

GIP – gastric inhibitory polypeptide

GxG – gene-gene interactions

GLP-1 – Glucagon-like peptide-1

GWAS – genome-wide association study

LEP – leptin-coding gene

LEPR – leptin receptor

LM – linear model

MAF – minor allele frequency

MBS – metabolic syndrome

MC4R – melanocortin four receptor

NE – norepinephrine

OXM – Oxyntomodulin

PCSK1 – pro-hormone convertase 1

POMC – proopiomelanocortin

PVN – paraventricular nucleus

PYY – Peptide YY

QTL(s) – quantitative trait locus (loci)

RYGB – Roux-en-Y Gastric Bypass

SAPU – Sensitive data analysis platform

SIM1 single-minded homolog 1

SNP(s) – single nucleotide polymorphism(s)

T2DM – type 2 diabetes mellitus

WHO – World Health Organization

WHR – waist-to-hip ratio

WL – weight loss

INTRODUCTION

Obesity is a rising global problem. According to the World Health Organisation (WHO), in 2022, 2.5 billion adults aged 18 and older were overweight (body mass index (BMI) >= 25 kg/m^2), and 16% were living with obesity (BMI >= 30 kg/m^2) (Okunogbe et al., 2022). This corresponds to a total of 43% of adults being overweight or living with obesity worldwide. The prevalence of overweight and obesity is increasing rapidly, making it a more significant concern, as overweight and obesity contribute to various comorbid diseases, such as diabetes, cardiovascular diseases, cancer, and Alzheimer's disease (Okunogbe et al., 2022), ("Worldwide Trends in Underweight and Obesity from 1990 to 2022," 2024).

Obesity is a complex disease with genetic and environmental factors affecting its development and progression (Albuquerque et al., 2017). Excessive body fat accumulation (higher BMI) results from a food intake and energy expenditure imbalance (Kloock et al., 2023). According to twin studies, BMI is estimated to be 40-90% heritable (Elks et al., 2012). The common genes significantly associated with BMI variability are FTO, MC4R, LEP, etc (Sandholt et al., 2010). Several studies and meta-analyses reported that the effect of genetic predisposition to obesity can be influenced by the environment (Flores-Dorantes et al., 2020; Tyrrell et al., 2017a). Two main sources of unexplained variance of BMI are gene-gene (GxG) and gene-environment (GxE) interactions (Zhang & Bell, 2024). They represent a case where individuals with a particular genotype can exhibit a variable phenotype as a result of genotype interactions with specific environmental exposures or other genetic factors (Zhang & Bell, 2024).

It is essential to study and understand the gene-environment interactions (GxE) in human obesity to detect specific environmental factors that can increase or decrease the susceptibility to obesity. Moreover, with knowledge about environmental factors, it is possible to establish preventative measures and find optimal surroundings to prevent the development of obesity (*Gene–Environment Interactions in Human Diseases | Nature Reviews Genetics*).

This study focused on testing and applying different methods to analyse the genotypic and phenotypic data from the Estonian biobank (EstBB) to detect significant variance differences in BMI caused by genetic and environmental traits/factors separately, using previously reported data, including the data from the UK Biobank. Furthermore, in this study, quantitative trait loci (QTLs) and environmental variables, such as physical activity, smoking, and education level, that caused significant variance heterogeneity of BMI, were combined

with genetic data for several SNPs that were previously reported to significantly influence BMI, and tested for gene-environment interactions.

1 LITERATURE REVIEW

1.1 Biological background of obesity

Obesity is a complex disease with genetic and environmental factors influencing its progression (Albuquerque et al., 2017). The common metric that is used to estimate the degree of severity of obesity is BMI. Based on the BMI value, people can be classified as underweight, of a healthy weight, overweight, people with obesity, and people with severe obesity (**Figure 1**) (Okunogbe et al., 2022; “Worldwide Trends in Underweight and Obesity from 1990 to 2022,” 2024).

Figure 1. Weight categories based on BMI. Five weight categories can be distinguished based on the BMI value: underweight, healthy weight, overweight, obesity, and severe obesity.

Excessive body fat accumulation leads to overweight and obesity and arises from the imbalance of food intake and energy expenditure regulation (Kloock et al., 2023). The hypothalamus and brainstem are the main centers of energy homeostasis regulation (Roh & Kim, 2016). They serve as integrators of metabolic signals that indicate the energy status of the body (**Figure 2**) (Roh & Choi, 2023). The inputs derived from these signals are integrated to produce a comprehensive physiological response that controls appetite and energy expenditure to ensure energy homeostasis (Roh & Choi, 2023; Roh & Kim, 2016). Next to the median eminence of the hypothalamus lies the arcuate nucleus (ARC) (Roh & Choi, 2023). The median eminence is characterized as a circumventricular organ that receives direct neural inputs for circulating hormones and nutrients (Fitch et al., 2024). The hormones and other signaling molecules come from different organs and tissues, such as the intestines (GLP-1,

GIP, PYY, CCK, OXM), stomach (Ghrelin - stimulates the appetite), pancreas (insulin, glucagon, amylin), liver (FGF21), and adipose tissue (Leptin) (Roh & Choi, 2023). They can affect the activity of ARC neurons directly by passing through the median eminence (Roh & Choi, 2023). Leptin and insulin are adiposity signals that circulate proportionately to the amount of stored fat and inform the brain about long-term energy storage (Fitch et al., 2024; Roh & Kim, 2016). Leptin receptors are highly expressed in several brain regions, including the hypothalamus, and are activated by leptin (Roh & Choi, 2023). Two populations of ARC neurons contribute to appetite regulation - anorexigenic proopiomelanocortin (POMC) neurons and orexigenic agouti-related peptide (AgRP) neurons that co-express neuropeptide Y (NPY), a very potent appetite signal (Fitch et al., 2024; Millington, 2007; Roh & Choi, 2023). Both POMC and AgRP NPY populations provide outputs to paraventricular nucleus (PVN) neurons expressing melanocortin four receptor (MC4R) (Fitch et al., 2024; Roh & Choi, 2023). POMC neurons inhibit appetite and stimulate energy expenditure via α-melanocyte-stimulating hormone (α-MSH), activating the MC4R signalling pathway (Millington, 2007; Roh & Choi, 2023). The anorexigenic effect of α-MSH is further promoted by serotonin at 5HT-2C receptors on POMC neurons (Berglund et al., 2013). On the contrary, AgRP/NPY neurons activate appetite by inhibiting POMC neurons and blocking α-MSH action at MC4R in a manner that inhibits melanocortin signalling (Fitch et al., 2024; Roh & Choi, 2023). Imbalance of this homeostatic system can result in the up- or down-regulation of the appetite, leading to overweight/obesity or underweight (Berglund et al., 2013; Fitch et al., 2024; Roh & Kim, 2016). Both genetic and environmental factors can impact the system imbalance, leading to the disruption of energy homeostasis (Bjørnland et al., 2017; Frayling et al., 2007; Tyrrell et al., 2017b).

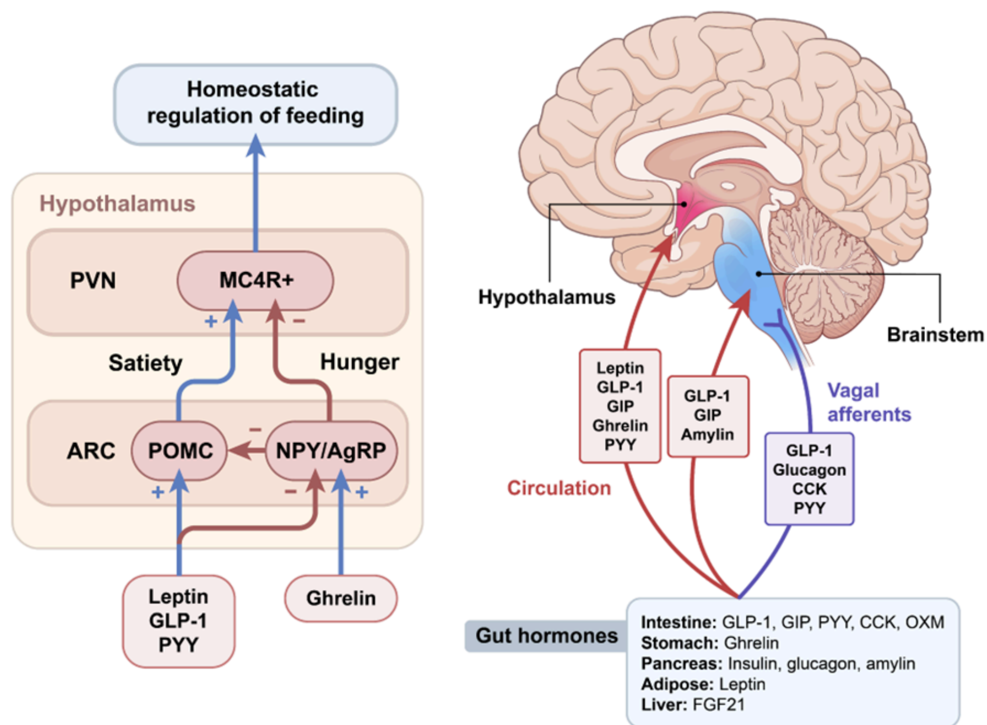

Figure 2. Regulation of food intake and energy expenditure. The main points of regulation are illustrated in the hypothalamus, brainstem, vagal afferents, and intestine. Key signaling molecules: MC4R, Leptin, GLP-1, PYY, and Ghrelin that participate in the regulatory network are shown (Roh & Choi, 2023).

1.2 Obesity treatment

Commonly used techniques to treat obesity are lifestyle changes (Tchang et al., 2000). They can be combined with specific pharmacological treatment to increase the effect on the weight loss outcome (Chakhtoura et al., 2023). However, in severe cases, bariatric surgery is performed to maintain the long-term outcome of the weight loss (Hopkins et al., 2016). Moreover, obesity comes with associated comorbidities. They include cardiovascular disease (CVD), type 2 diabetes mellitus (T2DM), certain types of cancer (especially esophagus, colon, rectum, and liver), chronic kidney disease (Afshin et al., 2017), hypertension, and dyslipidemia (Lee et al., 2016). Being overweight is generally associated with a shorter life expectancy, and it became evident through the COVID-19 pandemic (Bornstein et al., 2021). Such comorbid conditions significantly affect obesity treatment, including the type of surgery that needs to be performed on the patient or lifestyle modifications, and treatment outcomes (Kloock et al., 2023).

Conservative treatment includes changes in eating habits and the incorporation of sports into the daily life of a patient (**Figure 3**) (Kloock et al., 2023). If such treatment is not successful,

anti-obesity medications and/or bariatric surgery are the next options for obesity treatment (Kloock et al., 2023).

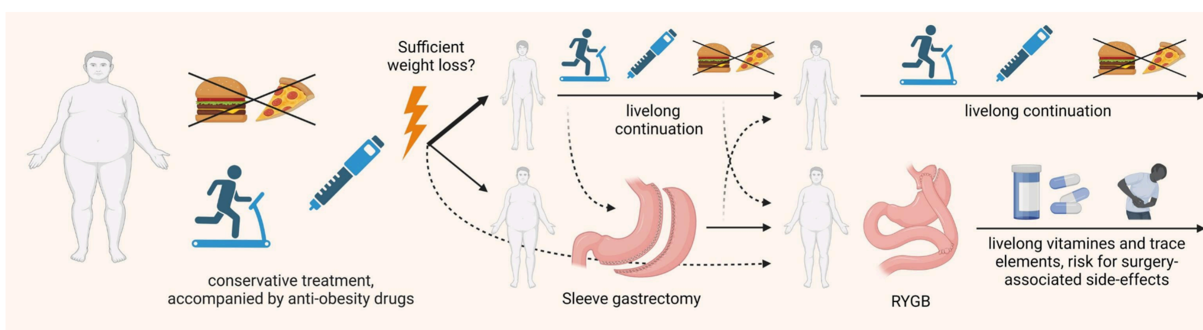

Figure 3. Treatment strategy for patients with severe obesity (BMI > 40 kg/m^2) (Kloock et al., 2023). If there are no metabolic comorbidities, then the first step is a conservative treatment that includes reduced calorie intake, lifestyle changes, such as incorporating sports on a regular basis, and anti-obesity medications (for example, GLP-1 analogues, metformin). In patients with obesity and metabolic disorders, metabolic surgery should be considered as an early-stage treatment option. A lifelong continuation of such a lifestyle is needed if sufficient weight loss is reached without invasive treatment. If conservative treatment is not successful, bariatric surgery should be the next option. Following surgery, vitamins and trace elements should be taken and monitored lifelong to prevent any risks for surgery-associated side effects.

1.2.1 Lifestyle changes

In non–severe cases of obesity, lifestyle changes should be considered as a primary treatment or preventative option. They include the reduction of caloric intake (diets) and an increase in physical exercise (energy expenditure). Calorie restriction is important for achieving clinically impactful weight loss (Wadden et al., 2020). According to Obesity Guidelines, consumption of a diet that is designed to obtain a deficit of 500-750 kcal/day, with a resulting mean weight loss of 0.5 - 0.75 kg per week, is recommended (Jensen et al., 2014; Wadden et al., 2020). Correspondingly, women are frequently prescribed 1200-1500 kcal/day and men 1500-1800 kcal/day (Jensen et al., 2014; Wadden et al., 2020). Higher target values of weight loss are applied for individuals who weigh more or with higher physical activity levels (Jensen et al., 2014; Wadden et al., 2020). Moreover, the diet composition does not influence the weight loss outcome, as long as a sufficient calorie deficit is present (Sacks et al., 2009).

The separate effect of physical exercise on weight loss was studied in the randomized controlled trial (Slentz et al., 2004). According to the study, which lasted more than three years, it was found that the effect of exercise on weight loss follows a dose-response manner in the absence of any dietary changes. The achieved mean weight loss was not greater than

2.9 kg in the high-intensity physical exercise group (Slentz et al., 2004). Moreover, advantageous effects (0.9 kg of weight loss) for moderate intensity physical exercises (30 min of walking per day) group were reported (Slentz et al., 2004).

The combined effect of both diet and physical exercise has been shown to be generally more effective on weight loss compared to diet only. Meta-analyses suggest that the pooled mean weight loss difference between diet and physical exercise group was greater than in the physical exercise only group (Foster-Schubert et al., 2012; Johns et al., 2014; Messier et al., 2013).

Moreover, in the Nutrition and Exercise in Women study, authors studied the effect of 12 months of a low-calorie diet (1200–2000 kcal/day, <30% fat), moderate intensity exercises (5 days/week, 225 min/week) or the combination of both programs on weight loss and body composition in overweight and obese postmenopausal women (Foster-Schubert et al., 2012).

It was found that the total weight loss was greater in the diet and exercise group (10.8%), compared to the diet alone (8.5%) and exercise alone (0.8%) groups (**Figure 4**) (Foster-Schubert et al., 2012). Similarly, in a different meta-analysis, it was reported that the combination of diet and physical exercise resulted in a greater weight loss (5.3 kg greater) compared to the diet alone in a course of 3-6 months of interventions (Johns et al., 2014). In the case of 12-18 months long interventions, a 6.3 kg greater weight loss was reported in the case of diet and physical exercises compared to the diet-only programs (Johns et al., 2014).

All together, combined diet and physical exercise result in a greater weight loss compared to physical exercise only (Foster-Schubert et al., 2012; Swift et al., 2018). Moreover, the diet-only approach can result in a clinically significant weight loss; the overall importance of physical activity for health, especially for cardiovascular diseases, should not be neglected (Ross et al., 2000).

Other lifestyle modification approaches involve lifestyle coaching and behavioral therapy to change eating habits, preferences, and other behavioral patterns (Wadden et al., 2020). Overall, lifestyle interventions can have a moderate effect on weight loss, and often are combined with pharmacological and/or invasive treatment (Wadden et al., 2005).

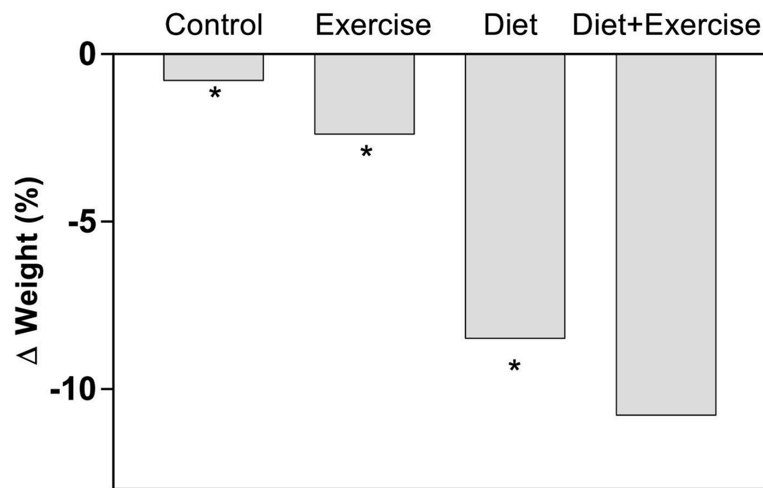

Figure 4. Changes in weight in response to 12 months of diet only, exercise only, and the combination of diet and exercise in postmenopausal overweight-to-obese women in the Nutrition and Exercise in Women study. *Shows significant difference compared to the diet and exercise group (Foster-Schubert et al., 2012; Swift et al., 2018).

1.2.2 Pharmacological treatment of obesity

In addition to lifestyle changes, there are different FDA-approved anti-obesity medications for non-syndromic obesity **(Table 1)**. These medications help to maintain the weight loss over an extended period of time, therefore, their intake can be lifelong in many cases of severe obesity (Chakhtoura et al., 2023; Tchang et al., 2000). They target peripheral (adipose tissue, gastrointestinal tract) and central (brain) signaling pathways of the hunger-satiety regulatory network to decrease food intake by reducing the appetite and increasing satiety (Chakhtoura et al., 2023; Tchang et al., 2000).

Table 1. Anti-obesity medications: Name (and a trade name), year of approval, mechanism of action. Adapted from (Chakhtoura et al., 2023).

Drug (trade name)	Approval FDA/EMA (year)	Mechanism of action
Orlistat (Xenical, Alli)	FDA 1999	Gastric and pancreatic lipase inhibitor
	EMA 1998	
Phentermine/Topiramate (Qsymia)	FDA 2012	NE agonist/GABA agonist, glutamate antagonist
Naltrexone/Bupropion (Contrave/Mysimba)	FDA 2014	Opioid receptor antagonist/DA and NE reuptake inhibitor
	EMA 2015	
Liraglutide (Saxenda)	FDA 2014	GLP-1 analogue
	EMA 2015	
Semaglutide (Wegovy)	FDA 2021	GLP-1 analogue
	EMA 2021	
Setmelanotide (Imcivree)	FDA 2020	MC4R agonist
	EMA 2021	
Tirzepatidec	Under consideration by FDA	GIP/GLP-1 dual agonist

Each of the reported medications has its adverse events and contraindications, which are discussed in detail by Chakhtoura et al., 2023.

In the case of monogenic obesity, pharmacological treatment is directed to eliminate the effects of single-gene deficiencies and restore the normal regulation of hunger-satiety (Chakhtoura et al., 2023; Tchang et al., 2000).

The effect of anti-obesity medications depends on several factors, such as the severity of obesity, associated comorbidities, age of the patient, and previous medical history (Tchang et al., 2000). Thus, the combined approach of surgery, pharmacological treatment, and lifestyle changes needs to be performed to treat obesity and avoid potential complications successfully.

1.3 Genetics of obesity

According to twin and family studies, BMI is highly heritable: the heritability estimates range from 0.47 to 0.9 (twin studies), and 0.24 - 0.81 (family studies) (Elks et al., 2012).

Based on the genetics data, obesity can be classified as syndromic and non-syndromic (Fitch et al., 2024). Syndromic obesity is associated with intellectual disability, developmental delay, dysmorphological features, or abnormalities affecting different organs and systems, with low frequency, high variability, and a Mendelian pattern of inheritance (Bochukova et al., 2010; Fitch et al., 2024; Šket et al., 2022). For example, Prader–Willi, fragile X, Bardet–Biedl, Cohen, and Albright Hereditary Osteodystrophy syndromes are associated with early-onset obesity and developmental delay (Bochukova et al., 2010; Mahmoud et al., 2022).

Non-syndromic obesity can be divided into monogenic and polygenic. The most widely spread form of non-syndromic obesity is polygenic obesity (**Table 2**) (Fitch et al., 2024). It results from the cumulative effect of each gene that affects BMI and interactions of the genes with various environmental and behavioral factors, such as physical activity, smoking, eating habits, stress, etc (Fitch et al., 2024). More than 500 loci with varying effect sizes associated with BMI were identified in different GWAS studies (**Figure 5**) (Khera et al., 2019; Pulit et al., 2019; Speliotes et al., 2010).

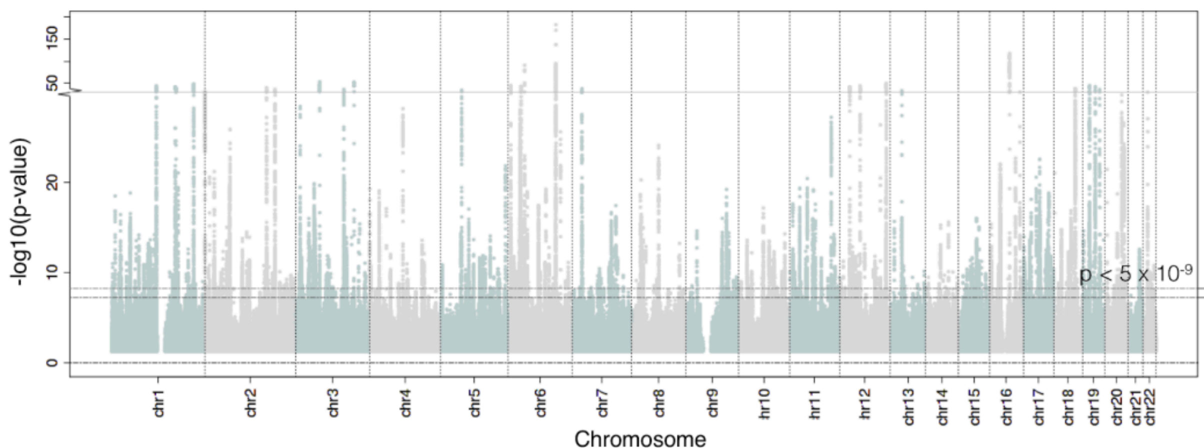

Figure 5. Manhattan plot for BMI (max N = 806,834) from a large meta-analysis of fat distribution and obesity phenotypes (Pulit et al., 2017). Meta-analysis of UK Biobank GWAS with existing GWAS data generated by the GIANT consortium was performed. Y-axes are disrupted at p < 1 x 10-50. Genome-wide significance was set at p < 5 x 10-9 to illustrate the SNP density of the UK Biobank data. Traditional genome-wide significance (p < 5 x 10-8) is indicated by the lower horizontal line (Pulit et al., 2019).

Monogenic obesity is linked to single-gene deficiencies primarily within the hypothalamic MC4R pathway (**Table 2**), such as Fat mass and obesity-associated gene (FTO), melanocortin four receptor (MC4R), leptin (LEP), leptin receptor (LEPR), proopiomelanocortin (POMC), pro-hormone convertase 1 (PCSK1), single-minded homolog 1 (SIM1) (Fitch et al., 2024; Mahmoud et al., 2022; Pulit et al., 2019). Identification of such variants allows prediction of the individual's risk for developing obesity using polygenic risk scores (Khera et al., 2019).

Table 2. Comparison of polygenic and monogenic obesity. Adapted from (Fitch et al., 2024).

<table>
<tr><th></th><th>Polygenic</th><th>Monogenic and syndromic obesity</th></tr>
<tr><td>Etiology</td><td>Cumulative effects of multiple genetic variants interact with environmental factors that may facilitate development of obesity</td><td>Variants in ≥1 gene and/or chromosomal deletion with a strong effect</td></tr>
<tr><td>Primary cause(s) of obesity</td><td>Multifactorial: Diet, Physical activity, Aging, Pregnancy, Abstinence from substance use, Medications, Stress, Circadian disruption, Microbiome</td><td>Genetic variants</td></tr>
<tr><td>Age of obesity onset</td><td>Any age</td><td>Early in life (before 5 years of age)</td></tr>
<tr><td>Feeding behavior</td><td>Phenotypic variability (eg., poor satiety, increased hunger)</td><td>Negative impact on patient and caregiver quality of life</td></tr>
<tr><td>Efficacy of lifestyle-based management strategies</td><td>Partially effective; recommended to be used with pharmacotherapy</td><td>Not effective; strategies do not target underlying hyperphagia</td></tr>
<tr><td rowspan="2">Response to medical or surgical management strategies</td><td rowspan="2">Heterogeneous responses</td><td>Limited research and long-term efficacy for traditional pharmacologic agents, GLP-1 R agonists, and MBS; strategies do not target hyperphagia</td></tr>
<tr><td>Efficacious therapy with MC4R agonism in patients with POMC deficiency, LEPR deficiency, or BBS</td></tr>
</table>

1.3.1 Common genes related to BMI variation and obesity

FTO was the first obesity-related gene that was discovered through a GWAS performed on European patients who had type 2 diabetes (Frayling et al., 2007). Multiple SNPs in the first intron of the FTO gene have been shown to be significantly associated with BMI variation in children and adults, specifically, the rs9939609 and rs9930506 variants (Frayling et al., 2007). It was reported that FTO plays an important role in food intake regulation through the MCR4 pathway (Ahmad et al., 2011; Sonestedt et al., 2009). People with FTO risk alleles tend to

have a higher intake of fat- and protein-rich food, reduced satiety, which results in overeating and increased BMI (Ahmad et al., 2011; Sonestedt et al., 2009).

Another gene that is significantly associated with higher BMI and obesity is MC4R (Bjørnland et al., 2017). It encodes the MC4R protein, which plays a crucial role in food intake and energy expenditure regulation (Fatima et al., 2022). In the case of risk variants of MC4R, people tend to accumulate more body fat, as the regulation of appetite is disrupted: MC4R is not associated with α-MSH, therefore, the satiety effect is not achieved, and food intake is increased (Fatima et al., 2022). Common risk variants for MCR4 are rs6567160, rs17782313 (Lazopoulou et al., 2015; Pulit et al., 2019).

Leptin is a protein that is secreted by the white adipose tissue, passes the blood-brain barrier, and binds to the presynaptic GABAergic neurons of the hypothalamus, causing a decrease in appetite and an increase in energy expenditure (Obradovic et al., 2021). The common risk variants of Leptin that are associated with increased BMI are rs7799039 (LEP) and rs1137101 (LEPR) (Mahmoudi et al., 2016).

POMC is another example of an obesity-related gene (Millington, 2007). POMC is an appetite-inhibitory protein that impacts the leptin-melanocortin pathway (Dubern et al., 2008; Millington, 2007). Decreased levels of POMC lead to the absence of alpha-MSH and adrenocorticotropic hormone (ACTH), as these molecules are cleaved from POMC, resulting in hyperphagia, lower resting metabolic rate, and severe obesity (Millington, 2007).

1.4 Gene-environment interactions

Polygenic risk scores explain only ~6% of BMI variation in the population, taking into account only the genetic effect on obesity (Yengo et al., 2018). Possible sources of unexplained BMI variance are rare gene variants with strong effect sizes and/or gene-gene (GxE) and gene-environment (GxE) interactions (Yengo et al., 2018; Zhang & Bell, 2024). They cause BMI variance heterogeneity across genotype groups (Flores-Dorantes et al., 2020; Zhang & Bell, 2024).

Gene-environment interactions are defined as the case where the dependency of the trait on the environment is different for different genotypes (and vice versa) (**Figure 6**) (Anholt & Mackay, 2018; Zhang & Bell, 2024).

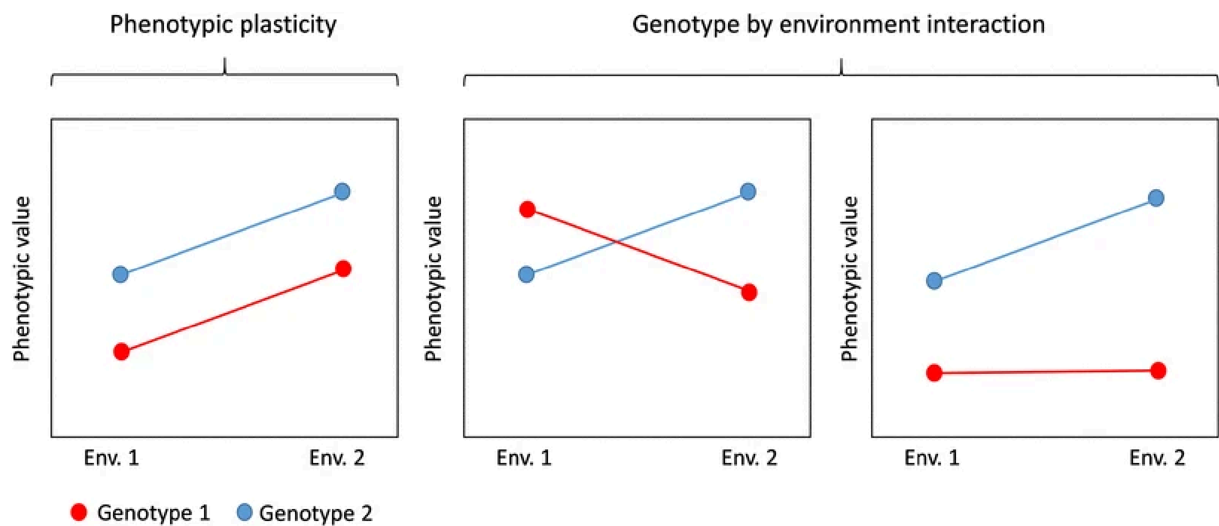

Figure 6. Phenotypic plasticity and GxE interaction (Anholt & Mackay, 2018). The phenotypic norm reaction is shown for two genotypes. In the case of phenotypic plasticity, the response of each genotype to the environment is similar (lines are parallel). In contrast, when the GxE is present, two genotypes can respond differently (lines in the opposite direction), or one genotype can respond to a different environment, while the other will be stable.

The example of the GxE interactions is the FTO interaction with physical activity, smoking, and artificially sweetened beverages (**Figure 7**) (Bjørnland et al., 2017). It was demonstrated that the effect of FTO on BMI and WHR in the age group 20-40 years is negated by the increasing effect of physical activity (**Figure 7A, B**) (Bjørnland et al., 2017). Precisely, the individuals with higher physical activity index have similar WHR and BMI, regardless of the FTO genotype (Bjørnland et al., 2017). The results are similar between males and females (**Figure 7A, B**). In contrast, in the case of interaction with the intake of artificially sweetened beverages, BMI is similar across FTO genotype groups when the consumption is zero for both males and females (age 40-60) (**Figure 7C**) (Bjørnland et al., 2017). However, when the consumption of beverages increases, BMI also increases, although in a different pattern for the three FTO genotypes in males (**Figure 7C**) (Bjørnland et al., 2017). In females, there was no significant interaction between FTO and artificially sweetened beverages (**Figure 7C**) (Bjørnland et al., 2017).

In the case of FTO interaction with smoking (pack years), for males who had one or two FTO risk alleles, the increase in pack years was significantly associated with increased BMI, although for males with no FTO risk alleles, there was no interaction detected (**Figure 7D**) (Bjørnland et al., 2017). For females, however, the effect was the opposite: females with two FTO risk alleles showed decreased BMI with increased pack years (**Figure 7D**) (Bjørnland et

al., 2017). For females with no or one FTO risk allele, an increase in pack years was associated with increased BMI (**Figure 7D**) (Bjørnland et al., 2017).

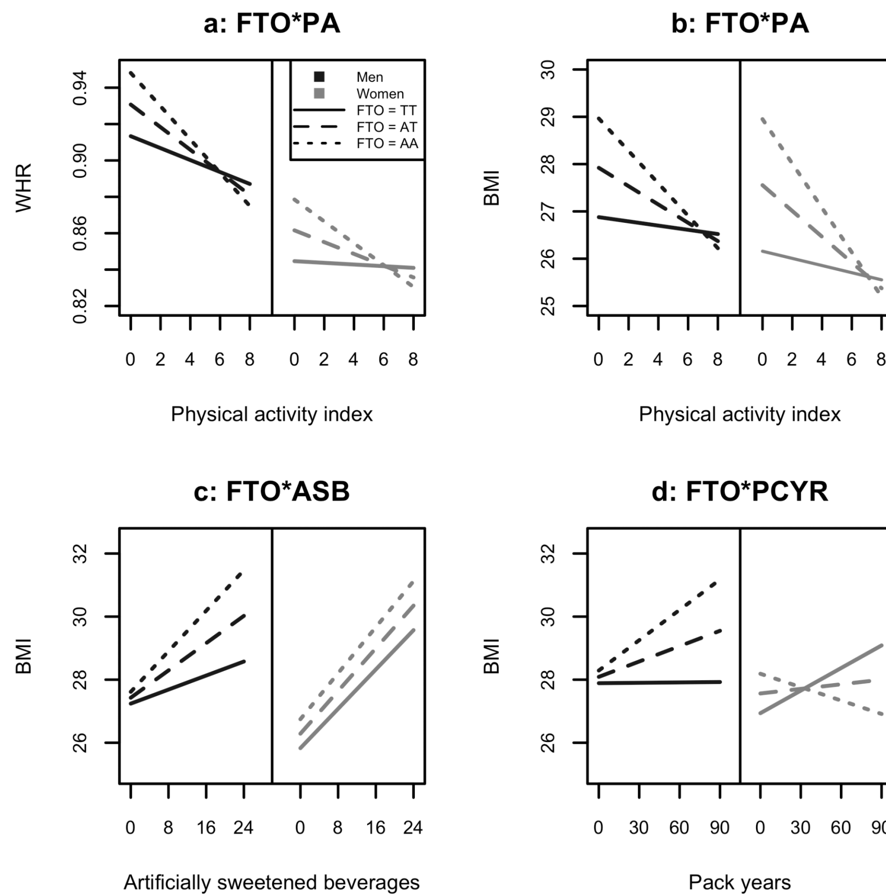

Figure 7. Gene-environment interactions for the FTO genotypes (Bjørnland et al., 2017). Estimated WHR (a) and BMI (b) for all FTO genotypes in the 20-40 years age group and for increasing levels of physical activity (PA) for females and males separately. Estimated BMI for all FTO genotypes for males and females separately in the 40-60 age group for increasing intake of artificially sweetened beverages (c) and increasing pack years (d).

Another large study in the UK Biobank showed the evidence for gene-environment interactions for BMI genetic risk scores consisting of 69 known variants that influence BMI with different environmental factors, such as fizzy drink consumption, sedentary time, percentage fat consumption, etc, in the UK population (**Figure 8**) (Tyrrell et al., 2017b).

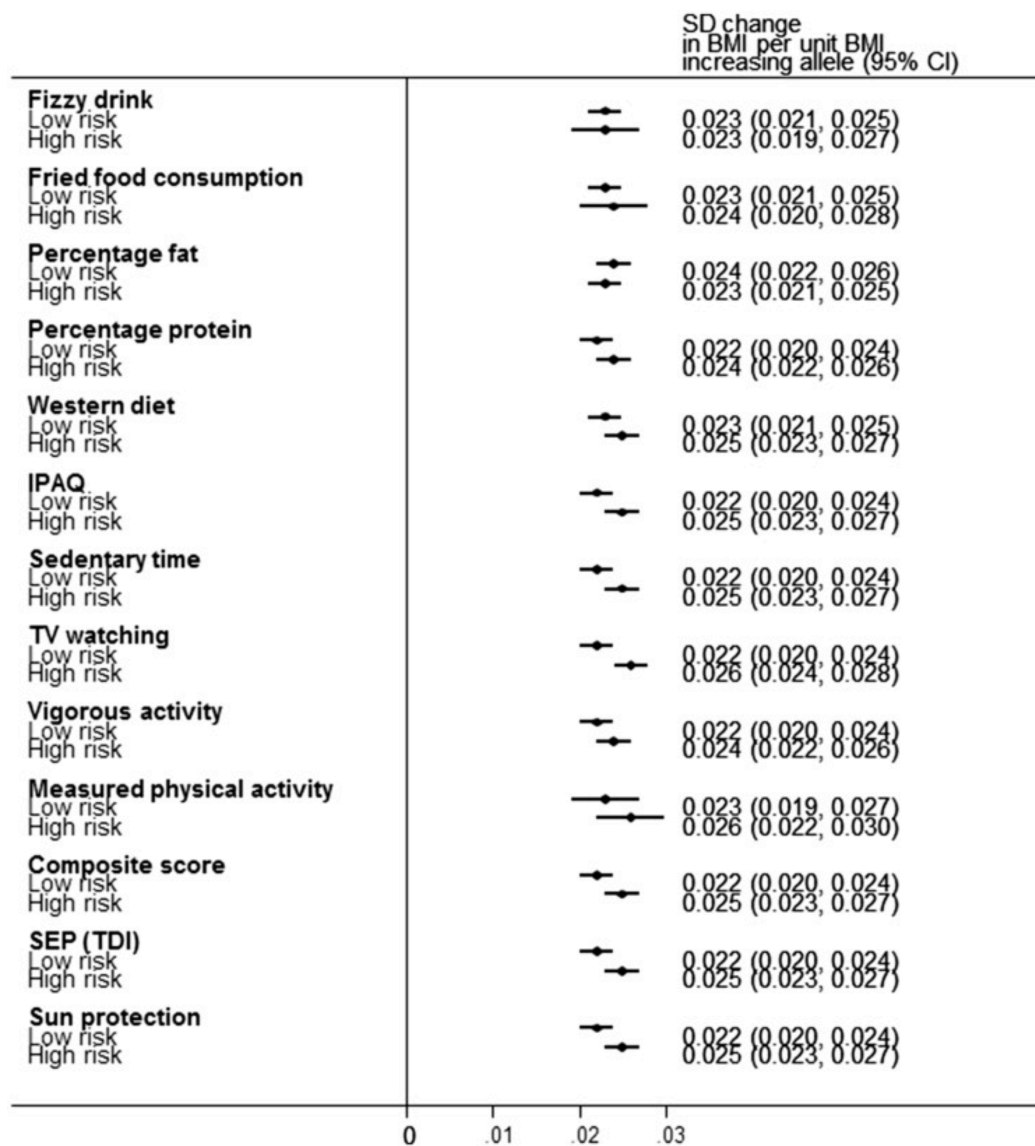

Figure 8. Forest plot showing the change in BMI per-allele increase in BMI genetic risk score for the 12 different obesogenic environments and a negative control on standardized inverse normalized scale (Tyrrell et al., 2017). BMI was corrected for sex, age, ancestry principal components, and assessment centre location prior to calculating residuals.

Assessing gene-environment interactions is relevant for designing the individualized treatment and prophylaxis against obesity, as it allows for the detection of specific environmental variables that might have a significant effect on an individual's BMI, and the development and progression of obesity (Bjørnland et al., 2017; Tyrrell et al., 2017b).

1.5 Methods to detect variance differences across different study groups

Detection of gene-environment interactions is essential for establishing specific environmental factors that can influence the progression and development of obesity and other complex traits in humans (Bjørnland et al., 2017; Tyrrell et al., 2017b). Genetic variants that are associated with phenotypic trait variability are called variable quantitative trait loci (vQTLs) (**Figure 9A**) (Zhang & Bell, 2024). Genetic variants that participate in GxE are the types of vQTLs (Bjørnland et al., 2017; Zhang & Bell, 2024). Detecting vQTLs instead of directly assessing the effects of GxE is more expedient since it doesn't require previous knowledge about the genetic variants' interaction with environmental factors, or doesn't require a comprehensive multidimensional search of interactions among multiple factors (Zhang & Bell, 2024).

The general goal of the methods for capturing GxE is to detect phenotypic variance changes across different genotype groups (Zhang & Bell, 2024). Such methods are categorized into variance-only and mean-variance joint tests, based on their ability to detect both mean-controlling QTLs and vQTLS (**Figure 9A, B**) (Zhang & Bell, 2024).

Variance-only methods contain homogeneity of variance tests and association tests (Figure 11B). Most mean-variance joint methods include the use of likelihood ratio tests to study the role of genetic variants (**Figure 9B**). In the study, the methods for detecting vQTLs were compared using the previously reported data to test their ability of replicating the results using genotype and gene expression data from the lymphoblastoid cell line samples from 765 samples from the TwinsUK cohort (Brown et al., 2014; Zhang & Bell, 2024).

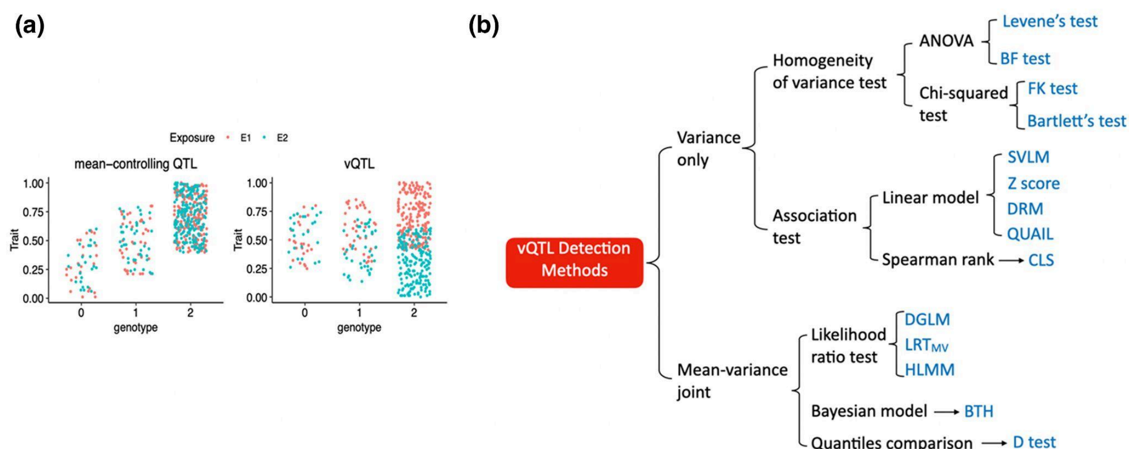

Figure 9. Overview of methods for vQTLs detection (Zhang & Bell, 2024). a) Example of traits associated with a mean-controlling QTL (left) and a vQTL (right). b) Overview of vQTL identification methods.

The performance of each method was evaluated using the false positive rate and the discovery rate based on the simulation results(Zhang & Bell, 2024). The datasets used for the method evaluation had the variable sample sizes (500, 1,000, 5,000, 10,000, 20,000, 50,000, 100,000, and 200,000), minor allele frequencies (MAFs) (0.05, 0.1, 0.2, and 0.3), environmental factor distributions (binary, normal, and uniform distributions), and phenotype distributions (normal, Chi-squared, and gamma distributions) (Zhang & Bell, 2024).

The FPRs of Bartlett's test, CLS, DGLM, FK, Levene's test, and QUAIL were near 0.05 when the error followed a Normal distribution, but were greater when the error was non-normally distributed (Zhang & Bell, 2024). The SVLM, BF, and DRM methods showed a consistently low FPR, regardless of the error distribution (Zhang & Bell, 2024)

In the case of GxE with a discrete environmental variable, SVLM is optimal with normally distributed data (Zhang & Bell, 2024). The optimal model is shifted from SVLM to DRM when the sample size is increased and the data are non-normally distributed (Zhang & Bell, 2024). BF is one of the optimal methods when MAF is 0.2 or greater, regardless of data distribution (Zhang & Bell, 2024).

In the case of a G×E interaction with continuous exposure, such as BMI, SVLM, DRM, and BF methods have low FPR (Zhang & Bell, 2024). At low MAFs, the SVLM shows the highest discovery rate when the phenotypic data follow the normal distribution, and DRM performs the best in the case phenotype data is non-normally distributed (Zhang & Bell, 2024). BF outperforms the other two methods when the MAF is increased (**Figure 10**) (Zhang & Bell, 2024).

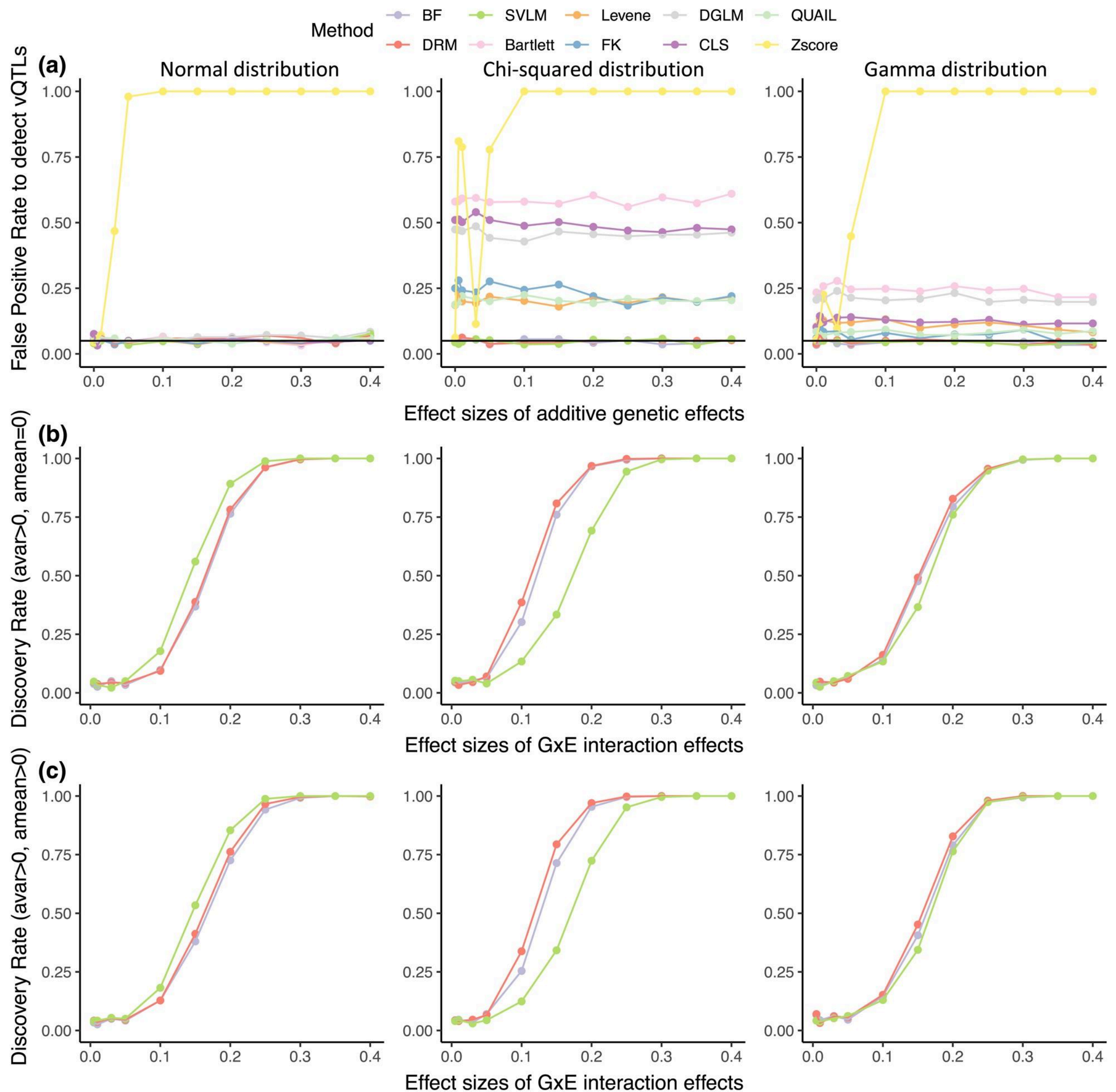

Figure 10. Method performance for uniformly distributed continuous exposure (Zhang & Bell, 2024). The discovery rate of methods under the scenario that a single genetic variant affects a) trait level only ($a_{mean} \geq 0$, $a_{var} = 0$), which includes the situation where there is no genetic effect ($a_{mean} = 0$, $a_{var} = 0$), and b) trait level and variance ($a_{mean} = 0.1$, $a_{var} > 0$). The black horizontal line corresponds to a rate of 0.05. The trait is also affected by noise and environmental factors. The first, second, and third columns represent errors generated from normal, Chi-squared, and gamma distributions, respectively (Zhang & Bell, 2024).

Thus, the BF test is the best approach for detecting variance heterogeneity in the genotype groups for both discrete and continuous variables, and is more robust against non-normally distributed data, compared to Levene's test (Wang et al., 2017; Zhang & Bell, 2024)

2 THE AIMS OF THE THESIS

Previous studies have identified the gene candidates that highly influence BMI variance. Moreover, some of the GxE interactions of these genes with obesogenic environmental factors have also been described, along with particular methods to detect and study them. However, most of the studies were performed with the UK Biobank data, limiting the study population to the UK region. This study aims to investigate the potential of different methods to detect variance heterogeneity and possible GxE in the population data from the Estonian Biobank. Furthermore, this work aims to study a new approach of using the variance heterogeneity test (Brown-Forsythe test) on environmental variables to detect the ones that are significantly associated with BMI variability, thus serving as potential candidates for GxE.

Overall, the main aim of the study is to explore the potential of several methods for detecting the phenotypic variance heterogeneity using the data from the Estonian Biobank and to develop a methodological approach for testing possible GxE interactions in human obesity.

The objectives of this work are:

- Identify the optimal method for testing for variance heterogeneity using the previously published studies
- Perform the random permutation test to validate the identified method for variance heterogeneity testing using the training dataset from Kaggle and the EstBB data
- Identify variance QTLs that are associated with BMI variability in the EstBB population data using the variance heterogeneity approach and previously published studies
- Identify environmental factors that are associated with BMI variability in the EstBB population based on the phenotypic variance heterogeneity method and previously published data
- Test the identified genes and environmental factors for gene-environment interactions

The thesis is structured as follows. First, because getting access to data from the Estonian biobank was a lengthy process, a publicly available dataset from the Kaggel repository, containing BMI measurements and lifestyle variables, was used to explore the properties of published methods for detecting heteroscedasticity and investigate these signals in a real-world dataset. As previous studies had indicated that some methods can generate excess false positives when presented with data from skewed distributions (such as BMI), the permutation-based test was performed for bias in the p-values reported by the tests.

Second, the signals of heteroscedasticity in BMI and waist-to-hip ratio in the Estonian Biobank data were investigated. They were first studied separately for BMI quantitative trait loci (QTL) SNPs and selected lifestyle variables, and gene-environment effects were tested by joint statistical modelling for pairs of SNPs and lifestyle variables that both showed evidence of heteroskedasticity. The analyses were structured by sex and age.

3 EXPERIMENTAL PART

3.1 MATERIALS AND METHODS

3.1.1 Data

3.1.1.1 Kaggle dataset

The dataset from Kaggle contained the data of 2402 individuals with or without metabolic syndrome (https://www.kaggle.com/datasets/antimoni/metabolic-syndrome). The data included BMI, age, and sex, along with triglycerides, glucose, cholesterol, and other metabolite levels in blood for each patient. Moreover, the dataset contained information about whether the individual had the metabolic syndrome or not. The data was used to analyze BMI variance difference across the study groups based on the environmental variables, such as triglycerides or glucose levels in blood. Moreover, the dataset was used to evaluate the Brown-Forsythe test for variance heterogeneity by performing the random permutation test and analysing the null distribution of the test's p-value null distribution.

3.1.1.2 Dataset from the Estonian Biobank

Estonian Biobank is a population-based data bank that contains the data for more than 200,000 individuals who were genotyped with genome-wide arrays. Data reflects the age, sex, and geographical distribution of the adult Estonian population (*Estonian Biobank*, 2024).

Data collection was performed according to the Estonian Human Genes Research Act (HGRA, https://www.riigiteataja.ee/en/eli/531102013003/consolide). All participants, whose data was collected, signed the consent form allowing the use of the data in various studies (*Estonian Biobank*, 2024).

All participants were recruited randomly by general practitioners and physicians in hospitals. Individuals have filled out the questionnaire, which included the following entries:

- personal data (place of birth, place(s) of living, nationality etc.),
- genealogical data (family history of medical conditions spanning four generations),
- educational and occupational history,
- lifestyle data (physical activity, dietary habits - FFQ, smoking, alcohol consumption, women's health, quality of life) (*Estonian Biobank*, 2024).

Ethical approval for the study was obtained from EBIN (Eesti bioeetika ja inimuuringute nõukogu, approval no. 1.1-12/327), and the data was released from the Estonian Biobank. The dataset included phenotype, genotype, and chromosomal position data (including the

chromosome number, rsIDs of the SNPs, and their position on the chromosome). The initial data was preprocessed to include the genotype data for 668 SNPs that were shown to be significantly associated with BMI variability, according to the large meta-analysis study (Pulit et al., 2019). Phenotypic data was provided for 129115 individuals based on the questionnaire entries, such as age, sex, BMI, WHR, educational, occupational data, health, and lifestyle data, such as physical activity per week, smoking, alcohol consumption, sleeping schedule, etc. The vkood column in the phenotype data referred to the unique ID of each participant.

3.1.2 Methods

3.1.2.1 Software and data processing

The R Studio software (R version 4.3.2, R Studio 2023.09.1 + 494 "Desert Sunflower" release version for Linux operating system x86_64) was used for working with the data, plotting, and performing the statistical tests.

The data was made available in the Sensitive Data Analysis Platform (SAPU) provided by the University of Tartu High-Performance Computing Centre through a web browser interface, and all analysis of the Estonian Biobank data (https://docs.hpc.ut.ee/public/services/SAPU/) was carried out in this environment.

The function *read.table()* was used to load the separate .txt data files into the R environment. Function parameters were specified for each of the data types. For phenotypic data, parameters were set to header = TRUE, sep = "\t", quote = " ", fill = TRUE. For genotype data, the parameters were as follows: header = FALSE, sep = " ". Lastly, for the position data, the function parameters were set to header = TRUE, sep = " ". These parameters ensured the correct reading and loading of the data. In the case of the Kaggle dataset, the function *read.csv* was used to read and load the data into the R environment.

Furthermore, the function *na.omit()* was used to get rid of the possible rows with the NA data entries. Basic subsetting commands in R were used to separate the dataset into age groups and to get the data for male and female individuals separately. The function *cbind()* was used to combine the specific SNP columns with the phenotype and position data.

The WHR was calculated for each individual by dividing the waist measurement by the hip measurement.

For plotting the data, the *ggplot2* package was used along with the base R functions *plot()* and *hist()* (https://cran.r-project.org/web/packages/ggplot2/index.html).

Lastly, the package *car* was used to perform the Brown-Forsythe test by specifying the center = "median" parameter in the *leveneTest* function, to turn it into the Brown-Forsythe test (https://cran.r-project.org/web/packages/car/index.html). For the Kolmogorov-Smirnov test, the *ks.test()* function was used with the generated null distribution of the BF test p-values. For comparison, the randomly generated uniform distribution was used (*runif()* function).

In the appendix, the sample code for performing the BF test on phenotypic groups, along with validation of the BF test and plotting the GxE interaction plot, is provided (**Appendix, Figure 1, Figure 2**).

3.1.2.2 Variance analysis with the Brown-Forsythe test

The new approach for selecting the environmental variables that are associated with high BMI and WHR variability was used in this study. The candidate environmental factors were selected based on the previously published results. In particular, the environmental variables from EstBB included physical activity (hours per week), smoking (categorical variable with "Current", "Never", "Former" groups), working_in_shifts_last3months ("true", "false", and alcohol_first_age_consumption (numeric variable). In the case of the Kaggle dataset, the variables were triglyceride and glucose levels in blood.

The study groups were created for BMI and WHR based on the environmental exposure, for example, high and low physical activity groups based on the data spread and/or the quantile selection (lowest and highest quantile groups), "Current", "Never", "Former" smoker groups, etc. Physical activity data from the EstBB were more concentrated in the range from 0 to 5 hours per week; therefore, the two groups, 0 and <= 5 hours per week of physical activity, were created for the analysis.

To correct for the effect of age and gender on the analysis, four age groups were created for this study: the first age group from 18 to 30 years, the second from 31 to 50, the third from 51 to 65, and the fourth group over 65 years. Additionally, the analysis was performed for males and females separately.

The Brown-Forsythe test was performed to detect the significant variance difference in BMI or WHR across different study groups. Compared to Levene's test for variance homogeneity, the BF test is more robust in the case of skewed data, as it uses the median statistic instead of the mean (Wang et al., 2017; Zhang & Bell, 2024). The BF test was performed according to the following formula:

$$\text{BF test}: F = \frac{N-G}{G-1} \times \frac{\sum_{g=1}^{G} N_g \times (Z_g - Z_{..})^2}{\sum_{g=1}^{G}\sum_{i=1}^{N_g} (Z_{gi} - Z_g)^2}$$

N - number of samples from all study groups together,

G - number of study groups,

Ng shows the number of samples in gth group,

Zgi - absolute value of the deviation of the data points from the 10%-trimmed median value of the particular group

Zg - the mean of Zgi within a group

Z.. - the weighted mean of Zg

The BF test results were summarized in a separate table for different age groups, and males and females separately. If the difference was considered statistically significant, the phenotypic trait was further used in the GxE analysis.

3.1.2.3 Brown-Forsythe test validation

To determine if the BF test worked correctly and the output of the test was valid, a random permutation test was performed. The steps for performing the random permutation test are shown in **Figure 11.**

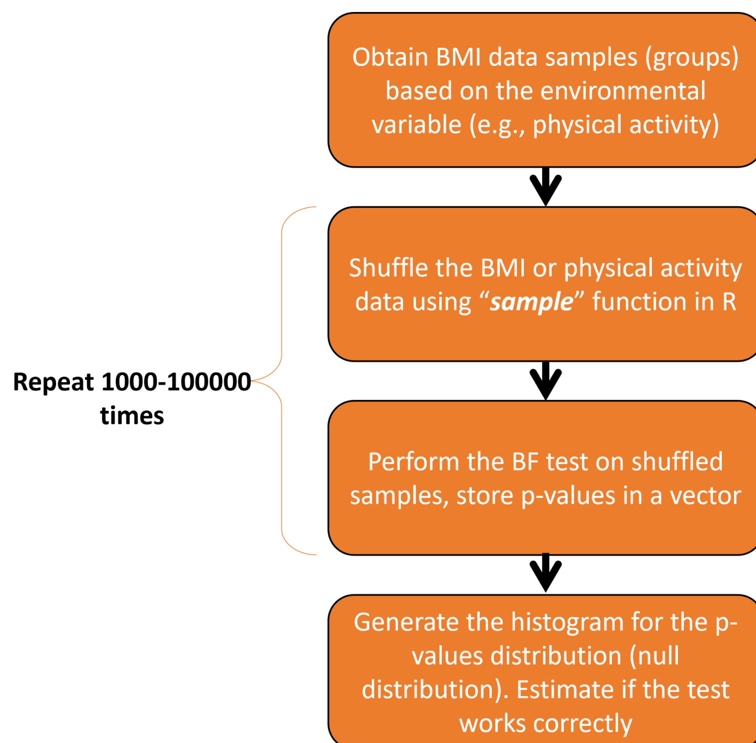

Figure 11. The overview of the steps for performing the random permutation test in R. Firstly, the groups based on the environmental variable (such as physical activity) are generated for the outcome variable (BMI). Next, the data is shuffled using the sample function in R, the BF test is performed, and the results are stored in a vector. The process is repeated 1000-100000 times. Lastly, the generated distribution of the p-values is compared to the random uniform distribution using the Kolmogorov-Smirnov test.

If the difference between the generated p-value distribution and a random uniform distribution is not statistically significant, the BF test works correctly, and its p-value under the random permutation test follows the uniform distribution.

3.1.2.4 Selection of phenotypes and genotypes based on variance heterogeneity

The BF test was carried out between different phenotype and genotype groups (based on the SNP data from the Estonian Biobank) for the outcome variable - BMI. The environmental factors and SNPs that were significantly associated with BMI variability were selected for further analysis to detect possible GxE.

3.1.2.5 Linear model to capture GxE

To visualize possible gene-environment interactions, the environmental trait (e.g., physical activity, smoking) was plotted on the x-axis, and the BMI (or WHR) on the y-axis. Such plots were generated for all possible SNP variants of the FTO and MCR4 genes as the main candidates for GxE interactions.

Furthermore, the linear model in the base R (*lm* function) was used to estimate the trends of BMI (WHR) changes in response to the environmental exposure based on the genotype groups (FTO, MC4R). The interaction terms were added to model the GxE, for example, *lm(BMI~physical activity *SNP)*, which estimates the interaction of physical activity and BMI for each of the SNPs.

The slopes of the linear regression lines were extracted from the models, and p-values were analysed. This approach allows for the detection of the presence or absence of GxE.

3.2 RESULTS

3.2.1 Validation of the BF test on the training dataset from Kaggle

The BF test was validated using the random permutation test on the training dataset from Kaggle. (**Figure 1**). BMI groups were selected based on the triglycerides quantiles (lowest and highest) (**Figure 1A**). The BF test was used to calculate the difference between BMI variance in the triglyceride groups. The difference was shown to be statistically significant ($p < 0.05$).

Next, the random permutation test was performed, and the distribution of the p-values of this test was obtained (**Figure 1B**). The distribution of p-values followed the uniform distribution (Kolmogorov-Smirnov test p-value = 0.71).

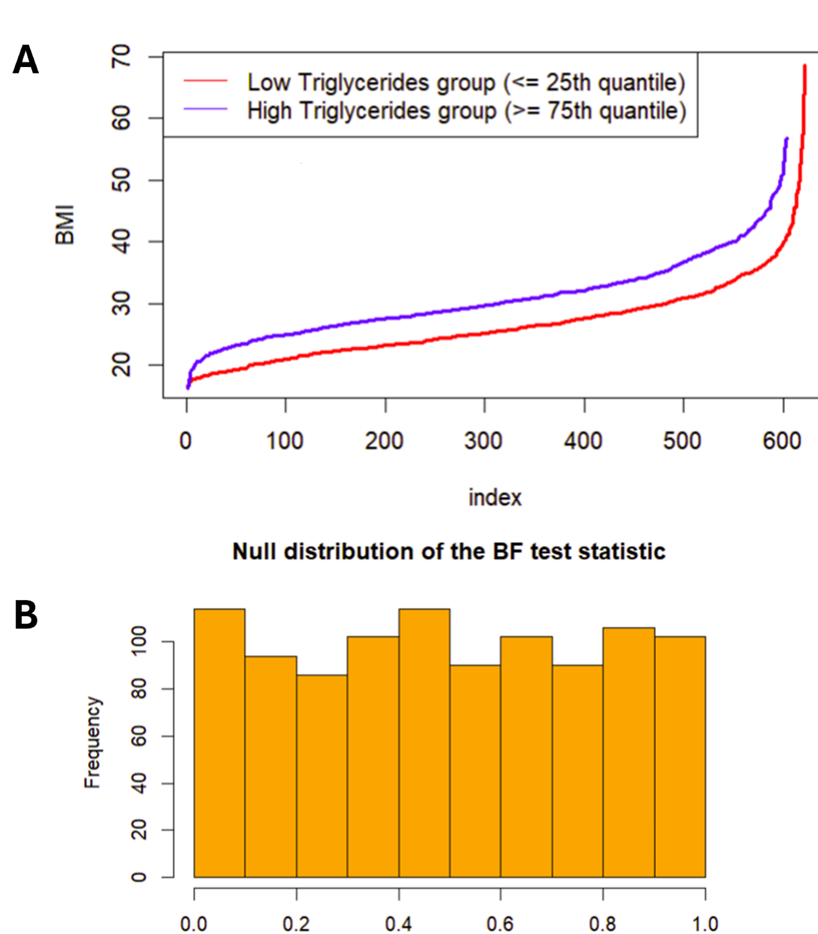

Figure 1. Validation of the BF test on the Kaggle data. The quantile selection, along with the difference between the highest and lowest quantiles, is shown (A). Null distribution of the BF test p-value is shown on (B). It follows the uniform normal distribution with a p-value = 0.71 of the Kolmogorov-Smirnov test.

3.2.2 Validation of the BF test on the Estonian Biobank data.

Next, the BF test was validated using the dataset from the Estonian Biobank. The null distribution showed that the p-value followed the uniform distribution, and the Kolmogorov-Smirnov test confirmed it (**Figure 2**).

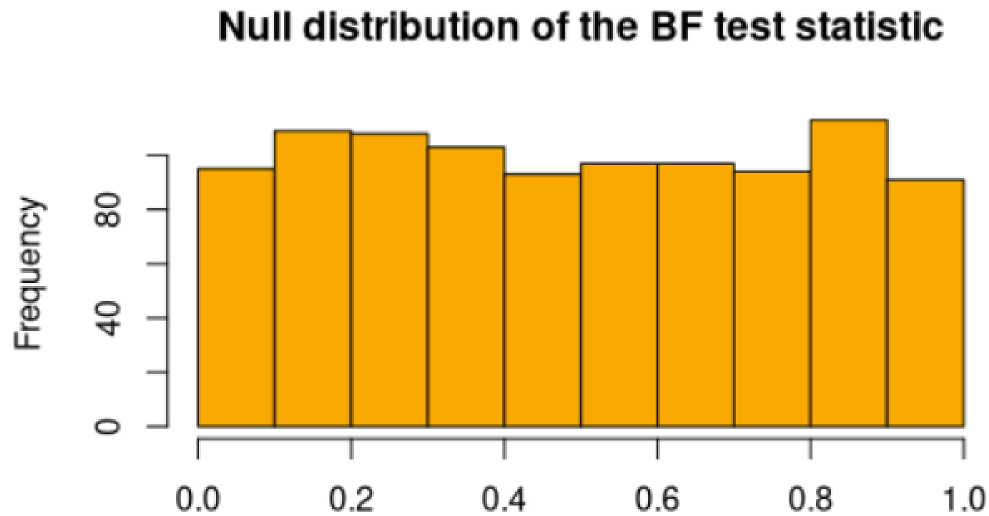

Figure 2. Validation of the BF test on the Estonian Biobank data. The null distribution (repeated 1000 times) is shown for the age group 31-50 females, based on the low physical activity group (0 hours per week) and high physical activity group (5 hours per week). P-value = 0.8593 from Kolmogoro-Smirnov test .

Therefore, the BF test was shown to be a valuable tool for detecting variance differences across the study groups for the trait with skewed distribution, such as BMI or WHR.

3.2.3 EstBB dataset summary statistics

After validating the BF test, the general overview of the Estonian Biobank data was obtained (**Table 1**). The dataset was separated into 4 age groups and two genders: males and females. This was done to ensure that the effect of age and gender is studied separately.

Table 1. Distribution of the individuals in the Estonian Biobank data separately for age groups and gender.

Age group	Female	Male
18-30 years	6963	4763
31-50 years	10040	4688
51-65 years	6512	3152
over 65 years	1829	1077

3.2.4 Genotype selection based on the BF test

The next step was to apply the BF test to select genotypes that were significantly associated with BMI variance in the Estonian Biobank dataset. (**Table 2**). Due to time constraints, the two genes that were previously reported to be highly associated with BMI variance were selected for further analysis: FTO (**rs9937053**), MC4R (**rs6567160**).

Table 2. Bf test p-value for some of the genes that were significantly associated with high BMI variance across four age groups. The names of genes from the genotype data are in format of V+serial number of the gene. V570 and V615 are FTO and MC4R correspondingly.

18-30 years	BF test p-value	31-50 years	BF test p-value	51-65 years	BF test p-value	>65 years	BF test p-value
V1	2.71E-02	V14	2.84E-02	V17	2.58E-02	V4	0.015213117
V17	1.91E-02	V15	3.49E-02	V25	9.07E-03	V18	0.041239636
V24	1.99E-02	V16	8.39E-04	V56	1.19E-02	V22	0.010992753
V25	1.20E-03	V23	1.87E-03	V59	9.11E-03	V36	0.040962399
V35	2.85E-04	V25	2.21E-06	V69	1.89E-02	V56	0.004752166
V36	5.50E-03	V26	4.09E-05	V93	3.11E-02	V61	0.028542826
V40	1.98E-02	V34	2.45E-02	V118	3.56E-02	V70	0.036634781
V44	8.79E-04	V35	6.07E-05	V139	4.34E-02	V76	0.049355505
V53	2.56E-02	V38	5.03E-03	V162	1.73E-02	V112	0.047389734
V56	3.31E-04	V39	4.51E-02	V165	4.18E-02	V136	0.031713029
V66	1.93E-02	V40	4.11E-05	V177	5.44E-03	V150	0.009810232
V72	1.01E-02	V44	3.86E-03	V199	1.96E-02	V155	0.00167514
V76	2.90E-02	V54	1.35E-02	V200	4.16E-02	V160	0.024915118
V86	2.77E-02	V56	1.31E-08	V210	2.42E-04	V187	0.005387607
V88	3.10E-02	V57	2.99E-02	V232	2.40E-03	V201	0.018426486
V100	3.53E-02	V61	1.96E-03	V252	1.88E-03	V202	0.043209456
V110	1.80E-02	V66	1.37E-02	V261	1.95E-02	V238	0.024254371
V116	3.66E-02	V73	4.79E-02	V309	4.27E-02	V252	0.012006261
V123	4.12E-03	V86	3.18E-02	V326	2.48E-02	V253	0.042440401
V126	4.73E-02	V93	4.58E-02	V334	4.09E-02	V261	0.025692804
V127	1.93E-02	V94	3.97E-02	V375	1.53E-02	V262	0.043033487
V128	2.74E-02	V97	1.20E-02	V382	3.99E-02	V287	0.007274869
V132	2.95E-02	V108	3.79E-02	V400	2.19E-03	V295	0.041324656
V133	1.29E-02	V110	3.14E-02	V403	5.57E-03	V305	0.003117648
V135	2.54E-02	V112	2.55E-03	V407	4.51E-03	V316	0.000477658
V153	4.12E-02	V114	5.49E-04	V412	2.73E-03	V326	0.042806099
V155	2.25E-02	V124	2.19E-03	V416	2.30E-02	V336	0.043438057
V160	1.45E-02	V131	2.32E-02	V426	4.55E-02	V362	0.016718905
V165	7.37E-04	V137	4.82E-02	V430	3.35E-02	V364	0.025155886
V570	**5.54E-12**	**V570**	**3.12E-18**	**V570**	**3.97E-10**		
V615	**2.39E-04**	**V615**	**5.70E-11**	**V615**	**2.86E-03**		

3.2.5 Phenotypic traits' variance difference between defined groups

To select the phenotypes and genotypes that were associated with high BMI variance, the BF test was used to test for variance heterogeneity of BMI and WHR across different phenotype groups (**Tables 3,4,5,6**). The phenotypes that were tested were Physical activity, smoking, working in shifts last 3 months, and education level. The analysis was carried out for different age groups and males and females separately. In the case of working in shifts for the last three months, there was no significant difference in BMI and WHR variance, hence, it was not tested for GxE further (**Table 5**). However, the other phenotypic traits showed a significant variance difference in BMI at some age groups for both BMI and WHR. Furthermore, in the case of physical activity, the difference in variance was significant for all of the age groups and genders for BMI (**Table 3**). In the case of education level, the first age group was formed from the individuals of 25-30 years for the sake of correct and comfortable separation into educational level categories, as the individuals from the age of 25 will represent people who should have definitely finished the higher education in most cases (**Table 6)**.

Table 3. BMI and WHR variance difference based on physical activity. Orange cells indicate the statistically significant difference in BMI and WHR without the Bonferroni correction. Blue cells indicate the statistically significant results for the Bonferroni-adjusted p-value.

Diffrence in BMI variance based on the physical activity			
Age group	**Females**	**Males**	
18-30 years	1.27E-11	3.36E-05	**BF test p-values**
31-50 years	< 2.2e-16	4.66E-09	
51-65 years	1.85E-10	0.002428	
over 65 years	0.0005495	0.04024	

Diffrence in WHR variance based on the physical activity			
Age group	**Females**	**Males**	
18-30 years	1.54E-04	3.93E-01	**BF test p-values**
31-50 years	1.67E-10	1.99E-04	
51-65 years	3.40E-06	3.10E-05	
over 65 years	0.2678	0.1006	

Table 4. BMI and WHR variance difference based on smoking. Orange cells indicate the statistically significant difference in BMI and WHR without the Bonferroni correction. Blue cells indicate the statistically significant results for the Bonferroni-adjusted p-value.

Diffrence in BMI variance based on smoking			
Age group	**Females**	**Males**	
18-30 years	9.54E-04	1.85E-01	**BF test p-values**
31-50 years	0.2705	2.29E-07	
51-65 years	9.04E-03	0.2625	
over 65 years	0.7322	0.5808	

Diffrence in WHR variance based on smoking			
Age group	**Females**	**Males**	
18-30 years	0.001226	0.7989	**BF test p-values**
31-50 years	0.1684	2.31E-02	
51-65 years	0.5651	0.00115	
over 65 years	0.1597	0.4717	

Table 5. BMI and WHR variance difference based on the working in shift last 3 months variable. Orange cells indicate the statistically significant difference in BMI and WHR without the Bonferroni correction. Blue cells indicate the statistically significant results for the Bonferroni-adjusted p-value.

Diffrence in BMI variance based on working in shifts in the last 3 months			
Age group	**Females**	**Males**	
18-30 years	0.2296	0.5275	**BF test p-values**
31-50 years	0.0612	0.5805	
51-65 years	0.5226	0.6895	
over 65 years	0.8508	0.2455	

Diffrence in WHR variance based on working in shifts in the last 3 months			
Age group	**Females**	**Males**	
18-30 years	0.4337	0.09	**BF test p-values**
31-50 years	0.2848	0.7654	
51-65 years	0.5197	0.9203	
over 65 years	0.5505	0.2084	

Table 6. BMI and WHR variance difference based on the education level variable. Orange cells indicate the statistically significant difference in BMI and WHR without the Bonferroni correction. Blue cells indicate the statistically significant results for the Bonferroni-adjusted p-value.

Diffrence in BMI variance based on the education level			
Age group	**Females**	**Males**	
25-30 years	3.61E-06	3.17E-02	**BF test p-values**
31-50 years	4.78E-08	0.002552	
51-65 years	1.66E-07	0.02317	
over 65 years	0.03853	0.007771	

Diffrence in WHR variance based on the education level			
Age group	**Females**	**Males**	
25-30 years	1.08E-02	1.64E-01	**BF test p-values**
31-50 years	0.02784	1.77E-01	
51-65 years	3.45E-01	0.09886	
over 65 years	0.2761	0.3088	

In the case of the smoking variable, three categories were created ("Former", "Current", "Never"). The BF-test showed a significant variance difference in the case of females (18-30 years group) and the 51-65 years group for BMI (**Table 4**). In case of males, only the 31-50

years group showed significant results (**Table 4**). In contrast, for WHR, females (18-30 years group) and males (31-50, 51-65 years) showed significant results.

3.2.6 GxE of the strongest candidates from the phenotypic variables

After selecting the candidate phenotypes and genes, the GxE testing was performed using linear modeling to detect significant variance differences.

3.2.6.1 FTO and physical activity

Physical activity vs BMI plots for separate age groups, separated by gender (**Figure 3**), were created for both MC4R and FTO. Physical activity ranged from 0 to 5 hours per week, as most data points were located in that range. In the case of male individuals, average BMI is different across the genotype groups at 0 hours of physical activity, with the higher BMI corresponding to the carriers of two FTO risk alleles (genotype 2) (**Figure 3E, F, G**). With the increasing physical activity, BMI stays stable for the genotypes 0 and 1; however, for the genotype 2, it decreases over time, suggesting that the effect of the genetic predisposition to higher BMI is negated with the high physical activity levels. A similar pattern can be observed for females as well (**Figure 3 B, C, D)**

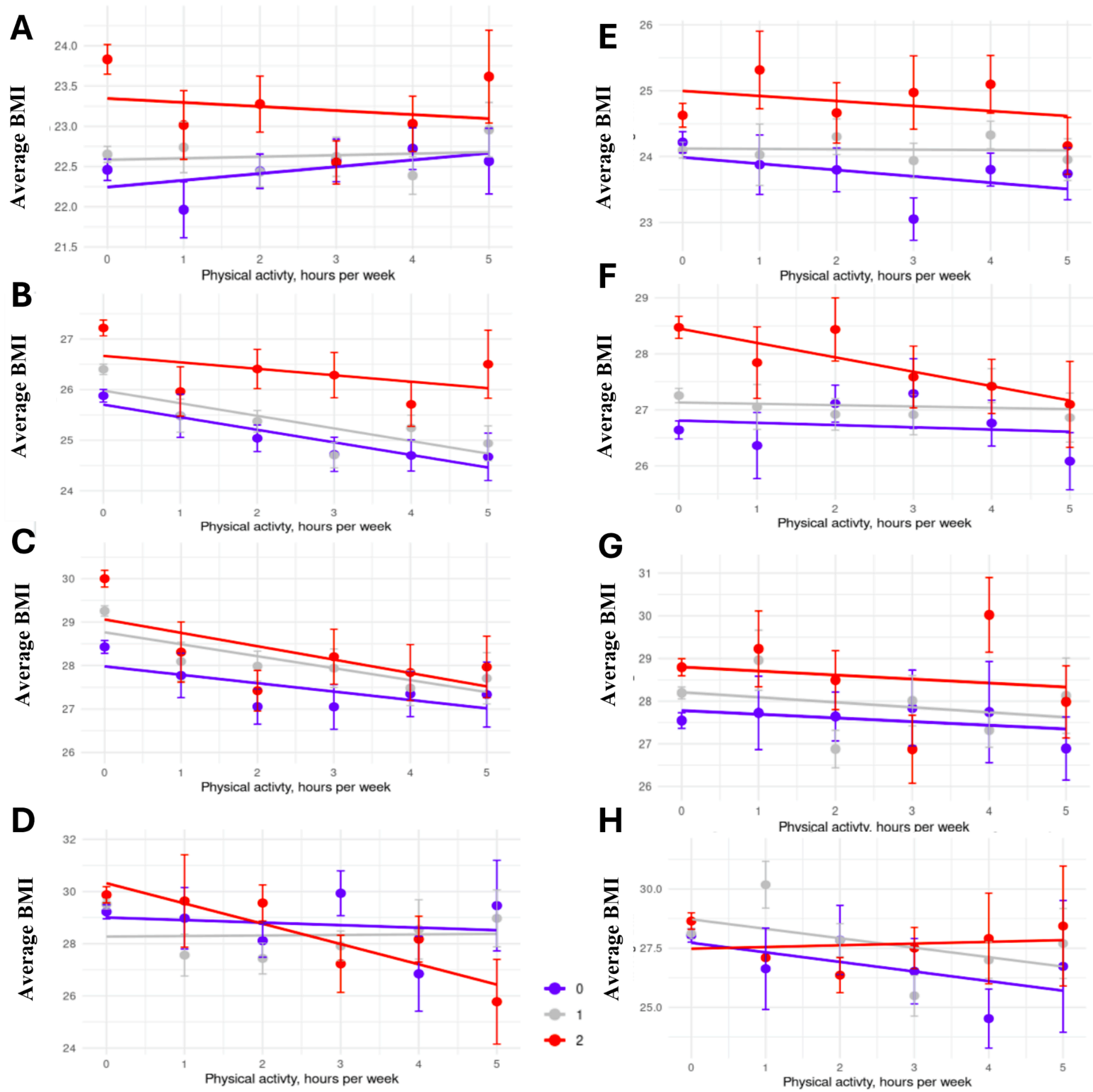

Figure 3. Physical activity and average BMI for different age groups, for females(A-D) and males (E-H) for each of the FTO genotypes (0,1,2). Age groups were defined as follows: age group (18 - 30 y.o.) (A, E); age group (31 - 50 y.o.) (B, F); age group (51 - 65 y.o.) (C, D); age group (> 65 y.o.) (D, H). Colored lines indicate linear model trend lines. Error bars indicate the standard error (SE) of the mean.

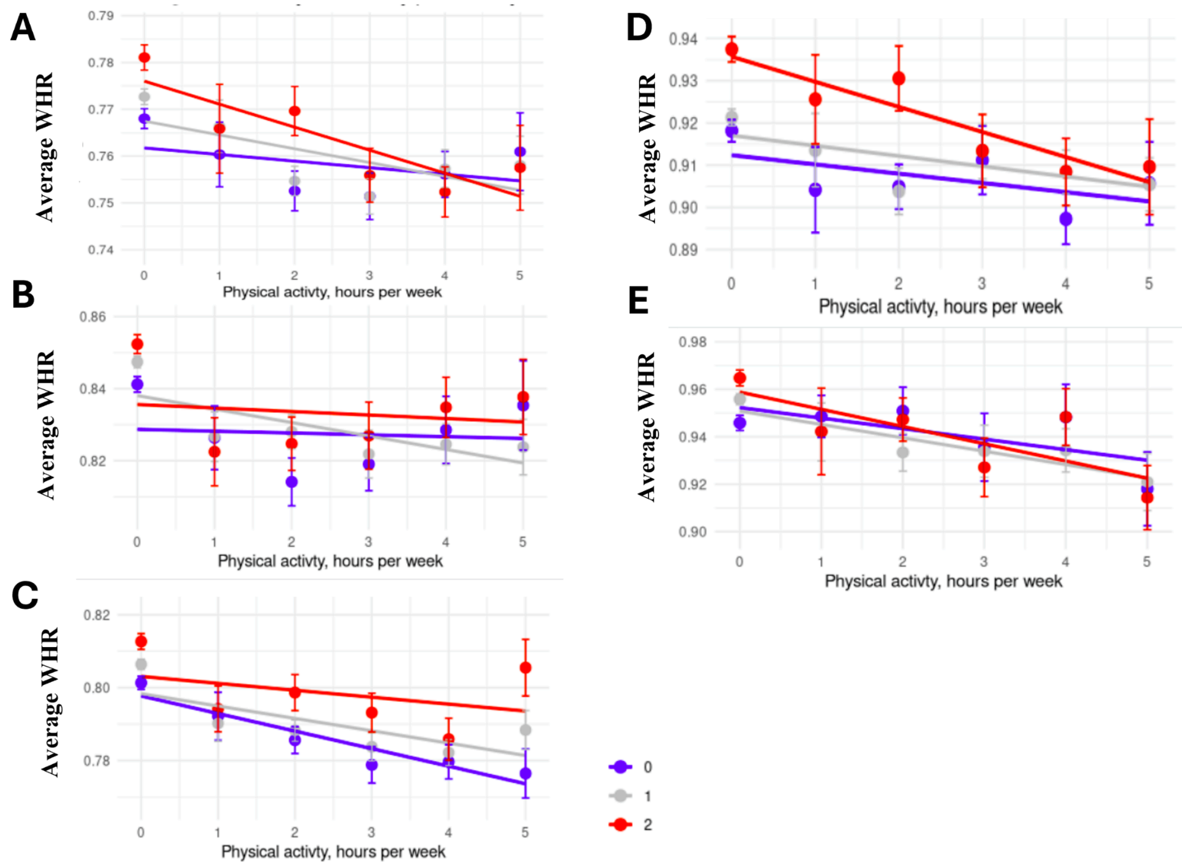

Figure 4. Physical activity and average WHR for different age groups, for females (A-C) and males (D-E) for each of the FTO genotypes (0,1,2). Age groups were as follows: 18-30 (A), 31-50 (B), 51-65(C) for females. For males, the age groups were 31-50 and 51-65 years. Colored lines indicate linear model trend lines. Error bars indicate the standard error (SE) of the mean.

In the case of physical activity and FTO, the significant interaction was detected for females (age 18-30) (p = 0.0106) for BMI (**Figure 3A**) in the case of SNP2 compared to SNP0. Furthermore, in the case of male individuals, the significant interaction was detected for the age group 31-50 years, for FTO and physical activity and BMI, with p = 0.03465 **(Figure 3F)** in the case of SNP2 compared to SNP0**.** Similarly, in the case of WHR as the outcome, for FTO and physical activity, the significant interaction was detected for females (age group 18-30) with p = 0.03969 (**Figure 4A**). There was no significant interaction for the rest of the age groups, for both females and males (**Figure 3, Figure 4**).

3.2.6.2 MC4R vs Physical activity

The next combination for the interaction testing was MC4R and physical activity for both BMI (**Figure 5**) and WHR (**Figure 6**). Significant interaction was detected for the MC4R and

physical activity for BMI in the case of males of the 18-30 age group (**Figure 5E**). The detected p-value was 0.0244 (for SNP2 compared to SNP0). No significant interaction was detected for the rest of the study groups, for both BMI (**Figure 5A, B, C, D, F, G, H**) and for the WHR (**Figure 6**).

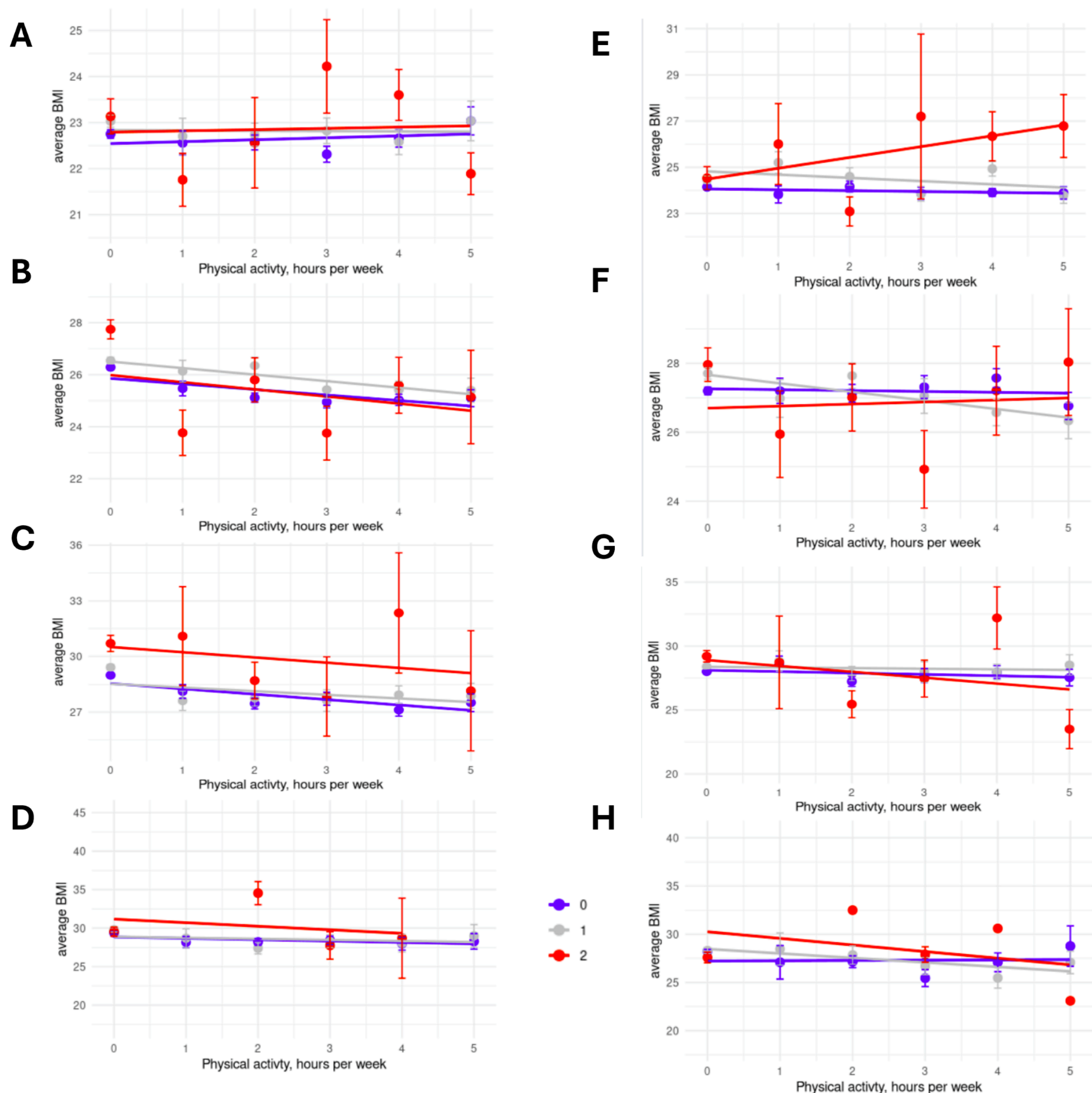

Figure 5. Physical activity and average BMI for different age groups, for females (A-D) and males (E-H) for each of the MC4R genotypes (0,1,2). Age groups were defined as follows: age group (18 - 30 y.o.) (A, E); age group (31 - 50 y.o.) (B, F); age group (51 - 65 y.o.) (C, D); age group (> 65 y.o.) (D, H). Colored lines indicate linear model trend lines.

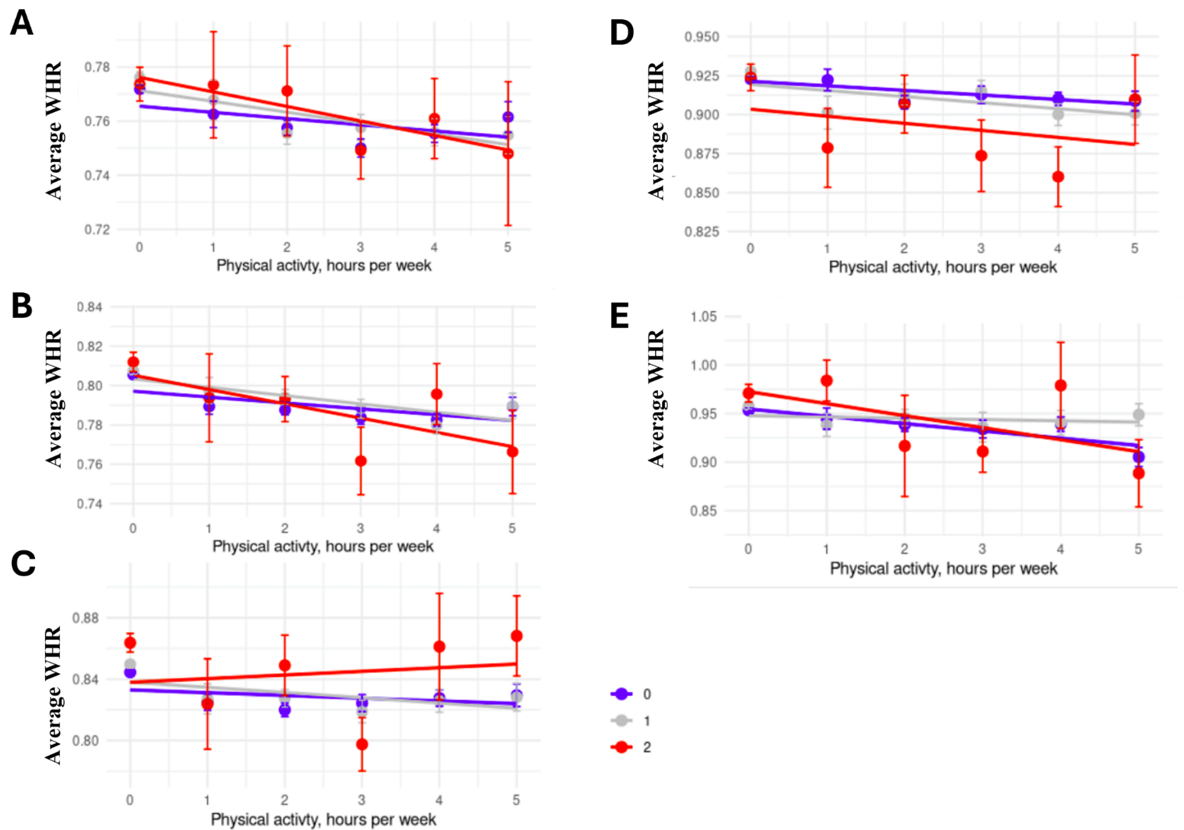

Figure 6. Physical activity and average WHR for different age groups, for females (A-C) and males (D-E) for each of the MC4R genotypes (0,1,2). Age groups were as follows: 18-30 (A), 31-50 (B), 51-65(C) for females. For males, the age groups were 31-50 and 51-65 years. Colored lines indicate linear model trend lines. Error bars indicate the standard error (SE) of the mean.

3.2.6.3 FTO and Smoking

Next, the interaction between FTO and smoking for both BMI and WHR was tested (**Figure 7, Figure 8**).

The significant interaction of FTO and smoking was observed for males in the age group of 31-50 years in the case of BMI for the SNPs 0 and 2 (homozygotes for the reference and risk alleles) **(Figure 7C**)

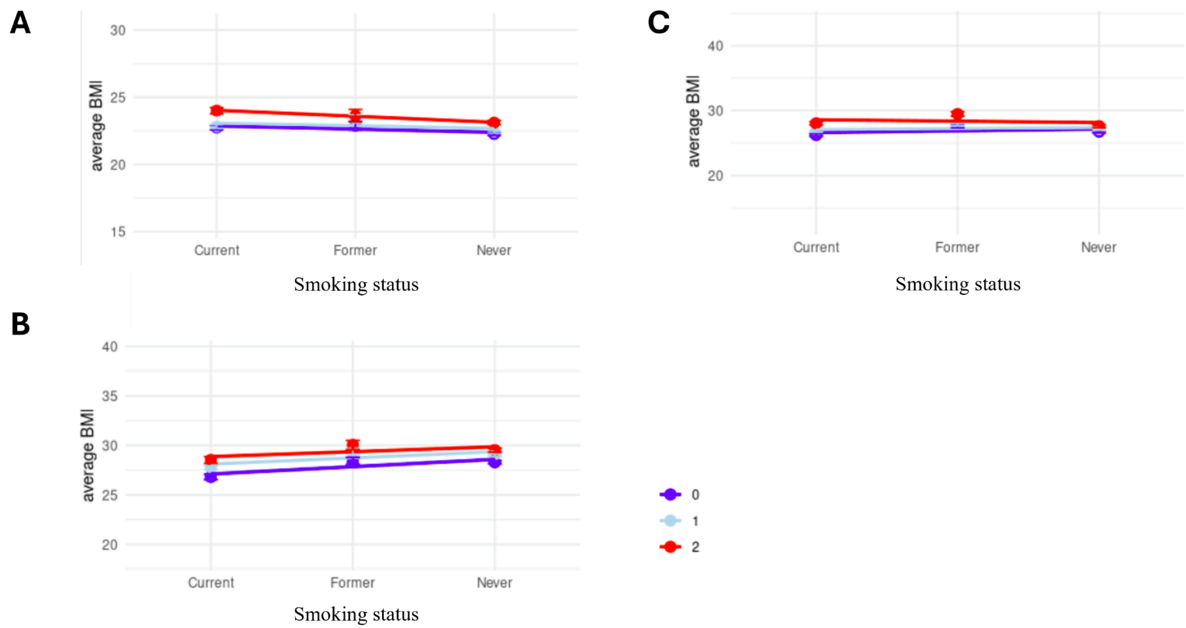

Figure 7. Smoking status vs average BMI for females (A-B) and males (C) for each of the FTO genotypes (0,1,2). Age groups were defined as follows: age group 18-30 (A), 51-65 (B), 31-50 (C). Colored lines indicate linear model trend lines. Error bars indicate the standard error (SE) of the mean.

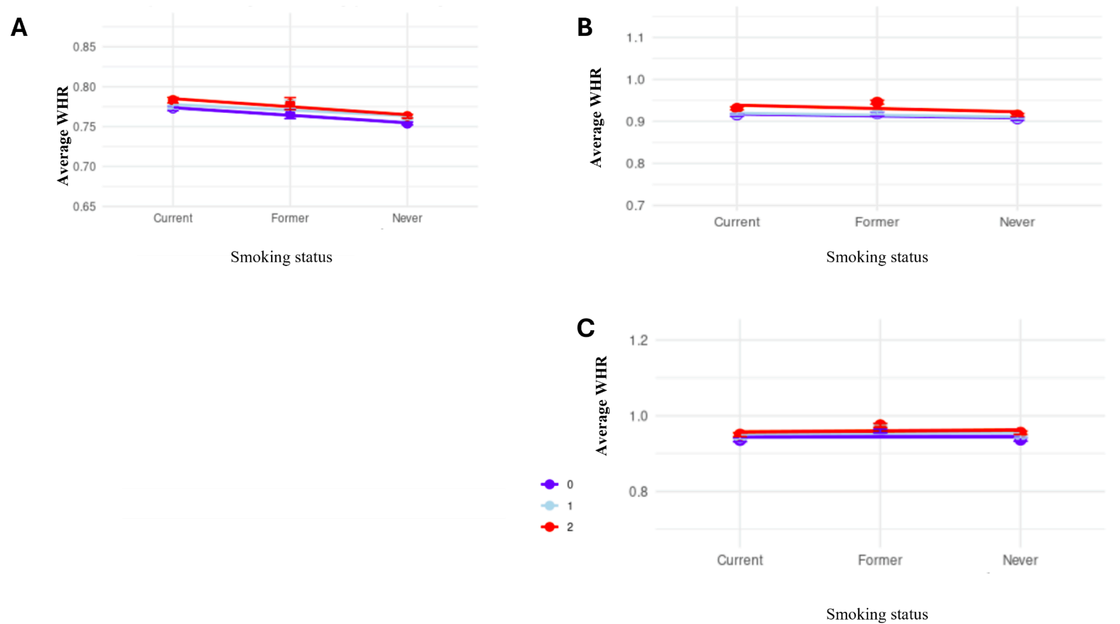

Figure 8. Smoking status vs average WHR for females (A) and males (B, C) for each of the FTO genotypes (0,1,2). Age groups were as follows: 18-30 (A), 31-50 (B), 51-65 (C). Error bars indicate the standard error (SE) of the mean.

The was no evidence for the significant GxE for the rest of the groups of the FTO and smoking, neither for BMI, nor for WHR (**Figure 7, Figure 8**).

3.2.6.4 MC4R and Smoking

As the next step, the interaction of MC4R and smoking was tested for BMI and WHR.

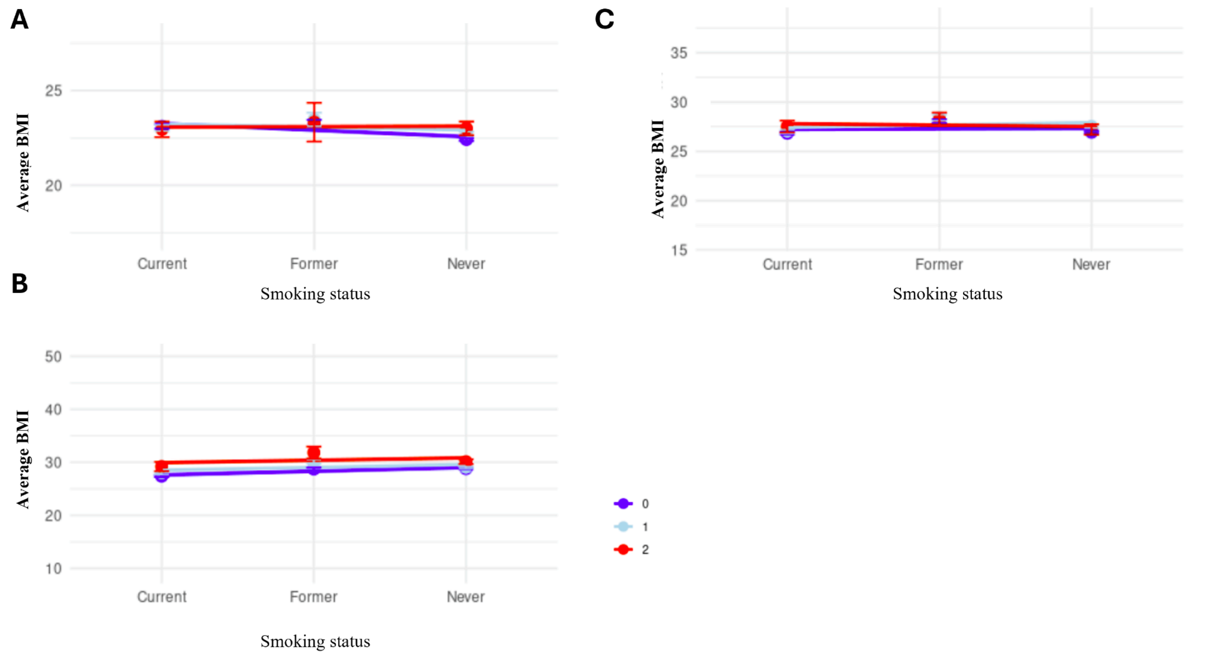

Figure 9. Smoking status vs average BMI for females (A,B) and males (C) for each of the MC4R genotypes (0,1,2). Age groups were defined as follows: age group 18-30 (A), 51-65 (B), 31-50 (C). Error bars indicate the standard error (SE) of the mean.

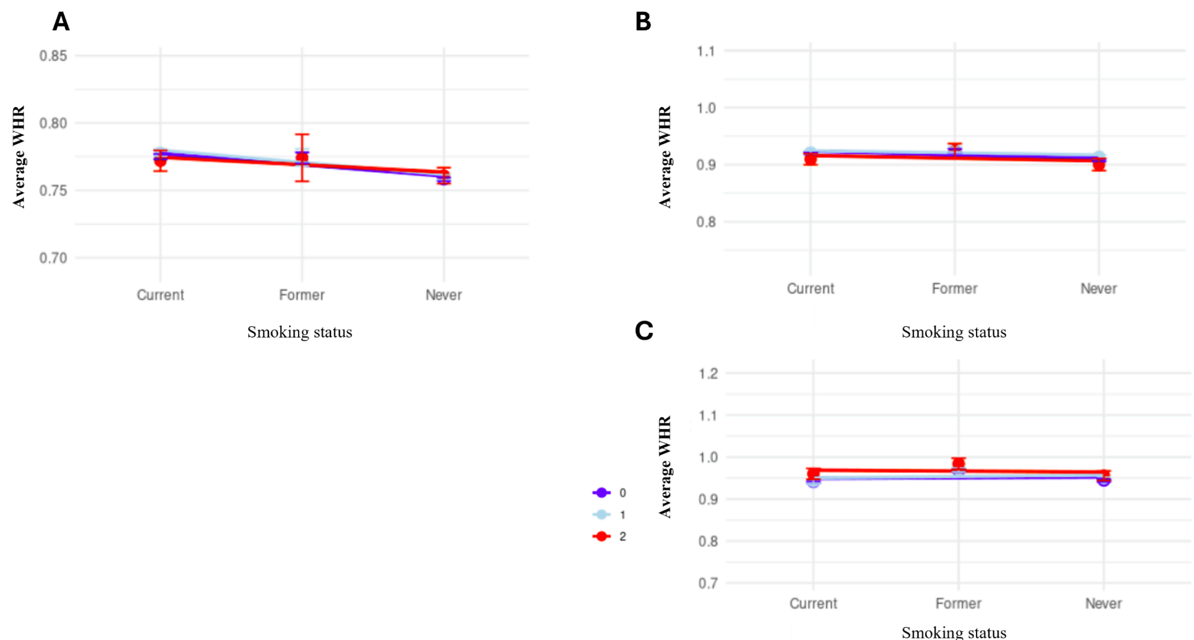

Figure 10. Smoking status vs average WHR for females (A) and males (B, C) for each of the FTO genotypes (0,1,2). Age groups were as follows: 18-30 (A), 31-50 (B), 51-65 (C). Colored lines indicate linear model trend lines. Error bars indicate the standard error (SE) of the mean.

No significant interaction was detected for MC4R and smoking, neither for females nor for males (**Figure 9, Figure 10**). The difference between genotypes is not visible on the plots.

3.2.6.5 FTO and Education level

The final phenotypic trait for interaction testing was the Education Level. It was divided into 3 categories: “Finished community college”, “Finished higher Education”, “Have completed the education up to college level”. The was no significant difference in variance detected for FTO and education level interaction, regardless of the outcome variable (**Figure 11, Figure 12**). The linear model output was insignificant for every combination of the age group and gender.

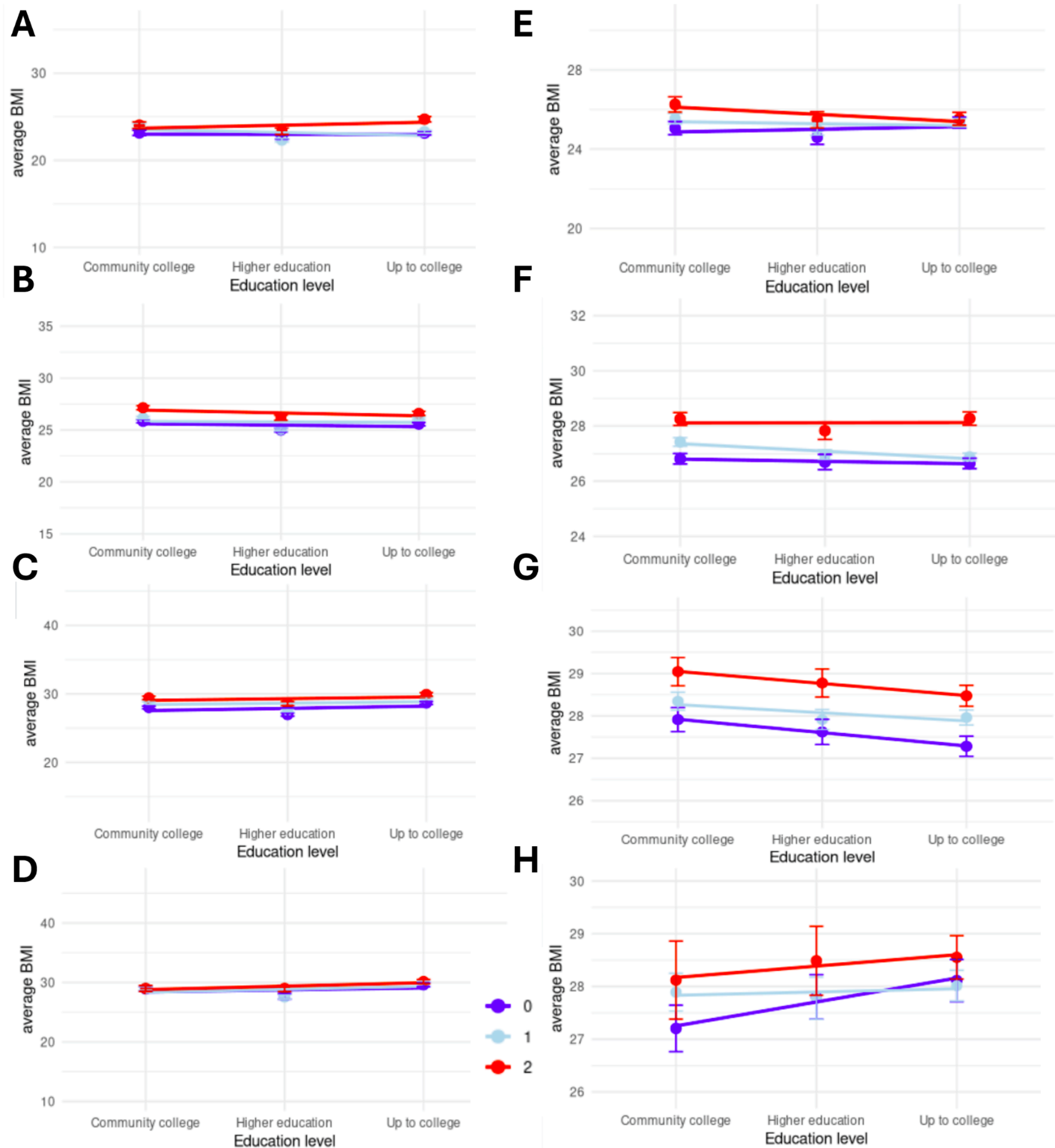

Figure 11. Education levels and average BMI for different age groups, for females (A-D) and males (E-H) for each of the FTO genotypes (0,1,2). Age groups were defined as follows: age group (18 - 30 y.o.) (A, E); age group (31 - 50 y.o.) (B, F); age group (51 - 65 y.o.) (C, D); age group (> 65 y.o.) (D, H). Colored lines indicate linear model trend lines. Error bars indicate the standard error (SE) of the mean.

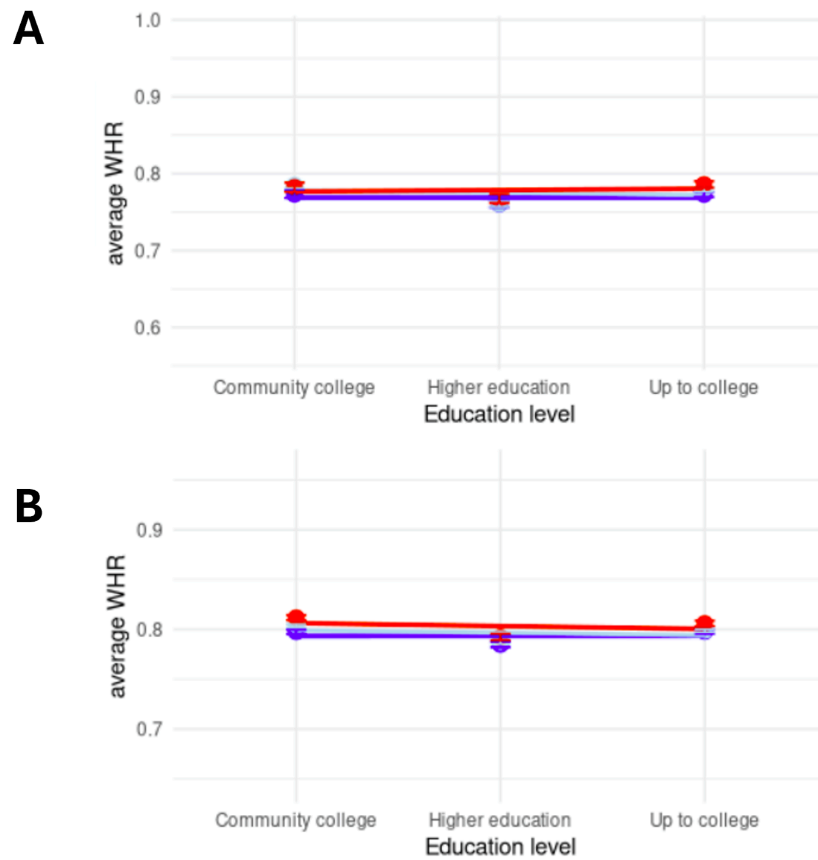

Figure 12. Education levels vs average WHR for females 25-30 years (A), and 31-50 (B) for three of the FTO genotypes (0,1,2). Colored lines indicate linear model trend lines. Error bars indicate the standard error (SE) of the mean.

3.2.6.6 MC4R and Education Level

Finally, the MC4R and the education level were tested for possible GxE interactions for both BMI (**Figure 13**) and WHR(**Figure 14**).

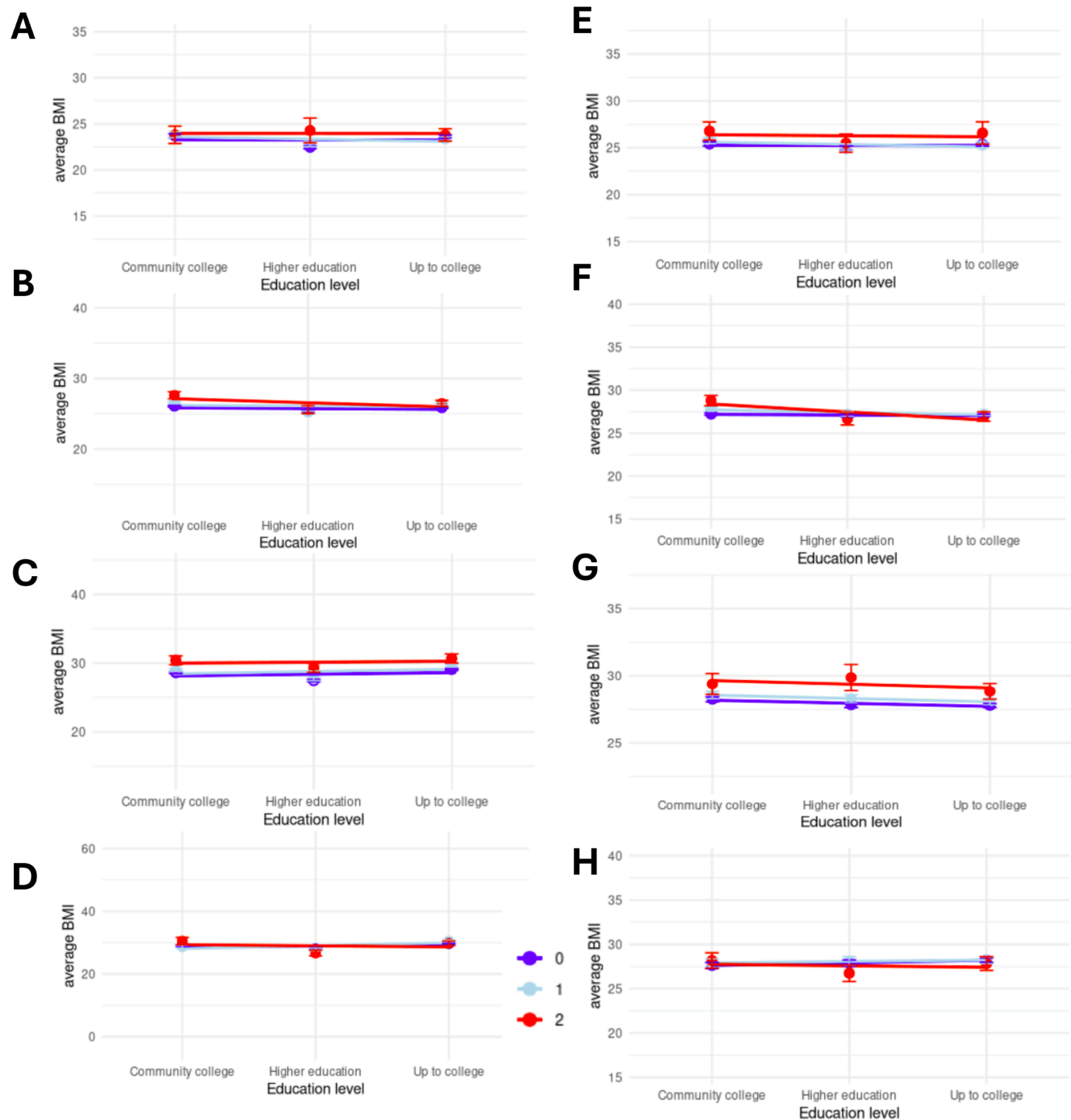

Figure 13. Education levels and average BMI for different age groups, for females (A-D) and males (E-H) for each of the MC4R genotypes (0,1,2). Age groups were defined as follows: age group (18 - 30 y.o.) (A, E); age group (31 - 50 y.o.) (B, F); age group (51 - 65 y.o.) (C, D); age group (> 65 y.o.) (D, H). Colored lines indicate linear model trend lines. Error bars indicate the standard error (SE) of the mean.

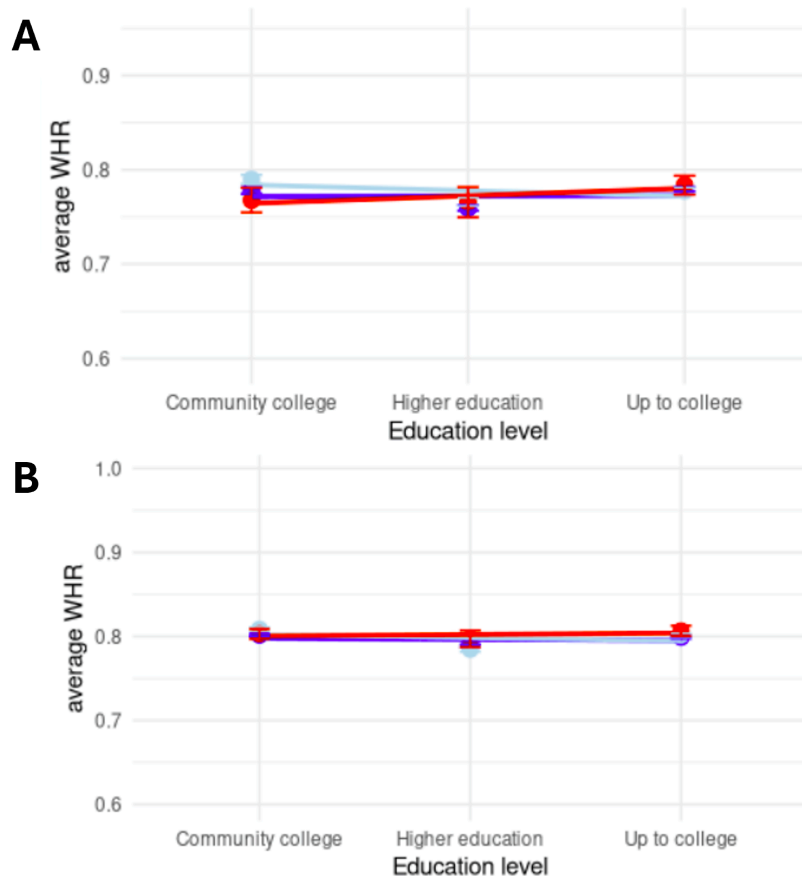

Figure 14. Education levels vs average WHR for females 25-30 years (A), and 31-50 (B) for three of the MC4R genotypes (0,1,2). Colored lines indicate linear model trend lines. Error bars indicate the standard error (SE) of the mean.

In the case of males in the age group of 31-50 years, the significant interaction was detected for the BMI as the outcome variable (p = 0.03213) (**Figure 13F**). In any other cases, the was no significant difference in the BMI or WHR variances based on the MC4R and education level. The difference between genotypes cannot be distinguished visually form the plots.

3.3 DISCUSSION

3.3.1 The Brown-Forsythe test proved to be robust against skewed data

In this work, the Brown-Forsythe test was tested for its ability to detect heteroskedasticity in the phenotypic traits based on different group factors, such as another phenotypic trait or genotypes. The test has been shown to be effective in detecting significant BMI or WHR variance difference, and was not affected by the skewed distributions of BMI and WHR. The BF test is more robust in the case of non-normally distributed (skewed) data compared to Levene's test. The reason behind this is that the BF test utilizes the median value of the group, while Levene's test uses the mean value (Wang et al., 2017; Zhang & Bell, 2024), making the BF test a valuable tool for detecting the variance heterogeneity in non-normally distributed data.

3.3.2 Difference in response to the environment between males and females across different age groups

The observed heteroscedasticity of BMI and WHR varied between age groups and genders. For example, in the case of physical activity, the significance of variance difference decreased with increasing age in both males and females. This effect could be explained by the overall decreased level of physical activity in individuals of higher age groups. Furthermore, the changes in metabolic state of the organism with age can account for the difference in metabolic rates and homeostatic reactions in the organism, resulting in less significant differences in variance in such groups. Moreover, the difference in BMI variance was overall more significant for females, compared to males. A possible explanation can be connected with different hormonal content and differences in metabolism among males and females (Muscogiuri et al., 2024).

3.3.3 Possible biological explanations of heteroscedasticity

Next, in this study, the genotypes and phenotypic traits that were associated with significant variance differences were examined. Phenotypic traits included physical activity, smoking, and education level. And the genotypes were FTO and MC4R, which were previously reported to be significantly associated with BMI variability. One possible explanation for the observed heteroscedasticity in BMI is that there is a genetic or environmental factor that influences the complex regulatory network of appetite and energy expenditure such that the strength and effectiveness of the homeostatic responses to the environment decrease. As a result, the balances maintained by homeostatic mechanisms are more easily perturbed by diverse environmental influences, resulting in increased variance of BMI and WHR.

Another explanation is that the variance heteroskedasticity could be caused by gene-environment interactions (GxE). GxE represents the case where the variance is different depending on the particular genotype for the specific gene that is highly associated with BMI variance. In this study, the test for GxE was done for FTO and MC4R and several environmental exposures, such as physical activity, smoking, and education level. The evidence for significant GxE was observed for the FTO and smoking in the age group 31-50 years for the SNPs 0 and 2 (homozygotes for the reference and risk alleles). Furthermore, the effect of FTO genotypes on BMI was negated with the increased physical activity levels for males and females, which is similar to the effect shown by Bjørnland et al., 2017 for FTO and physical activity for the age group 31-50. Furthermore, the evidence for significant interaction of FTO with physical activity for Females (18-30 age group) and males (31-50 age group) in the case of BMI as the outcome variable. Moreover, the evidence for significant interaction of FTO and physical activity was found for females (age group 18-30) for the WHR as the outcome variable. Furthermore, the significant interaction was observed in the case of MC4R and physical activity for BMI in males (18-30 year old group). Additionally, it was shown that the significant interaction between MC4R and education level for males (31-50 age group) was present for BMI.

Moreover, the estimated GxE was not significant for the low and high activity groups for both FTO and MC4R, which is different from the results reported by Bjørnland et al., 2017.

Thus, gene-environment interactions may be more important in younger age groups, whereas in older age groups, the main contribution from differences in trait variance may come from general dysregulation of energy homeostasis.

3.3.4 Limitations of the study

A fundamental limitation of GxE modeling is that they do not fully represent the concept of the biology that maintains the stability of the internal processes, such as BMI regulation, thermal regulation, glucose levels, and energy expenditure regulation. It opens up the potential for establishing and designing new experiments in vivo to better understand the internal regulatory mechanisms in connection with GxE modeling. However, for now, it is not clear what types of modeling and experiments can be suitable for such an approach.

In this work, linear modeling was used to estimate the GxE interactions on the biobank data. However, in several cases, the data did not follow the simple linear trend, which made it harder to estimate the GxE. In such cases, more complex and non-linear models could be used

to further investigate the potential GxE and account for non-linear trends. Furthermore, the number of participants in the age groups was not equal among the study population, and some data contained outliers, which could affect the statistical analyses.

Additionally, due to time constraints, only two significant genes (FTO, MC4R) that were previously reported to be associated with high BMI variability were used to study and model GxE.

3.3.5 Future perspectives

Future work would include more precise statistical modeling, including non-linear models, to better fit the data and carefully investigate possible GxE.

Furthermore, more gene candidates can be investigated for potential GxE along with different environmental exposures, which were not investigated in this work. For example, the alcohol consumption and sleeping time can be tested for variance heteroscedasticity. However, more data preprocessing could be done for some environmental variables, such as sleeping time, to make it suitable for variance testing and group separation.

Finally, more research on expanding the understanding of internal regulatory mechanisms, such as BMI regulation, glucose, and thermal regulation, etc., in connection with GxE modeling can be done. This will allow for more precise modeling that could more comprehensively describe the effect of GxE and internal mechanisms on response to the environment.

SUMMARY

Obesity is a rising global problem, clinically defined as having a body mass index (BMI) > 30 or a waist-to-hip ratio >0.9 for males and >0.85 for females. It is a complex disease with both genetic and environmental factors affecting its development and progression. However, although BMI is highly heritable (60-90% according to twin studies), only ~6% of the variance is explained by additive genetic variation discovered in large genome-wide association studies. Possible sources of unexplained BMI variance are gene-gene (GxG) and gene-environment (GxE) interactions, but to date, only a few studies have addressed this question, and mostly focused on the UK biobank and data from smaller longitudinal studies

This study used data from the Estonian biobank to investigate the potential of using the Brown-Forsythe (BF) test to detect heteroscedasticity in BMI based on the genotype and phenotype groups. Moreover, in this work, the BF test was used to select genotypes and environmental variables that were significantly associated with BMI and WHR variance. The study demonstrated sex and age-dependent differences in heteroscedasticity, with higher levels of heteroscedasticity in females compared to males and decreasing heteroscedasticity with increasing age for both sexes.

Next, the gene-environment interactions for the outcome variables BMI and WHR were modeled and studied for two genes (FTO and MC4R) and environmental factors, such as physical activity, smoking, and education level. The study found evidence for significant interaction of FTO with smoking for the age group of 31-50 years for males, along with significant interaction of FTO with physical activity for Females (18-30 age group) and males (31-50 age group) in the case of BMI as the outcome variable. Moreover, the evidence for significant interaction of FTO and physical activity was found for females (age group 18-30) for the WHR as the outcome variable. Furthermore, the significant interaction was observed in the case of MC4R and physical activity for BMI in males (18-30 year old group). Additionally, it was shown that the significant interaction between MC4R and education level for males (31-50 age group) was present for BMI.

Finally, the study outlined the ground for future research work in the area of gene-environment interactions and statistical modeling, indicating potential environmental factors that can be studied for possible gene-environment interactions.

REFERENCES

Afshin, A., Forouzanfar, M. H., Reitsma, M. B., Sur, P., Estep, K., Lee, A., Marczak, L., Mokdad, A. H., Moradi-Lakeh, M., Naghavi, M., Salama, J. S., Vos, T., Abate, K. H., Abbafati, C., Ahmed, M. B., Al-Aly, Z., Alkerwi, A., Al-Raddadi, R., Amare, A. T., … GBD, 2015 Obesity Collaborators. (2017). Health effects of overweight and obesity in 195 countries over 25 years. *New England Journal of Medicine*, *377*(1), 13–27. Scopus. https://doi.org/10.1056/NEJMoa1614362

Ahmad, T., Lee, I.-M., Paré, G., Chasman, D. I., Rose, L., Ridker, P. M., & Mora, S. (2011). Lifestyle interaction with fat mass and obesity-associated (FTO) genotype and risk of obesity in apparently healthy U.S. women. *Diabetes Care*, *34*(3), 675–680. https://doi.org/10.2337/dc10-0948

Albuquerque, D., Nóbrega, C., Manco, L., & Padez, C. (2017). The contribution of genetics and environment to obesity. *British Medical Bulletin*, *123*(1), 159–173. https://doi.org/10.1093/bmb/ldx022

Anholt, R. R. H., & Mackay, T. F. C. (2018). The road less traveled: From genotype to phenotype in flies and humans. *Mammalian Genome*, *29*(1), 5–23. https://doi.org/10.1007/s00335-017-9722-7

Berglund, E. D., Liu, C., Sohn, J.-W., Liu, T., Kim, M. H., Lee, C. E., Vianna, C. R., Williams, K. W., Xu, Y., & Elmquist, J. K. (2013). Serotonin 2C receptors in pro-opiomelanocortin neurons regulate energy and glucose homeostasis. *The Journal of Clinical Investigation*, *123*(12), 5061–5070. https://doi.org/10.1172/JCI70338

Bjørnland, T., Langaas, M., Grill, V., & Mostad, I. L. (2017). Assessing gene-environment interaction effects of FTO, MC4R and lifestyle factors on obesity using an extreme phenotype sampling design: Results from the HUNT study. *PLOS ONE*, *12*(4), e0175071. https://doi.org/10.1371/journal.pone.0175071

Bochukova, E. G., Huang, N., Keogh, J., Henning, E., Purmann, C., Blaszczyk, K., Saeed, S., Hamilton-Shield, J., Clayton-Smith, J., O'Rahilly, S., Hurles, M. E., & Farooqi, I. S. (2010). Large, rare chromosomal deletions associated with severe early-onset obesity. *Nature*, *463*(7281), 666–670. https://doi.org/10.1038/nature08689

Bornstein, S. R., Guan, K., Brunβen, C., Mueller, G., Kamvissi-Lorenz, V., Lechler, R., Trembath, R., Mayr, M., Poston, L., Sancho, R., Ahmed, S., Alfar, E., Aljani, B., Alves, T. C., Amiel, S., Andoniadou, C. L., Bandral, M., Belavgeni, A., Berger, I., … Ziegler, C. G. (2021). The transCampus Metabolic Training Programme Explores the Link of SARS-CoV-2 Virus to Metabolic Disease. *Hormone and Metabolic Research*,

53(3), 204–206. Scopus. https://doi.org/10.1055/a-1377-6583

Brown, A. A., Buil, A., Viñuela, A., Lappalainen, T., Zheng, H.-F., Richards, J. B., Small, K. S., Spector, T. D., Dermitzakis, E. T., & Durbin, R. (2014). Genetic interactions affecting human gene expression identified by variance association mapping. *eLife*, *3*, e01381. https://doi.org/10.7554/eLife.01381

Chakhtoura, M., Haber, R., Ghezzawi, M., Rhayem, C., Tcheroyan, R., & Mantzoros, C. S. (2023). Pharmacotherapy of obesity: An update on the available medications and drugs under investigation. *eClinicalMedicine*, *58*. https://doi.org/10.1016/j.eclinm.2023.101882

Dubern, B., Lubrano-Berthelier, C., Mencarelli, M., Ersoy, B., Frelut, M.-L., Bouglé, D., Costes, B., Simon, C., Tounian, P., Vaisse, C., & Clement, K. (2008). Mutational Analysis of the Pro-opiomelanocortin Gene in French Obese Children Led to the Identification of a Novel Deleterious Heterozygous Mutation Located in the α-Melanocyte Stimulating Hormone Domain. *Pediatric Research*, *63*(2), 211–216. https://doi.org/10.1203/PDR.0b013e31815ed62b

Elks, C. E., Den Hoed, M., Zhao, J. H., Sharp, S. J., Wareham, N. J., Loos, R. J. F., & Ong, K. K. (2012). Variability in the Heritability of Body Mass Index: A Systematic Review and Meta-Regression. *Frontiers in Endocrinology*, *3*. https://doi.org/10.3389/fendo.2012.00029

Estonian Biobank. (2024, October 29). https://genomics.ut.ee/en/content/estonian-biobank

Fatima, M. T., Ahmed, I., Fakhro, K. A., & Akil, A. S. A. (2022). Melanocortin‐4 receptor complexity in energy homeostasis,obesity and drug development strategies. *Diabetes, Obesity & Metabolism*, *24*(4), 583–598. https://doi.org/10.1111/dom.14618

Fitch, A. K., Malhotra, S., & Conroy, R. (2024). Differentiating monogenic and syndromic obesities from polygenic obesity: Assessment, diagnosis, and management. *Obesity Pillars*, *11*, 100110. https://doi.org/10.1016/j.obpill.2024.100110

Flores-Dorantes, M. T., Díaz-López, Y. E., & Gutiérrez-Aguilar, R. (2020). Environment and Gene Association With Obesity and Their Impact on Neurodegenerative and Neurodevelopmental Diseases. *Frontiers in Neuroscience*, *14*, 863. https://doi.org/10.3389/fnins.2020.00863

Foster-Schubert, K. E., Alfano, C. M., Duggan, C. R., Xiao, L., Campbell, K. L., Kong, A., Bain, C. E., Wang, C.-Y., Blackburn, G. L., & McTiernan, A. (2012). Effect of Diet and Exercise, Alone or Combined, on Weight and Body Composition in Overweight-to-Obese Postmenopausal Women. *Obesity*, *20*(8), 1628–1638.

https://doi.org/10.1038/oby.2011.76

Frayling, T. M., Timpson, N. J., Weedon, M. N., Zeggini, E., Freathy, R. M., Lindgren, C. M., Perry, J. R. B., Elliott, K. S., Lango, H., Rayner, N. W., Shields, B., Harries, L. W., Barrett, J. C., Ellard, S., Groves, C. J., Knight, B., Patch, A.-M., Ness, A. R., Ebrahim, S., … McCarthy, M. I. (2007). A Common Variant in the FTO Gene Is Associated with Body Mass Index and Predisposes to Childhood and Adult Obesity. *Science (New York, N.Y.)*, *316*(5826), 889–894. https://doi.org/10.1126/science.1141634

Gene–environment interactions in human diseases | Nature Reviews Genetics. (n.d.). Retrieved April 24, 2025, from https://www.nature.com/articles/nrg1578#Bib1

Hopkins, J. C. A., Blazeby, J. M., Rogers, C. A., & Welbourn, R. (2016). The use of adjustable gastric bands for management of severe and complex obesity. *British Medical Bulletin*, *118*(1), 64–72. https://doi.org/10.1093/bmb/ldw012

Jensen, M. D., Ryan, D. H., Apovian, C. M., Ard, J. D., Comuzzie, A. G., Donato, K. A., Hu, F. B., Hubbard, V. S., Jakicic, J. M., Kushner, R. F., Loria, C. M., Millen, B. E., Nonas, C. A., Pi-Sunyer, F. X., Stevens, J., Stevens, V. J., Wadden, T. A., Wolfe, B. M., Yanovski, S. Z., … Obesity Society. (2014). 2013 AHA/ACC/TOS guideline for the management of overweight and obesity in adults: A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and The Obesity Society. *Circulation*, *129*(25 Suppl 2), S102-138. https://doi.org/10.1161/01.cir.0000437739.71477.ee

Johns, D. J., Hartmann-Boyce, J., Jebb, S. A., & Aveyard, P. (2014). Diet or Exercise Interventions vs Combined Behavioral Weight Management Programs: A Systematic Review and Meta-Analysis of Direct Comparisons. *Journal of the Academy of Nutrition and Dietetics*, *114*(10), 1557–1568. https://doi.org/10.1016/j.jand.2014.07.005

Khera, A. V., Chaffin, M., Wade, K. H., Zahid, S., Brancale, J., Xia, R., Distefano, M., Senol-Cosar, O., Haas, M. E., Bick, A., Aragam, K. G., Lander, E. S., Smith, G. D., Mason-Suares, H., Fornage, M., Lebo, M., Timpson, N. J., Kaplan, L. M., & Kathiresan, S. (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

Kloock, S., Ziegler, C. G., & Dischinger, U. (2023). Obesity and its comorbidities, current treatment options and future perspectives: Challenging bariatric surgery? *Pharmacology & Therapeutics*, *251*, 108549.

https://doi.org/10.1016/j.pharmthera.2023.108549

Lazopoulou, N., Gkioka, E., Ntalla, I., Pervanidou, P., Magiakou, A.-M., Roma-Giannikou, E., Chrousos, G. P., Papassotiriou, I., Dedoussis, G., & Kanaka-Gantenbein, C. (2015). The combined effect of MC4R and FTO risk alleles on childhood obesity in Greece. *Hormones (Athens, Greece)*, *14*(1), 126–133. https://doi.org/10.14310/horm.2002.1524

Lee, S. K., Heo, Y., Park, J. M., Kim, Y. J., Kim, S. M., Park, D. J., Han, S. M., Shim, K. W., Lee, Y. J., Lee, J. Y., & Kwon, J. W. (2016). Roux-en-Y Gastric Bypass vs. Sleeve Gastrectomy vs. Gastric Banding: The First Multicenter Retrospective Comparative Cohort Study in Obese Korean Patients. *Yonsei Medical Journal*, *57*(4), 956–962. https://doi.org/10.3349/ymj.2016.57.4.956

Mahmoud, R., Kimonis, V., & Butler, M. G. (2022). Genetics of Obesity in Humans: A Clinical Review. *International Journal of Molecular Sciences*, *23*(19), 11005. https://doi.org/10.3390/ijms231911005

Mahmoudi, T., Farahani, H., Nobakht, H., Dabiri, R., & Zali, M. R. (2016). Genetic Variations in Leptin and Leptin Receptor and Susceptibility to Colorectal Cancer and Obesity. *Iranian Journal of Cancer Prevention*, *9*(3), e7013. https://doi.org/10.17795/ijcp-7013

Messier, S. P., Mihalko, S. L., Legault, C., Miller, G. D., Nicklas, B. J., DeVita, P., Beavers, D. P., Hunter, D. J., Lyles, M. F., Eckstein, F., Williamson, J. D., Carr, J. J., Guermazi, A., & Loeser, R. F. (2013). Effects of Intensive Diet and Exercise on Knee Joint Loads, Inflammation, and Clinical Outcomes Among Overweight and Obese Adults With Knee Osteoarthritis: The IDEA Randomized Clinical Trial. *JAMA*, *310*(12), 1263–1273. https://doi.org/10.1001/jama.2013.277669

Millington, G. W. (2007). The role of proopiomelanocortin (POMC) neurones in feeding behaviour. *Nutrition & Metabolism*, *4*(1), 18. https://doi.org/10.1186/1743-7075-4-18

Obradovic, M., Sudar-Milovanovic, E., Soskic, S., Essack, M., Arya, S., Stewart, A. J., Gojobori, T., & Isenovic, E. R. (2021). Leptin and Obesity: Role and Clinical Implication. *Frontiers in Endocrinology*, *12*, 585887. https://doi.org/10.3389/fendo.2021.585887

Okunogbe, A., Nugent, R., Spencer, G., Powis, J., Ralston, J., & Wilding, J. (2022). Economic impacts of overweight and obesity: Current and future estimates for 161 countries. *BMJ Global Health*, *7*(9), e009773. https://doi.org/10.1136/bmjgh-2022-009773

Pulit, S. L., Stoneman, C., Morris, A. P., Wood, A. R., Glastonbury, C. A., Tyrrell, J., Yengo, L., Ferreira, T., Marouli, E., Ji, Y., Yang, J., Jones, S., Beaumont, R., Croteau-Chonka, D. C., Winkler, T. W., GIANT Consortium, Hattersley, A. T., Loos, R. J. F., Hirschhorn, J. N., … Lindgren, C. M. (2019). Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Human Molecular Genetics*, *28*(1), 166–174. https://doi.org/10.1093/hmg/ddy327

Roh, E., & Choi, K. M. (2023). Hormonal Gut–Brain Signaling for the Treatment of Obesity. *International Journal of Molecular Sciences*, *24*(4), Article 4. https://doi.org/10.3390/ijms24043384

Roh, E., & Kim, M.-S. (2016). Brain Regulation of Energy Metabolism. *Endocrinology and Metabolism*, *31*(4), 519–524. https://doi.org/10.3803/EnM.2016.31.4.519

Ross, R., Dagnone, D., Jones, P. J. H., Smith, H., Paddags, A., Hudson, R., & Janssen, I. (2000). Reduction in Obesity and Related Comorbid Conditions after Diet-Induced Weight Loss or Exercise-Induced Weight Loss in Men. *Annals of Internal Medicine*, *133*(2), 92–103. https://doi.org/10.7326/0003-4819-133-2-200007180-00008

Sacks, F. M., Bray, G. A., Carey, V. J., Smith, S. R., Ryan, D. H., Anton, S. D., McManus, K., Champagne, C. M., Bishop, L. M., Laranjo, N., Leboff, M. S., Rood, J. C., Jonge, L. de, Greenway, F. L., Loria, C. M., Obarzanek, E., & Williamson, D. A. (2009). Comparison of Weight-Loss Diets with Different Compositions of Fat, Protein, and Carbohydrates. *New England Journal of Medicine*, *360*(9), 859–873. https://doi.org/10.1056/NEJMoa0804748

Sandholt, C. H., Sparsø, T., Grarup, N., Albrechtsen, A., Almind, K., Hansen, L., Toft, U., Jørgensen, T., Hansen, T., & Pedersen, O. (2010). Combined Analyses of 20 Common Obesity Susceptibility Variants. *Diabetes*, *59*(7), 1667–1673. https://doi.org/10.2337/db09-1042

Šket, R., Kotnik, P., Bizjan, B. J., Kocen, V., Mlinarič, M., Tesovnik, T., Debeljak, M., Battelino, T., & Kovač, J. (2022). Heterozygous Genetic Variants in Autosomal Recessive Genes of the Leptin-Melanocortin Signalling Pathway Are Associated With the Development of Childhood Obesity. *Frontiers in Endocrinology*, *13*, 832911. https://doi.org/10.3389/fendo.2022.832911

Slentz, C. A., Duscha, B. D., Johnson, J. L., Ketchum, K., Aiken, L. B., Samsa, G. P., Houmard, J. A., Bales, C. W., & Kraus, W. E. (2004). Effects of the amount of exercise on body weight, body composition, and measures of central obesity:

STRRIDE--a randomized controlled study. *Archives of Internal Medicine*, *164*(1), 31–39. https://doi.org/10.1001/archinte.164.1.31

Sonestedt, E., Roos, C., Gullberg, B., Ericson, U., Wirfält, E., & Orho-Melander, M. (2009). Fat and carbohydrate intake modify the association between genetic variation in the FTO genotype and obesity. *The American Journal of Clinical Nutrition*, *90*(5), 1418–1425. https://doi.org/10.3945/ajcn.2009.27958

Speliotes, E. K., Willer, C. J., Berndt, S. I., Monda, K. L., Thorleifsson, G., Jackson, A. U., Allen, H. L., Lindgren, C. M., Luan, J., Mägi, R., Randall, J. C., Vedantam, S., Winkler, T. W., Qi, L., Workalemahu, T., Heid, I. M., Steinthorsdottir, V., Stringham, H. M., Weedon, M. N., … Loos, R. J. F. (2010). Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature Genetics*, *42*(11), 937–948. https://doi.org/10.1038/ng.686

Swift, D. L., McGee, J. E., Earnest, C. P., Carlisle, E., Nygard, M., & Johannsen, N. M. (2018). The Effects of Exercise and Physical Activity on Weight Loss and Maintenance. *Progress in Cardiovascular Diseases*, *61*(2), 206–213. https://doi.org/10.1016/j.pcad.2018.07.014

Tchang, B. G., Aras, M., Kumar, R. B., & Aronne, L. J. (2000). Pharmacologic Treatment of Overweight and Obesity in Adults. In K. R. Feingold, S. F. Ahmed, B. Anawalt, M. R. Blackman, A. Boyce, G. Chrousos, E. Corpas, W. W. de Herder, K. Dhatariya, K. Dungan, J. Hofland, S. Kalra, G. Kaltsas, N. Kapoor, C. Koch, P. Kopp, M. Korbonits, C. S. Kovacs, W. Kuohung, … D. P. Wilson (Eds.), *Endotext*. MDText.com, Inc. http://www.ncbi.nlm.nih.gov/books/NBK279038/

Tyrrell, J., Wood, A. R., Ames, R. M., Yaghootkar, H., Beaumont, R. N., Jones, S. E., Tuke, M. A., Ruth, K. S., Freathy, R. M., Davey Smith, G., Joost, S., Guessous, I., Murray, A., Strachan, D. P., Kutalik, Z., Weedon, M. N., & Frayling, T. M. (2017a). Gene–obesogenic environment interactions in the UK Biobank study. *International Journal of Epidemiology*, *46*(2), 559–575. https://doi.org/10.1093/ije/dyw337

Wadden, T. A., Berkowitz, R. I., Womble, L. G., Sarwer, D. B., Phelan, S., Cato, R. K., Hesson, L. A., Osei, S. Y., Kaplan, R., & Stunkard, A. J. (2005). Randomized Trial of Lifestyle Modification and Pharmacotherapy for Obesity. *New England Journal of Medicine*, *353*(20), 2111–2120. https://doi.org/10.1056/NEJMoa050156

Wadden, T. A., Tronieri, J. S., & Butryn, M. L. (2020). Lifestyle Modification Approaches for the Treatment of Obesity in Adults. *The American Psychologist*, *75*(2), 235–251. https://doi.org/10.1037/amp0000517

Wang, Y., Rodríguez De Gil, P., Chen, Y.-H., Kromrey, J. D., Kim, E. S., Pham, T., Nguyen, D., & Romano, J. L. (2017). Comparing the Performance of Approaches for Testing the Homogeneity of Variance Assumption in One-Factor ANOVA Models. *Educational and Psychological Measurement*, *77*(2), 305–329. https://doi.org/10.1177/0013164416645162

Worldwide trends in underweight and obesity from 1990 to 2022: A pooled analysis of 3663 population-representative studies with 222 million children, adolescents, and adults. (2024). *Lancet (London, England)*, *403*(10431), 1027–1050. https://doi.org/10.1016/S0140-6736(23)02750-2

Yengo, L., Sidorenko, J., Kemper, K. E., Zheng, Z., Wood, A. R., Weedon, M. N., Frayling, T. M., Hirschhorn, J., Yang, J., Visscher, P. M., & GIANT Consortium. (2018). Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Human Molecular Genetics*, *27*(20), 3641–3649. https://doi.org/10.1093/hmg/ddy271

Zhang, X., & Bell, J. T. (2024). Detecting genetic effects on phenotype variability to capture gene-by-environment interactions: A systematic method comparison. *G3 Genes|Genomes|Genetics*, *14*(4), jkae022. https://doi.org/10.1093/g3journal/jkae022

APPENDIX

```r
#BF test validation and implementation
#dataframe contains the data frame for analysis
library(car)
dataframe$PEgroup = 0
dataframe$PEgroup[dataframe$PE != 0] = 1
bmi_test = rnorm(length(dataframe$BMI))
df <- data.frame( values = c(bmi_test, dataframe$PE),
                  group = factor(c(rep("PE group"), length(bmi_test)), rep("PE group"), length(dataframe$PE)))
#BF test implementation
n_boot = 1000;
p_values <-c();
f_val <- c();
f_val[1] = 0; f_val[n_boot] = 0;
p_values[1] = 0; p_values[n_boot] = 0;

df = dataframe
for (i in 1:n_boot) {
  df$BMI = sample(dataframe$BMI, size = length(dataframe$BMI))
  df$PEgroup = sample(dataframe$PEgroup, size = length(dataframe$BMI))
  test_result <- leveneTest(df$BMI ~ as.factor(PEgroup), data = df, center = "median")
  p_values[i] <- test_result[1,"Pr(>F)"]
  f_val[i] <- test_result[1,"F value"]}
#hist(f_val)
hist(p_values, col = "orange", main = "Null distribution of the BF test statistic", x_lab = NULL)
compare = runif(length(p_values))
ks.test(p_values, compare)
```

Figure 1. BF test implementation and validation code in R

```r
pos_data[pos_data$CHR == 1,]
MC4R_position <- subset(pos_data, SNP == "rs6567160")
MC4R_data = data_genotype[, "V615"]
age_gap1 = 18
age_gap2 = 30
dataframe = data.frame(SNP = MC4R_data, gender = data$gender, WHR = data$whr,
                       PE = data$physicalExercisePerWeek, age = data$age, BMI = data$bmi)
dataframe = dataframe[dataframe$PE >= 0 & dataframe$PE <=5,]
dataframe = dataframe[dataframe$gender == "Male",]
dataframe = na.omit(dataframe)
dataframe = dataframe[dataframe$age >= age_gap1 & dataframe$age <= age_gap2,]
df_avg <- dataframe %>%
  group_by(SNP, PE) %>%
  summarize (mean_WHR = mean(WHR, na.rm = TRUE), se_sum = sd(WHR, na.rm = TRUE), meansd_l = mean_WHR - (se_sum/sqrt(length(WHR))),
             meansd_u = mean_WHR + (se_sum/sqrt(length(WHR))))
plt <- ggplot(df_avg, aes(x = PE, y = mean_WHR, color = factor(SNP), group = SNP)) + geom_point(size = 3) +
  labs (title = "Average WHR vs Physical activity per week 18-30 age group", x = "Physical activty, hours per week", y = "average WHR")
+ scale_color_manual (values = c("blue", "grey", "red"))+
  geom_smooth(method = "lm", fill = NA) +
  theme_minimal() + theme(legend.title = element_blank())
sE_add <- plt + geom_errorbar(aes(ymin = meansd_l, ymax = meansd_u), width = 0.1)
sE_add
model_inter <- lm(WHR~PE*(as.factor(SNP)), data = dataframe)
model_inter$coefficients
summary(model_inter)
```

Figure 2. R code for plotting the GxE of the desired gene and environmental variable (for the separate age group and gender)

NON-EXCLUSIVE LICENCE TO REPRODUCE THESIS AND MAKE THESIS PUBLIC

Non-exclusive licence to reproduce the thesis and make the thesis public

I, Stepan Strelchenko ,
(author´s name)

1. grant the University of Tartu a free permit (non-exclusive licence) to

   reproduce, for the purpose of preservation, including for adding to the digital archives of the University of Tartu until the expiry of the term of copyright, my thesis

   Genetic and environment interactions in human obesity ,
   (title of thesis)

   supervised by Jon Anders Eriksson ;
   (supervisor´s name)

2. grant the University of Tartu a permit to make the thesis specified in point 1 available to the public via the web environment of the University of Tartu, including via the digital archives, under the Creative Commons licence CC BY NC ND 4.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright;

3. am aware of the fact that the author retains the rights specified in points 1 and 2;

4. confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Stepan Strelchenko
20/05/2025