UNRAVELLING PUTATIVE ETHNIC-SPECIFIC OBESITY GENETIC LOCI IN MALAYSIAN YOUTHS: A GENOME-WIDE ASSOCIATION STUDY

Muhammad Bani Yamin[1], Fazleen Haslinda Mohd Hatta[2], Mohd Shafiq Aazmi[3], Saiful Effendi Syafruddin[4], Michelle Yee Shuen Lee[5], Nurul Hanis Ramzi[6], Nikman Adli Nor Hashim[1,7] and Nazia Abdul Majid[1,7]

[1]Institute of Biological Sciences, Faculty of Science, Universiti Malaya, Kuala Lumpur, Malaysia; [2]Faculty of Pharmacy, Universiti Teknologi MARA Cawangan Selangor, Kampus Puncak Alam; [3]Microbiome Health and Environment (MiHeaRT) Research Interest Group, Faculty of Applied Sciences, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia; [4]UKM Medical Molecular Biology Institute (UMBI), Jalan Yaacob Latif, Bandar Tun Razak, Cheras; [5]School of Postgraduate Studies, International Medical University; [6]Institute for Research Development and Innovation, International Medical University; [7]Centre for Drug Research in Systems Biology, Structural Bioinformatics and Human Digital Imaging (CRYSTAL), Universiti Malaya, Kuala Lumpur, Malaysia

Abstract. Obesity is a major global health problem and increases the risk of various chronic diseases, such as cancer and type 2 diabetes. Understanding the genetic factors contributing to obesity is essential for developing effective intervention strategies. Thus, the study aimed to identify obesity-associated genetic loci in the Malaysian youth population. We conducted a genome-wide association analysis of 203 Malaysian youths (18-30 years of age) to identify susceptibility loci associated with body mass index-based obesity using a genotyped dataset imputed from a list of Asian-specific, obesity-associated loci, and the 1000 Genomes data. Statistical analysis, including linear regression and case-control association, was conducted to determine the association between single nucleotide polymorphisms (SNPs) and obesity. While no single SNP achieved genome-wide significance ($p = 5\times10^{-8}$), 34 SNPs exhibited suggestive significance ($p\leq5\times10^{-5}$). Of note, three chromosomal regions stood out (3q29, 7p11.2 and 17p13.1), harboring multiple suggestive significant associated SNPs. Although we were not able to identify the precise mechanisms and/or metabolic pathways underlying obesity in this particular target population, the study advances our knowledge of the genetic aspects of obesity in a Malaysian multi-ethnic population, which provides a baseline dataset for future studies.

Keywords: body mass index, genome-wide association study, obesity, single nucleotide polymorphism, youth

Correspondence: Nazia Abdul Majid, Microbiology and Molecular Genetics, Institute of Biological Sciences, Faculty of Science, Universiti Malaya, Lengkungan Budi, Kuala Lumpur 50603, Malaysia
Tel : +60 3 79675833, +60 17 2121934 E-mail : nazia@um.edu.my
and
Nikman Adli Nor Hashim, Centre for Drug Research in Systems Biology, Structural Bioinformatics and Human Digital Imaging (CRYSTAL), Universiti Malaya, Lengkungan Budi, Kuala Lumpur 50603, Malaysia
Tel: +60 3 79677189 E-mail: nikmanadli@um.edu.my

INTRODUCTION

Obesity is a chronic disease caused by excessive fat deposition (WHO, 2024). This condition primarily results from an imbalance between energy intake and energy expenditure. Obesity has become a global epidemic in the past decades (WHO, 2024) and imposes a heavy burden, such as increased healthcare costs and reduced productivity, on countries worldwide (Purnell, 2023).

Obesity is a multifactorial health problem influenced by both lifestyle and environmental factors. An obesogenic environment, which consists of calorie-rich foods, low levels of physical activity and stressful lifestyles, is a major contributor to the rapid increase in obesity. However, not everyone who is exposed to the same unhealthy environment becomes obese, suggesting the role of genetics in obesity. Twin studies are often used to investigate the relative contributions of genetic and environmental factors to various phenotypes. Many twin studies have shown that body weight has a high degree of heritability, with estimates ranging from 40 to 80% for body mass index (BMI) (Silventoinen *et al*, 2017; Silventoinen and Konttinen, 2020). These findings suggest that genetics plays a significant role in modulating the risk of obesity.

In Malaysia, obesity has become a serious health issue as obesity prevalence has risen rapidly from 11.9% in 2015 to 21.8% in 2023 (Institute for Public Health, 2024). A study in Malaysia found that 21.2% of university students were overweight and 16.3% were obese (Wan Mohamed Radzi *et al*, 2019). This reflects the increasing trend of obesity among youths and adults in Malaysia. The same study also reported that the prevalence of obesity is highest among PhD students (18.4%) followed by the Bachelor's (17.6%) and then the Master's degree students (14.3%) (Wan Mohamed Radzi *et al*, 2019). However, these comparisons may not be accurate, as postgraduate students may have a wider range of ages than Bachelor's degree students and consist of a sizeable proportion of international students. Therefore, the data for Bachelor's degree students serve as a better indicator of obesity severity among Malaysia's youths.

Obesity research in Malaysia is still in its infancy, as a scoping review reported that only 22 out of 188 publications on obesity in 2018 are related to obesity genes and biomarkers (Mohamad Nor *et al*, 2018). Most of these studies focus on the melanocortin-4 receptor (MC4R), a key regulator of leptin-mediated appetite control (Mohamad Nor *et al*, 2018). However, the association between *MC4R* variants and obesity in Malaysia is unclear, as different studies reported conflicting results (Chua *et al*, 2012; Apalasamy *et al*, 2013). Several studies have investigated other genes involved in satiety regulation and pathways, such as leptin and its receptor, an important hormone that signals satiation (Peikin, 1989; Austin and Marks, 2009; Yeung and Tadi, 2023). Prevalence and associations of polymorphisms in these genes with obesity were studied in a multi-ethnic Malaysian suburban population, but the results are inconsistent and dependent on ethnicity (Fan and Say, 2014). Other than *MC4R* and leptin-related genes, those implicated in obesity-related pathways, such as fat mass and obesity-associated (FTO) gene, have also been examined for their single nucleotide polymorphism (SNP) associations with obesity,

but the findings are inconclusive (Apalasamy *et al*, 2015; Lim, 2016; Lim *et al*, 2020; Harun *et al*, 2023).

Given these inconclusive findings, we opted for a robust, genome-wide association (GWAS) approach tailored to the population of interest. Illumina Infinium™ Asian Screening Array (ASA) was selected for its suitability and practicality among East and Southeast Asian populations (Ren *et al*, 2024). The ASA array includes 743,722 gene variants, comprising 56.7% common, 30.8% low-frequency and 12.5% rare variants. In addition, ASA features a broad spectrum of pharmacogenomic markers from the ClinVar database, gnomAD database and unpublished results from various Asian consortia (Liu *et al*, 2021). Several studies using ASA reported new loci associated with a number of diseases in East and Southeast Asian populations (Liu *et al*, 2021; Namkoong *et al*, 2022; Sani *et al*, 2022). Thus, our GWAS aims to identify SNPs associated with obesity-related traits (BMI and obesity status) in the Malaysian youth population. We hypothesized that this population may have ethnic-specific obesity loci/SNPs that contribute to the obesity epidemic. We report several SNPs associated with obesity (BMI and obesity status) and its potential mechanisms contributing to obesity progression. It is hoped that these findings could contribute to current body of knowledge and improve our understanding of obesity etiology, which could enhance prevention efforts and treatment strategies in Southeast Asian population, specifically in Malaysia.

MATERIALS AND METHODS

Malaysian youth population

According to Daniel's sample size formula (Daniel, 1999; Pourhoseingholi *et al*, 2013), we should recruit at least 205 youths. Participants recruited between January and September 2022 consisted of Malaysian undergraduate university students (Universiti Teknologi MARA and Universiti Malaya) in Klang Valley, Kuala Lumpur and Selangor, comprising Chinese (n = 13), Indian (n = 15), Malay (n = 235), and other (n = 14) ethnic groups. Inclusion

criteria were: (i) between 18-30 years of age, (ii) no previous exposure to drug/disease/treatment altering weight gain/maintenance, and (iii) no restriction regarding sex and ethnicity. Individuals with body mass index (BMI) <25 kg/m^2 are defined as non-obese and those with BMI ≥25 kg/m^2 as obese, based on the WHO BMI classification for the Asian population (WHO WPRO, 2000).

Phenotypic and anthropometric data collection

A questionnaire was used to obtain anthropometric data and information affecting weight, such as daily routine, physical activity, diet, and medical history. Sex was based on assignment at birth (male/female). No changes in gender identities or socially constructed roles were observed or taken into consideration. Physical activities were self-reported comprising three categories, namely low, moderate or high. Low physical activity level mainly consists of sedentary tasks, such as using motor vehicles for transport, taking the elevator instead of stairs, and occasional sports. Moderate level of physical activity mainly consists of walking or cycling to work, taking the stairs instead of the elevator, and engaging in regular recreational exercise (at least 30 minutes a day). High level of physical activity consists of regular physical activity in daily life (*eg*, walking, cycling, climbing stairs), endurance training (3 times weekly - at least 20-60 minutes) and strength training (two times weekly). Participant's medical history (not related to treatment for weight problems) was recorded regarding medicine taken during the past year as well as the family medical history.

GWAS protocol

DNA was isolated from a 3 ml aliquot of blood samples using MasterPure DNA Purification Kit for Blood Version II (Lucigen, Middleton, WI) according to the manufacturer's protocol. DNA sample (50 ng; $A_{260\,nm}$: $A_{280\,nm}$ = 1.8-2.0 was genotyped using Infinium™ ASA (Illumina, San Diego, CA) according to the manufacturer's protocol.

Prior to imputation and SNP

analysis, standard quality checks (QCs) consisting of sex discordance, missing rate, heterogeneity rate, and call rate were conducted using PLINK (v1.9) (Chang *et al*, 2015). Population stratification including inbreeding was carried out to remove closely related samples. As Malaysia is a multiethnic country, with a population from different ancestries, multidimensional scaling (MDS) was performed to compare our dataset with other populations in the 1000G August 2010 release dataset (The 1000 Genomes Project Consortium, 2015). Rare variants were called using the zCall tool (Goldstein *et al*, 2012).

In addition, we performed an additional imputation on the post-QC dataset, with a comprehensive list of Asian-specific, obesity-associated loci collected from previous reports. These SNP genotyping data were imputed with IMPUTE2 (Howie *et al*, 2009) using the 1000 Genomes Project Phase 3 build hg19 reference panel (released in October 2014) (The 1000 Genomes Project Consortium, 2015). The imputed dataset was filtered to exclude SNPs with an imputation quality score <0.8.

Statistical analysis

The imputed, cleaned dataset was subjected to several statistical tests using the PLINK (v1.9) tool (Chang *et al*, 2015). In brief, case-control (1 degree of freedom, chi-square allelic test) (obesity versus non-obesity), binomial logistic regression (obesity status as the dependent variable), and linear regression (BMI as the dependent variable) tests were conducted accordingly. Due to the diversity of the Malaysian population, binomial logistic the dataset was categorized into two groups: all ethnicity (n = 277) and Malay-only (n = 235), as Malays are the largest proportion of the multi-ethnic population in Peninsular Malaysia (Nagaraj *et al*, 2015). Each statistical test was conducted with and without physical activity, sex and ethnicity as covariates. Manhattan and Quantile-Quantile (Q-Q) plots were generated accordingly using the qqman package (Turner, 2018) in R (v4.2.2) (R Foundation, Vienna,

Austria). Consequently, genomic control (GC) model adjustment was applied to the dataset via PLINK (v1.9) (Chang *et al*, 2015; Simon *et al*, 2015). SNPs that reached $p \leq 5 \times 10^{-5}$ threshold when tested with BMI in case-control association and linear regression tests, were validated again using SKAT (v2.2.5) (Wu *et al*, 2011). Annotation of the SNPs was performed using ANNOVAR (Wang *et al*, 2010).

Ethical approval

The study protocol was approved by the UiTM Research Ethics Committee [REC/05/2022 (ST/MR/98)]. Prior written informed consent was obtained from each participant.

RESULTS

Malaysian youth cohort phenotype and anthropometric data

From the original group of recruited Malaysian youths (n = 277), DNA samples of individuals (n = 240) were genotyped. Despite this, only a smaller cohort (n = 203) was eligible for inclusion in the GWAS of obesity-related traits. Excluded individuals consisted of those (n = 34) with a call rate <95%, mean ± 4 SD of heterogeneity rate or gender of self-reported information discordant with birth sex, and those (n = 40) with missing anthropometric data, poor DNA quality and concentration. For population stratification, plots of the first and second MDS components of the samples in our dataset displayed a clustering similar to the Asian population from the 1000G dataset (The 1000 Genomes Project Consortium, 2015). The eligible participants were composed of Malays (n = 169), Chinese (n = 12), Indians (n = 10) and other ethnic groups (n = 12); females (n = 149) and males (n = 54); non-obese (n = 163) and obese (n = 40) groups; low physical activity (n = 35), moderate physical activity (n = 114), and high physical activity (n = 54) groups (Table 1).

GWAS data

The imputed genotype dataset contained SNPs from 19,898,421 autosomal and 20,970 X chromosomes for 203 individuals. All imputed SNPs that were

Table 1

Demographic and anthropometric data of study participants

Demographic characteristic	Frequency, *n* (%)	
	Obese (*n* = 40)	Non-obese (*n* = 163)
Sex		
Male	20 (50)	34 (21)
Female	20 (50)	129 (79)
Race		
Malay	33 (82)	136 (83)
Chinese	0 (0)	12 (7)
Indian	5 (13)	5 (4)
Other races	2 (5)	10 (6)
Physical activity		
Low	13 (32)	22 (13)
Moderate	20 (50)	94 (58)
High	7 (18)	47 (29)

Note: Physical activity was categorized as follows. Low: mainly seated activities, using motor vehicles for transport, using elevators instead of stairs, no sports or just occasionally; Medium: walking to work or taking the bicycle, using stairs instead of elevators, regular recreational exercise (at least 30 minutes a day); High: regular physical activity in daily life (walking, cycling, climbing stairs, and similar activities), endurance training (at least 20-60 minutes 3 times weekly), strength training (at least twice a week). These options were presented as multiple-choice questions in the questionnaire given during the blood sample collection process.

missing in at least 20% of the samples, were excluded from the analysis, resulting in SNPs from 19,401,323 autosomal and 20,970 X chromosomes SNPs (97.5%) for statistical analysis. Any SNP with a Hardy-Weinberg equilibrium of 1×10^{-5}, a minor allele frequency

(MAF) >5% and a genotyping call rate <99% were excluded. In total, 168 (0.001%) and 832 (0.004%) SNPs reached a threshold *p*-value $\leq 5\times10^{-5}$ when compared to BMI in case-control association and linear regression test respectively, which were validated using SKAT (v2.2.5) (Wu *et al*, 2011). Although none of the SNPs tested achieved a significant genome-wide association threshold value (*p*-value $\leq 5\times10^{-8}$), we discovered 34 SNPs that reached a suggestive significant threshold (*p*-value $\leq 5\times10^{-5}$). A flow diagram of the preprocessing and association analysis to arrive at these 34 SNPs is shown in Fig 1.

Quantitative traits

The obesity-related trait investigated in this study was BMI. The mean values of BMI adjusted for age and sex were compared using a generalized linear regression analysis to predict the relationship between each SNP and mean BMI. A quantile-quantile (Q-Q) linear regression plot of the expected versus observed SNP p-values was constructed of all individuals' BMI values in our dataset (Fig 2A), together with a Manhattan plot of the analysis for linear regression of all individuals' BMI values (Fig 2B). All SNPs with *p*-values $\leq 5\times10^{-5}$ were subsequently validated using single-variant tests under an additive genetic model via the SKAT tool (Wu *et al*, 2011) in R (R Foundation, Vienna, Austria).

Our quantitative trait analysis of BMI values revealed 18 suggestive associated SNPs spanning 14 different chromosome regions. Analysis of the nearby genes flanking each SNP locus revealed two regions, namely 7p11.2 and 3q29, which harbor more than one such SNP (Table 2), and were thus considered as top-hit regions for further analysis.

Interestingly, the SNPs in the 7p11.2 region are located near two genes, *SEC61G-DT* and *EGFR*, that are implicated in adiposity and obesity regulation (Craig *et al*, 2023; Cao *et al*, 2022). Region 3q29 harbors five SNPs that are located close to *Hes1* and *LINC02036*, the former is involved in adiposity regulation in porcine species (Lei *et al*, 2013). Of note, one of the five

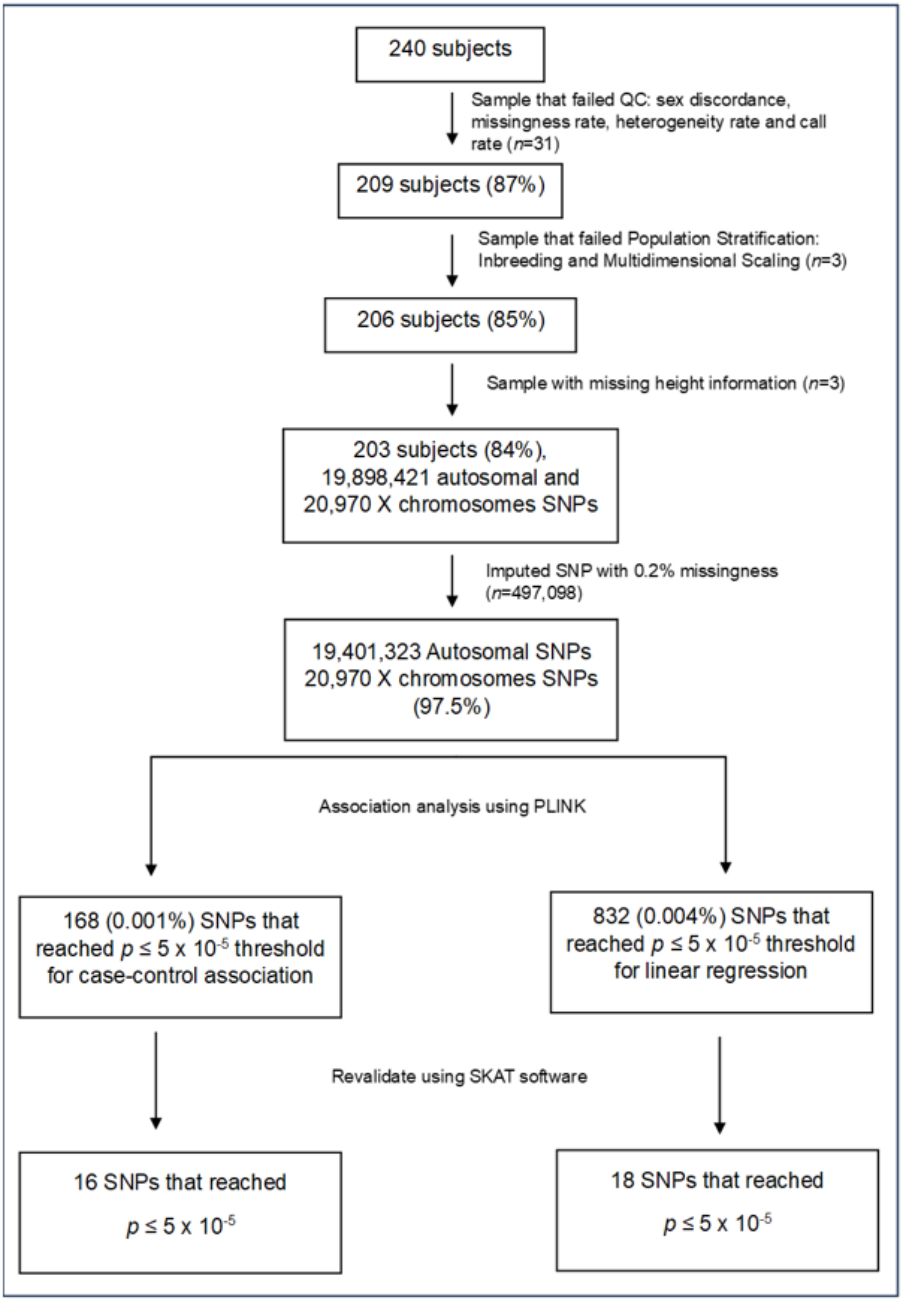

Fig 1 - Flow diagram of subjects and SNPs that underwent pre- and post-analysis to discover suggestive associated variants

QC: quality check; SNP: single nucleotide polymorphism

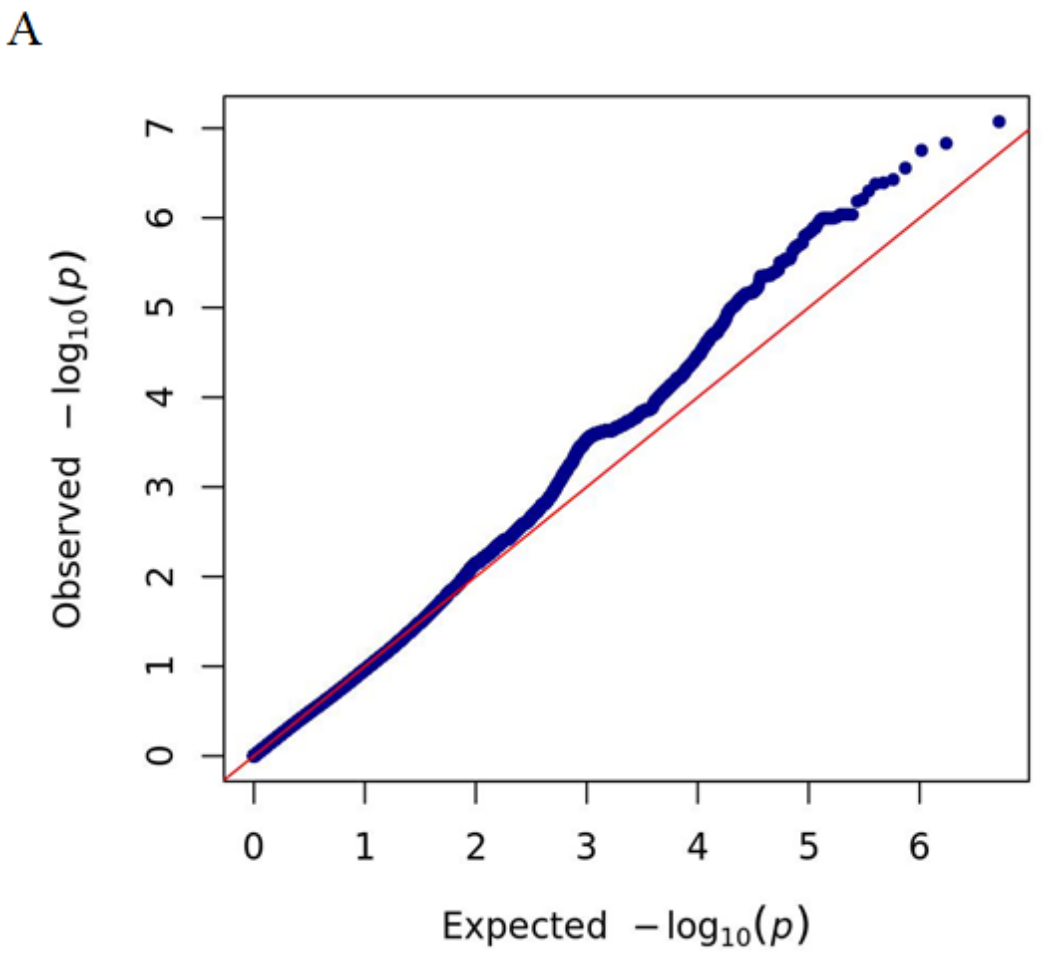

Fig 2 - Q-Q (A) and Manhattan (B) plots of body mass index quantitative trait study

A: Q-Q plot indicating expected and observed association values of linear regression model of BMI for all individuals. The results are in close agreement with the expected results, which minimized the effect of confounding population stratification. A GC model adjustment was applied.

BMI: body mass index; GC: genomic control; GWAS: genome-wide association study; -$\log_{10}(p)$: -$\log_{10}$ p-value; Q-Q plot: quantile-quantile plot; SNP: single nucleotide polymorphism

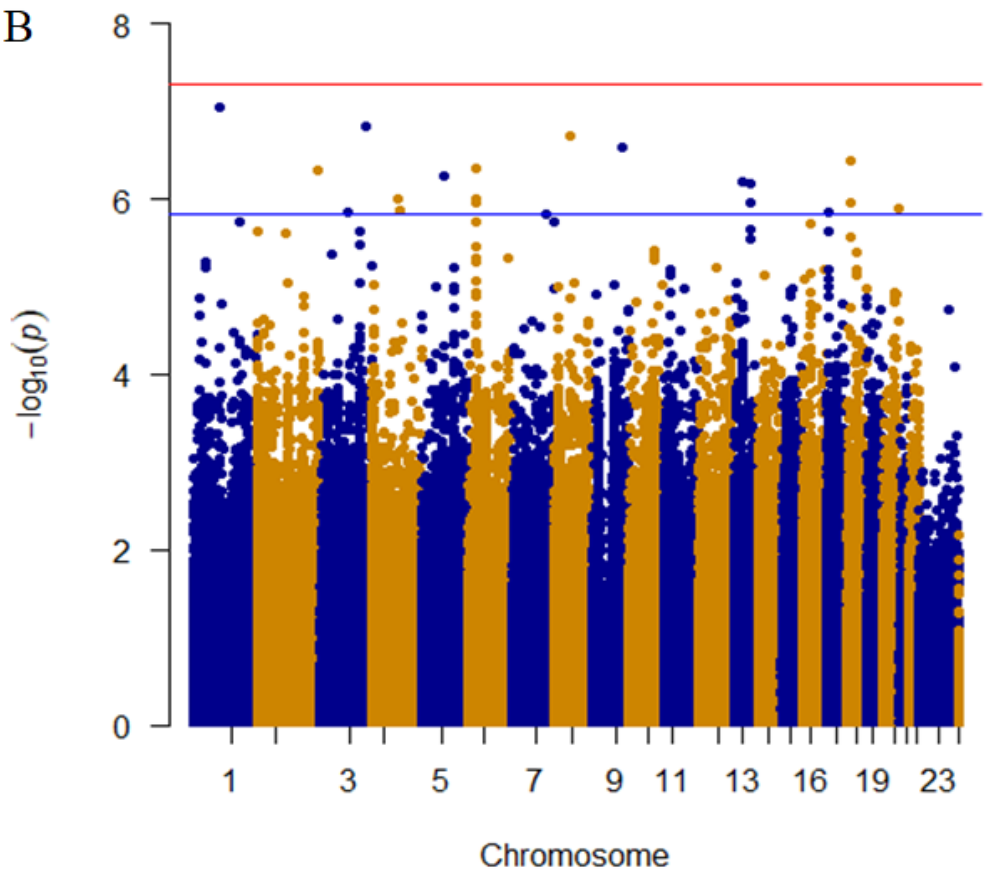

Fig 2 - Q-Q (A) and Manhattan (B) plots of body mass index quantitative trait study (cont)

B: Manhattan plot of the GWAS BMI quantitative trait of all individuals indicating regions of nominal significance. Each circle represents a SNP. The X-axis shows the chromosome position of each SNP. The Y-axis shows the $\log_{10}$ p-value of each SNP indicating the strength of association between the SNPs and BMI. The blue line indicates a suggestive significant threshold ($p = 5\times10^{-5}$). The red line indicates GWAS significant threshold ($p = 5\times10^{-8}$). The blue dots represent SNPs on chromosomes with odd numbers. The yellow dots represent SNPs on chromosomes with even numbers.

BMI: body mass index; GC: genomic control; GWAS: genome-wide association study; -$\log_{10}(p)$: -$\log_{10}$ p-value; Q-Q plot: quantile-quantile plot; SNP: single nucleotide polymorphism

Table 2

Single nucleotide polymorphisms (SNPs) from quantitative trait analysis of the study

SNP ID	Chromo-some	Type of association	*p*-value	β coefficient (95% CI)	Nearby gene (distance of SNP to gene)	Gene-associated trait	Allele frequency	Other allele frequency
rs28728404	3q29	Linear with covariates (Malay subgroup)	<0.001	0.396 (0.255-0.538)	*Hes1* (21.9 kbp), *LINC02036* (42.3 kbp)	*Hes1* - autism spectrum disorder (Anney *et al*, 2010), pulse wave analysis (Pilia *et al*, 2006)	C = 0.041	C = 0.435 (SGDP_PRJ), C = 0.928 (KRGDB)
		Linear	<0.001	0.318 (0.180-0.456)				
rs112690774	3q29	Linear with covariates (Malay subgroup)	<0.001	0.368 (0.231-0.504)	*Hes1* (21.2 kbp), *LINC02036* (43 kbp)	*Hes1* - autism spectrum disorder (Anney *et al*, 2010), pulse wave analysis (Pilia *et al*, 2006)	C = 0.039	G = 0.028 (Qatari), G = 0.046 (Vietnamese)
		Linear	<0.001	0.303 (0.169-0.438)				
rs112862622	3q29	Linear	<0.001	0.303 (0.170-0.436)	*Hes1*(18 kbp), *LINC02036* (45 kbp)	*Hes1* - autism spectrum disorder (Anney *et al*, 2010), pulse wave analysis (Pilia *et al*, 2006)	G = 0.042	G = 0.057 (Korea1K), G = 0.028 (Qatari)
rs112630848	3q29	Linear	<0.001	0.303 (0.170-0.436)	*Hes1* (19.1 kbp), *LINC02036* (45.1 kbp)	*Hes1* - autism spectrum disorder (Anney *et al*, 2010), pulse wave analysis (Pilia *et al*, 2006)	C = 0.042	C = 0.060 (KRGDB), C = 0.555 (SGDP_PRJ)

Table 2 (cont)

SNP ID	Chromo-some	Type of association	*p*-value	β coefficient (95% CI)	Nearby gene (distance of SNP to gene)	Gene-associated trait	Allele frequency	Other allele frequency
rs114506800	3q29	Linear	<0.001	0.275 (0.143-0.408)	*Hes1* (18.9 kbp), *LINC02036* (45.2 kbp)	*Hes1* - autism spectrum disorder (Anney *et al*, 2010), pulse wave analysis (Pilia *et al*, 2006)	A = 0.048	A = 0.033 (TOMMO), A = 0.040 (KRGDB)
rs79805190	1p21.2	Linear with covariates (Malay subgroup)	<0.001	0.371 (0.240-0.502)	*PLPPR4* (98 kbp), *LINC01708* (64.8 kbp)	*PLPPR4* - Alzheimer (Herold *et al*, 2016)	C = 0.032	C = 0.020 (TOPMED), C = 0.023 (Qatari)
		Linear (Malay subgroup)	<0.001	0.395 (0.256-0.535)				
		Linear	<0.001	0.364 (0.235-0.493)				
rs78644698	10p11.21	Linear with covariates (Malay subgroup)	<0.001	0.328 (0.194-0.462)	*PCAT5* (2.8 kbp), *ANKRD30A* (104.1 kbp)	*PCAT5* - body fat distribution* (Heid *et al*, 2010)	G = 0.019	G = 0.004 (TOMMO), G = 0.014 (Vietnamese)

Table 2 (cont)

SNP ID	Chromo-some	Type of association	*p*-value	β coefficient (95% CI)	Nearby gene (distance of SNP to gene)	Gene-associated trait	Allele frequency	Other allele frequency
rs77342483	8q12.1	Linear with covariates (Malay subgroup)	<0.001	0.338 (0.206-0.471)	*LOC100505501* (268.3 kbp), *CA8* (793.5 kbp)	*LOC100505501* - potassium (Province, 2011), *CA8* - BMI (Fox *et al*, 2007), schizophrenia (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014)	A = 0.024	A = 0.020 (Vietnamese), A = 0.051 (TOMMO)
		Linear	<0.001	0.354 (0.225-0.483)				
rs145763406	7p11.2	Linear with covariates (Malay subgroup)	<0.001	0.361 (0.225-0.497)	*SEC61G-DT* (118.4 kbp), *EGFR* (95.5 kbp)	*EGFR* - cholesterol*, HDL* (Province, 2011)	G = 0.049	G = 0.108 (TOMMO), G = 0.171 (Vietnamese)
rs118186417	7p11.2	Linear with covariates (Malay subgroup)	<0.001	0.365 (0.228-0.501)	*SEC61G-DT* (115.3 kbp), *EGFR* (98.6 kbp)	*EGFR* - cholesterol*, HDL* (Province, 2011)	G = 0.039	G = 0.108 (TOMMO), G =0.185 (Vietnamese)
rs118136016	7q31.33	Linear (Malay subgroup)	<0.001	0.385 (0.256-0.535)	*LOC101928283* (146 kbp), *GRM8* (913.2 kbp)	*GRM8* - body weight, BMI* (Fox *et al*, 2007)	T = 0.070	T = 0.052 (Vietnamese), T = 0.133 (TOMMO)

Table 2 (cont)

SNP ID	Chromo-some	Type of association	*p*-value	β coefficient (95% CI)	Nearby gene (distance of SNP to gene)	Gene-associated trait	Allele frequency	Other allele frequency
rs142743990	20q13.32	Linear (Malay subgroup)	<0.001	0.383 (0.243-0.524)	*PHACTR3* (0 bp)	*PHACTR3* - waist-hip ratio* (Heid *et al*, 2010)	A = 0.170	A = 0.034 (Vietnamese), A = 0.027 (TOMMO)
rs146077751	9q33.1	Linear (Malay subgroup)	<0.001	0.375 (0.234-0.516)	*ASTN2* (0 bp)	*ASTN2* - BMI (Fox *et al*, 2007), insulin response (Palmer *et al*, 2015)	C = 0.007	C = 0.009 (Vietnamese), C = 0.016 (TOMMO)
		Linear	<0.001	0.349 (0.220-0.479)				
rs148537777	2q37.1	Linear (Malay subgroup)	<0.001	0.388 (0.240-0.537)	*EIF4E2* (11.1 kbp), *EFHD1* (11.2 kbp)	*EIF4E2* - body height (Wood *et al*, 2014)	C = 0.011	C = 0.033 (Vietnamese), C = 0.059 (Korea1K)
		Linear	<0.001	0.364 (0.226-0.500)				
rs2052368	3q26.32	Linear	<0.001	0.357 (0.228-0.487)	*LINC02015* (126.6 kbp)	*LINC02015* -neuropsychological test* (Need *et al*, 2009)	G = 0.041	G = 0.023 (Vietnamese), G = 0.032 (Qatari)
rs76580608	18p11.21	Linear	<0.001	0.347 (0.217-0.467)	*FAM210A* (0 bp)	*FAM210A* - muscle mass (Chen *et al*, 2023)	A = 0.017	A = 0.001 (GnomAD), A = 0.001 (TOPMED)
rs74354207	13q31.1	Linear	<0.001	0.357 (0.221-0.492)	*LINC00564* (881.3 kbp), *SLITRK1* (1,768 kbp)	*SLITRK1* - schizophrenia (Legge *et al*, 2017)	C = 0.011	C = 0.014 (Vietnamese), C = 0.063 (KRGDB)

Table 2 (cont)

SNP ID	Chromosome	Type of association	*p*-value	β coefficient (95% CI)	Nearby gene (distance of SNP to gene)	Gene-associated trait	Allele frequency	Other allele frequency
rs2018417	4q23	Linear	<0.001	0.340 (0.211-0.471)	*ADH1B* (0 bp)	*ADH1B* - alcohol metabolism (Dasgupta, 2015), BMI (Polimanti and Gelernter, 2018), insulin resistance (Morales *et al*, 2021)	A = 0.192	A = 0.042 (Qatari), C = 0.500 (SGDP_PRJ)

Note: Type of association consists of Linear with covariates (Malay subgroup), Linear (Malay subgroup) and Linear. Linear with covariates (Malay subgroup): linear regression with sex and physical activity score covariates in Malay ethnic subgroup only; Linear (Malay subgroup): linear regression conducted in Malay ethnic subgroup only; Linear: linear regression conducted in all ethnic groups.

For gene associated traits, any trait found to associate significantly/suggestive significantly with the gene are presented.

Cohort consists of various cohort names. GnomAD: The Genome Aggregation Database; Korea1K: Korea 1K Human Genome Project; KRGDB: Korean Reference Genome Database; Qatari: whole genome sequencing of Qataris; SGDP_PRJ: Simons Genome Diversity Project; TOMMO: Tohoku Medical Megabank Project (also known as 14KJPN); TOPMED: Trans-Omics for Precision Medicine Program; Vietnamese: Vietnamese Genetic Variation Database

ADH1B: alcohol dehydrogenase 1B (class I), beta polypeptide; *ANKRD30A*: ankyrin repeat domain 30A; *ASTN2*: astrotactin 2; BMI: body mass index; bp: base pair; *CA8*: carbonic anhydrase 8; CI: confidence interval; *EGFR*: epidermal growth factor receptor; *EFHD1*: EF-hand domain family member D1; *EIF4E2*: eukaryotic translation initiation factor 4E family member 2; *FAM210A*: family with sequence similarity 210 member A; *GRM8*: glutamate metabotropic receptor 8; HDL: high density lipoprotein; *Hes1*: hairy and enhancer of split-1; kbp: kilobase pair; *LINC00564*: Long intergenic non-protein coding RNA 564; *LINC01708*: long intergenic non-protein coding RNA 1708; *LINC02015*: long intergenic non-protein coding RNA 2015; *LINC02036*: long intergenic non-protein coding RNA 2036; *LOC100505501*: uncharacterized *LOC100505501*; *LOC101928283*; uncharacterized *LOC101928283*; *PCAT5*: prostate cancer associated transcript 5; *PHACTR3*: phosphatase and actin regulator 3; *PLPPR4*: phospholipid phosphatase related 4; SEC61G-DT: SEC61 translocon subunit gamma divergent transcript; SLITRK1: SLIT and NTRK like family member 1; SNP: single nucleotide polymorphism; SNP ID: SNP identification number

*Suggestive significant association

SNPs in the 3q29 region, rs2872404, carries the highest association with BMI (adjusted β coefficient = 0.396) in the Malay subgroup compared to other SNPs in our dataset. Furthermore, our analysis showed that SNP rs28728404 was associated with a higher mean BMI of all individuals and Malay individuals (physical activity and sex as covariates) by 0.318 kg/m^2 (p-value <0.001, 95% CI: 0.180-0.456) and 0.396 kg/m^2 (p-value <0.001, 95% CI: 0.255-0.538) respectively (Table 2). The association of SNP rs28728404 also remained significant when the linear regression test was conducted in female individuals only (p-value <0.001, adjusted β coefficient = 0.637, 95% CI: 0.414-0.860).

SNP rs112690774 had the lowest p-value of 0.001 (adjusted β coefficient = 0.3676, 95% CI: 0.231-0.504) from the linear regression analysis of all individuals. In the Malay subgroup, rs112690774 is associated also with an increase in BMI value by 0.303 kg/m^2 (p-value <0.001, 95% CI: 0.231-0.504) (Table 2). Notably, SNP rs112690774 is present in the top-hit region, 3q29, which may lend support to the significance of this chromosome region.

Case-control studies

Our case-control analysis using chi-square allelic tests (1 degree of freedom) revealed 16 SNPs associated with obesity status, listed in Table 3 together with nearby genes flanking each locus. Another putative top-hit region, 17p13.1, which harbors 5 SNPs with a p-value <0.001 was identified. This region is associated with obesity-related traits only in the case-control approach (obesity vs non-obesity) and in the Malay-only (n =171) subgroup.

DISCUSSION

To the best of our knowledge, this GWAS of obesity is the first to be conducted in a cohort of the Malaysian youth population. Although no SNP reached the accepted GWAS significant association threshold ($p = 5\times10^{-8}$) (Fadista *et al*, 2016), 34 SNPs were identified with suggestive significance (p-value $\leq5\times10^{-5}$).

Table 3

Suggestive significant single nucleotide polymorphisms (SNPs) from case control analysis of the study

SNP ID	Region	Type of association	*p*-value	Odds ratio (95% CI)	Nearby gene (distance of SNP to gene)	Gene-associated trait	Allele frequency	Other allele frequency
rs114506800	3q29	Case control (Malay subgroup), Case control with covariates (Malay subgroup)	<0.001	15.890 (4.237-59.610)	*Hes1* (18.99 kbp), *CPN2* (184.98 kbp)	*Hes1* - autism spectrum disorder (Anney *et al*, 2010), pulse wave analysis (Pilia *et al*, 2006)	A = 0.039	A = 0.023 (Qatari), A = 0.040 (KRGDB)
rs57387722	4p15.1	Case control with covariates	<0.001	24.340 (3.484-37.490)	*PCDH7* (0 kbp)	*PCDH7* - BMI (Fox *et al*, 2007), cholesterol (Surakka *et al*, 2011)	C = 0.035	C = 0.049 (TOMMO), C = 0.028 (Vietnamese)
		Case control with covariates (Malay subgroup)	<0.001	12.280 (3.160-47.680)	*PCDH7* (0 kbp)	*PCDH7* - BMI (Fox *et al*, 2007), cholesterol (Surakka *et al*, 2011)	C = 0.033	C = 0.0487 (TOMMO), C = 0.052 (Qatari)
rs72696684	4p15.1	Case control (Malay subgroup), Case control with covariates (Malay subgroup)	<0.001	37.100 (4.552-302.400)	*HPF1* (0 kbp), *CLCN3* (15.78 kbp), *NEK1* (126.87 kbp)	*HPF1* - obesity** (Margulies and Ladurner, 2020), *CLCN3*, *NEK1* - cholesterol* (Heard-Costa *et al*, 2009)	C = 0.027	C = 0.048 (TOMMO), C = 0.065 (Qatari)

Table 3 (cont)

SNP ID	Region	Type of association	*p*-value	Odds ratio (95% CI)	Nearby gene (distance of SNP to gene)	Gene-associated trait	Allele frequency	Other allele frequency
rs13185666	5q12.1	Case control with covariates (Malay subgroup)	<0.001	4.321 (2.269-8.230)	*IPO11* (288 kbp), *DIMT1* (775 kbp)	*DIMT1* - insulin secretion (Verma *et al*, 2022)	C = 0.149	C = 0.047 (TOMMO), C = 0.111 (Vietnamese)
rs9501607	6p21.32	Case control with covariates	<0.001	6.137 (2.487-15.140)	*NOTCH4* (36.09 kbp)	*NOTCH4* - T2DM (Barrett *et al*, 2009)	T = 0.052	T = 0.033 (KRGDB), T = 0.052 (TOMMO)
rs9501398	6p21.32	Case control with covariates	<0.001	5.506 (2.285-13.260)	*NOTCH4* (10.75 kbp), TSBP-AS1 (20.35 kbp)	*NOTCH4* - T2DM (Barrett *et al*, 2009)	T = 0.0546	T = 0.052 (TOMMO) T = 0.042 (Vietnamese)
rs374231738	6p21.32	Case control with covariates	<0.001	5.541 (2.300-13.350)	*NOTCH4* (4.75 kbp), TSBP-AS1 (26.35 kbp)	*NOTCH4* - T2DM (Barrett *et al*, 2009)	G = 0.0545	G = 0.052 (TOMMO), G = 0.134 (Qatari)
rs1426713	6p21.32	Case control with covariates	<0.001	5.541 (2.300-13.350)	*NOTCH4* (6.20 kbp), TSBP-AS1 (24.9 kbp)	*NOTCH4* - T2DM (Barrett *et al*, 2009)	T = 0.0545	T = 0.033 (KRGDB), T = 0.052 (TOMMO)
rs4841405	8p23.1	Case control with covariates (Malay subgroup)	<0.001	11.880 (3.595-39.220)	*RP1L1* (0 kbp), PRSS55 (94.56 kbp)	*RP1L1* - obesity (Comuzzie *et al*, 2012)	A = 0.0417	A = 0.029 (Korea1K), A = 0.034 (TOMMO)

Table 3 (cont)

SNP ID	Region	Type of association	*p*-value	Odds ratio (95% CI)	Nearby gene (distance of SNP to gene)	Gene-associated trait	Allele frequency	Other allele frequency
rs189773	17p13.1	Case control, Case control with covariates	<0.001	4.660 (2.314-9.386)	*ASGR1* (25.23 kbp), *DLG4* (81.08 kbp)	*DLG4*** - obesity (Fadahunsi *et al*, 2024), *ASGR1*** - energy homeostasis (Svecla *et al*, 2024)	T = 0.091	T = 0.058 (KRGDB), T = 0.361 (Qatari)
rs220992	17p13.1	Case control with covariates	<0.001	4.672 (2.189-9.968)	*ASGR1* (26.96 kbp), *DLG4* (0 kbp)	*DLG4*** - obesity (Fadahunsi *et al*, 2024), *ASGR1*** - energy homeostasis (Svecla *et al*, 2024)	A = 0.081	A = 0.021 (TOMMO), A = 0.347 (Qatari)
rs4558460	17p13.1	Case control with covariates	<0.001	4.683 (2.145-10.230)	*ASGR1* (25.28 kbp) *DLG4* (0 kbp)	*DLG4*** - obesity (Fadahunsi *et al*, 2024), *ASGR1*** - energy homeostasis (Svecla *et al*, 2024)	T = 0.078	T = 0.036 (Korea1K), T = 0.023 (TOMMO)

Table 3 (cont)

SNP ID	Region	Type of association	*p*-value	Odds ratio (95% CI)	Nearby gene (distance of SNP to gene)	Gene-associated trait	Allele frequency	Other allele frequency
rs4436827	17p13.1	Case control with covariates	<0.001	4.683 (2.145-10.230)	*ASGR2* (33.37 kbp), *ASGR1* (25.23 kbp), *DLG4* (38.99 kbp)	*DLG4*** - obesity (Fadahunsi *et al*, 2024), *ASGR1*** - energy homeostasis (Svecla *et al*, 2024)	A = 0.078	A = 0.019 (TOMMO), A = 0.037 (KRGDB)
rs4148006	17p13.1	Case control with covariates	<0.001	5.088 (2.453-10.550)	*ABCA8* (0 kbp), *ABCA9* (91.21 kbp)	*ABCA8* - cholesterol (Willer *et al*, 2013)	A = 0.102	A = 0.108 (TOMMO) A = 0.088 (Vietnamese)
rs7258479	19p13.2	Case control with covariates	<0.001	8.800 (2.577-30.050)	*ARHGEF18* (0 kbp), *PEX11G* (25.66 kbp), *SAX05* (39.22 kbp)	*PEX11G*** - peroxisome elevation, fat loss (Liu and He, 2024)	A = 0.031	T = 0.036 (KRGDB), T = 0.019 (Vietnamese)

Table 3 (cont)

SNP ID	Region	Type of association	*p*-value	Odds ratio (95% CI)	Nearby gene (distance of SNP to gene)	Gene-associated trait	Allele frequency	Other allele frequency
rs75566654	19p13.2	Case control with covariates	<0.001	8.857 (2.594-30.240)	*ARHGEF18* (0 kbp), *PEX11G* (4.61 kbp), *SAX05* (25.28 kbp)	*PEX11G*** -peroxisome elevation, fat loss (Liu and He, 2024)	A = 0.030	A = 0.037 (TOMMO), A = 0.032 (Qatari)

Note: Type of association consists of Case control with covariates (Malay subgroup), Case control (Malay subgroup), Case control with covariates, and Case control. Case control with covariates (Malay subgroup): Case control with sex and physical activity score covariates in Malay ethnic subgroup only; Case control (Malay Subgroup): Case control study was conducted in Malay ethnic subgroup only; Case control with covariates: Case control with sex, ethnic and physical activity as covariates in all ethnic groups; Case control: Case control conducted in all ethnic groups.

For gene associated traits, any trait found to associate significantly/suggestive significantly with the gene is presented.

Cohort consists of various cohort names. Korea1K: Korea 1K Human Genome Project; KRGDB: Korean Reference Genome Database; TOMMO: Tohoku Medical Megabank Project (also known as 14KJPN); Qatari: whole genome sequencing of Qataris; Vietnamese: Vietnamese Genetic Variation Database

*Suggestive significant association; **No association test has been conducted to date

AB*CA8*: ATP binding cassette subfamily A member 8; *ABCA9*: ATP-binding cassette subfamily A member 8; *ARHGEF18*: Rho/Rac guanine nucleotide exchange factor 18; *ASGR1*: asialoglycoprotein receptor 1; *ASGR2*: asialoglycoprotein receptor 1; CI: confidence interval; *CLCN3*: chloride voltage-gated channel 3; *CPN2*: carboxypeptidase N subunit 2; *DIMT1*: DIM1 rRNA methyltransferase and ribosome maturation factor; *DLG4*: discs large MAGUK scaffold protein 4; *Hes1*: hairy and enhancer of split-1; *HPF1*: histone PARylation factor 1; *IPO11*: karyopherin subunit beta 1; *NEK1*: NIMA related kinase 1; *NOTCH4*: Notch receptor 4; *PCDH7*: protocadherin 7; *PEX11G*: peroxisomal biogenesis factor 11 gamma; PRSS55: serine protease 55; *RP1L1*: RP1 like 1; *SAX05*: stabilizer of axonemal microtubule; SNP ID: SNP identification number; TSBP-AS1: testis expressed basic protein 1-antisense RNA1

We further highlighted three chromosome regions, 3q29, 7p11.2 and 17p13.1, that harbor putative obesity-related genes, *eg*, *ASGR1*, *EGFR*, *DLG4*, and *Hes1*. Although our GWAS results could not indicate possible etiologies of obesity in the Malaysian youth population, the findings from our study have provided insights into how nearby genes flanking the SNP loci may have causal links to the development of obesity based on reports from the literature.

The 3q29 region contains SNP rs112690774 (the most significant SNP in our study) and SNP rs2872404 (with the highest association with the increase in BMI value (adjusted β coefficient = 0.396). Interestingly, the latter remains suggestive significant in females when stratified by sex. Furthermore, all suggestive significant SNPs identified in the 3q29 region in our study were located near *Hes1* and *LINC02036*.

Current understanding of the relationships among *Hes1*, Notch signaling pathway and obesity provides a plausible mechanism for the role of *Hes1* and obesity (Bi and Kuang, 2015). Notch is a plasma membrane receptor protein involved in a signaling pathway responsible for such cellular processes as proliferation and differentiation (Guruharsha *et al*, 2012). Its activation requires ligand binding, a two-stage proteolytic cleavage and a transfer of a transcriptional complex to the nucleus to activate the downstream *hairy/enhancer-of-split (Hes)* gene family (Bi and Kuang, 2015). Interestingly, a positive feedback loop that prolongs Notch signaling is via expression of a Notch target gene *Hes5* through repression of *FBW7* transcription (a mediator of the two-stage proteolytic cleavage) (Sancho *et al*, 2013). Therefore, we suggest that the *Hes* family plays a key role in the regulation of Notch signaling pathway.

In adipocyte thermogenesis, cold ambient temperature induces the sympathetic nervous system to release catecholamine, which binds with the β-adrenergic receptor to activate lipolysis (Bi and Kuang, 2015). For heat production, fatty acids directly activate the

uncoupling protein (UCP1) (Bi and Kuang, 2015). The *Hes1* gene, a target of Notch signaling, binds to the promoter region of *Pparγ*, *Ppargc1a*, and *Prdm16* genes, hindering their transcription (Bi and Kuang, 2015). Consequently, there is a reduction in mitochondria numbers and *UCP1* expression, promoting the production of proinflammatory cytokines through activation of NF-κB. Macrophages are consequently recruited causing low-grade systemic inflammation and worsening insulin resistance (Bi and Kuang, 2015). Thus, the relationship between *Hes1* and obesity development through the Notch signaling pathway is pertinent, further highlighting the importance of *Hes1* and 3q29 region in obesity. We have identified five suggestive significant SNPs located near *Hes1*: rs112630848, rs112690774, rs112862622, rs114506800, and rs28728404. Fig 3 summarizes the role of *Hes1* in obesity progression.

The 7p11.2 region contains several other genes known to be important in body fat and glycemic control, *eg*, *EGFR* and *SEC61G-DT* (Alamri *et al*, 2022). EGFR has been postulated to play important roles in adiposity and obesity regulation via the gut microbiota. EGFR activation, a key regulator of various cellular processes, influences the composition and metabolic activity of gut microbiota. Studies suggest a bidirectional relationship between EGFR activation and the composition of the gut microbiome (Buret *et al*, 1999; Dvorak, 2010; Knoop *et al*, 2020). This bidirectional relationship is pivotal in shaping the complex landscape of obesity development. EGFR signaling can influence microbial diversity, and conversely, the gut microbiota can modulate EGFR activation. EGFR activation induces changes in the gut microbiota composition by impacting intestinal epithelial cells and the mucosal environment. Alterations in EGFR signaling may lead to changes in expression of a variety of proteins (Tang *et al*, 2016) and in production of mucin (Clark *et al*, 2006; Damiano *et al*, 2015), thereby influencing the adhesion and growth of specific

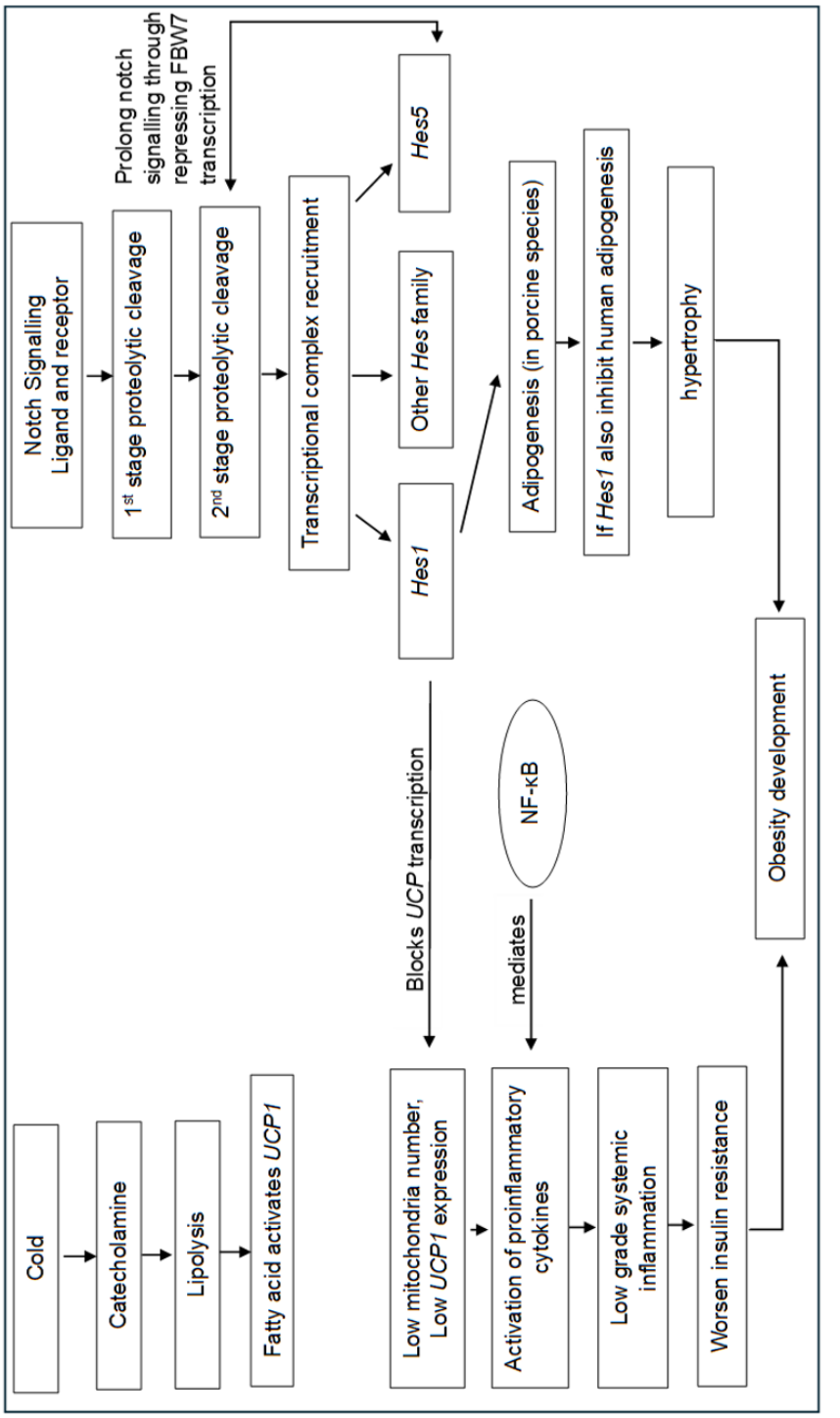

Fig 3 - Notch signaling pathway and *Hes1* role in obesity progression

The figure shows how a mutation at *Hes1* gene affects the *Hes* gene family expression, which plays a key role in prolonging Notch signaling and UCP transcription.

FBW7: F-box and WD repeat domain containing 7 gene; *Hes1:* hairy/enhancer split 1 gene; *Hes5*: hairy/enhancer split 5; NF-κB: nuclear factor kappa B; *UCP1*: uncoupling protein 1 gene

gut microbial species (Elliott *et al*, 2000; Wang *et al*, 2014). These changes in the microbiome have been associated with metabolic shifts thereby contributing to the development of obesity.

Interestingly, epidermal growth factors (EGFs) have been reported to show moderate positive correlations with Proteobacteria, a major phylum of Gram-negative bacteria, in women (Alamri *et al*, 2022). Furthermore, EGFR signaling pathways influence the metabolism of the gut microbiota as described above, impacting the production of metabolites such as short-chain fatty acids (SCFAs) (Morrison and Preston, 2016). SCFAs derived from commensal microbial fermentation of dietary fibers play a crucial role in energy regulation (Morrison and Preston, 2016). EGFR-mediated alterations in SCFA production can, therefore, have profound effects on host metabolism, contributing to the energy imbalance associated with obesity. These findings indicate that an understanding of the molecular intricacies of EGFR influence on gut microbiota and obesity is of paramount importance.

The association of 17p13.1 region, containing four SNPs bracketing *DLG4*, with obesity was only discovered through the case-control analysis. *DLG4* encodes PSD-95 protein of PSD-95 which is involved in glutamate receptor signaling of the brain. PSD-95 interacts with N-methyl-D-aspartate (NMDA) and α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptors, which are key regulators of neurotransmission and synaptic plasticity, and NMDA-ADPA interaction is also crucial in maintaining energy homeostasis (Fadahunsi *et al*, 2024). In addition, rodent studies showed that disrupting NMDA and AMPA receptors could mitigate food-motivated behavior and produce weight-lowering benefits (Fetterly *et al*, 2023; Fadahunsi *et al*, 2024). Sasaki *et al* (2016) also reported that NMDA and its co-agonist, D-serine, suppress food intake preference, supporting the notion of food intake regulation by *DLG4* (Fadahunsi *et al*, 2024). Another gene, *ASGR1*, located near *DLG4*, was postulated

to regulate food intake through the control of adipose tissue hormone distribution (Henry and Clarke, 2008). Specifically, adipose tissue regulates the levels of adiponectin, a hormone that regulates appetite, energy expenditure and insulin sensitivity level. Elevated adiponectin levels are associated with reduced appetite suggests that *ASGR1* plays a role in regulating food intake through control of adipose tissue hormone, such as adiponectin level. In addition, an *ASGR1* knockout mice model was reported to have an altered metabolic flexibility causing significant increases in visceral adipose and subcutaneous adipose tissues compared to wild-type mice (Svecla *et al*, 2024). However, the exact mechanisms and interplay among the factors involved remain elusive.

Other than the abovementioned SNPs, we also identified 21 other SNPs associated with obesity status and BMI. Previous studies support the relationship between obesity and genes located near these SNPs (Wang *et al*, 2015; Polimanti and Gelernter, 2018; Després and Lemieux, 2006; Topalidou *et al*, 2020; Cao *et al*, 2022; Chen *et al*, 2023). The postulated relationships among a selected group of genes located close to BMI-associated suggestive significant SNPs are shown in Fig 4. The reported SNPs are mainly connected with obesity development through Alzheimer, insulin resistance, and metabolism. Cumulative weight gain may lead to brain structure changes, potentially causing reverse causality where Alzheimer's contributes to obesity. Moreover, alteration of specific molecules metabolism and insulin resistance development were also highlighted.

Although we managed to identify 34 suggestive significant SNPs, albeit not reaching the GWAS significant threshold level of $p = 5\times10^{-8}$, this could be due to two reasons: i) low sample size, consisting only of 203 individuals; and ii) heterogeneity of samples, as participants were from various ethnic backgrounds, the majority being Malays and smaller numbers of Chinese, Indians and multi-

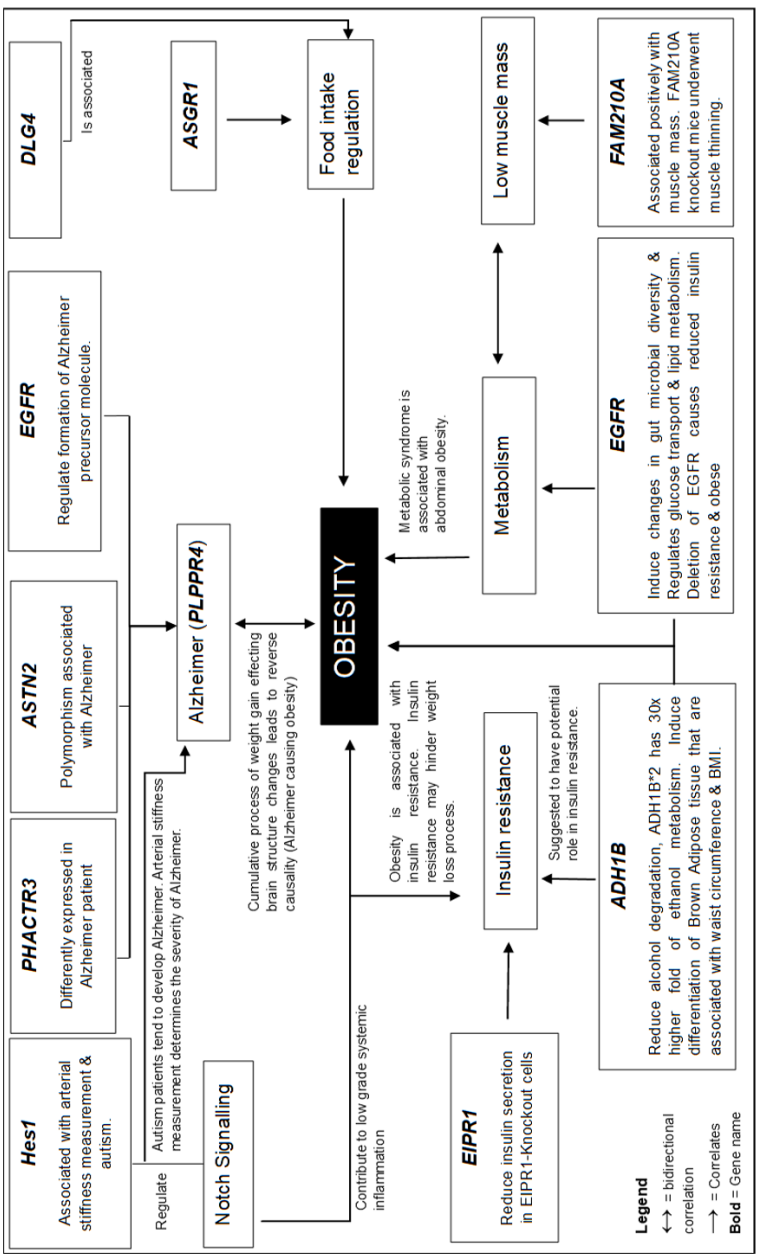

Fig 4 - Postulated relationships among genes in close proximity to BMI-associated suggestive significant SNPs

The diagram shows genes that are located near suggestive significant SNPs in our GWAS cohort. The postulated relationships depicted are based on reported results in the literature.

ADH1B: alcohol dehydrogenase 1B gene; *ASGR1*: asialoglycoprotein receptor 1 gene; *ASTN2*: astrotactin 2 gene; BMI: body mass index; *DLG4*: discs large MAGUK scaffold protein 4 gene; *EGFR*: epidermal growth factor receptor gene; *EIRP1*: EARP complex and GARP complex interacting protein 1 gene; *FAM210A*: family with sequence similarity 210 member A gene; GWAS: genome-wide association study; *Hes1*: hairy and enhancer split 1 gene; *PHACTR3*: phosphatase and actin regulator 3 gene; *PLPPR4*: phospholipid phosphatase related 4 gene; SNP: single nucleotide polymorphism

tribal East Malaysians. Studies reporting significant loci from GWAS normally require 5,000-100,000 individuals to obtain statistically robust association findings that fully represent a particular population (Liu *et al*, 2021; Namkoong *et al*, 2022; Verma *et al*, 2024). Despite these limitations, our study generated a dataset based on the most recent self-reported anthropometric dataset, containing information on general lifestyle, medical history and physical activity of a cohort of Malaysian youths compared to other studies that only include basic anthropometric data on obesity-related traits.

In conclusion, our study identified 34 SNPs that were associated with obesity in the Malaysian youth population. Among the 34 SNPs, there were three putative obesity regions (3q29, 7p11.2, and 17p13.1) that harbor several suggestive significant SNPs. However, the mechanisms for the development of obesity in this target population remain unclear. Nonetheless, it is hoped that these findings add to the body of knowledge and improve our current understanding of obesity etiology, particularly in multi-ethnic Southeast Asian populations such as Malaysia.

ACKNOWLEDGEMENTS

The authors thank Nuriannsyuhada and Nurul Izzah for their contributions in the collection of blood samples, all individuals who participated in the study, colleagues, families, and the whole team for their assistance in recruiting participants and sample collection. The work was funded by Fundamental Research Grant Scheme (FRGS), Ministry of Higher Education, Malaysia (grant no. FP040-2020).

CONFLICT OF INTEREST DISCLOSURE

All authors declare no conflict of interest.

DATA AVAILABILITY

Raw sequence data collected in the study are available at https://drive.google.com /drive/

folders/1gi1lXru0D. Access will be granted upon reasonable request.

REFERENCES

Alamri T, Alkhaldy AA, Gauthaman K, *et al*. Growth factors in relation to obesity, food habits, and microbiota among healthy Saudis: preliminary results. *Eur Rev Med Pharmacol Sci* 2022; 26(24): 9311-26.

Anney R, Klei L, Pinto D, *et al*. A genome-wide scan for common alleles affecting risk for autism. *Hum Mol Genet* 2010; 19(20): 4072-82.

Apalasamy YD, Ming MF, Rampal S, Bulgiba A, Mohamed Z. Association of melanocortin-4 receptor gene polymorphisms with obesity-related parameters in Malaysian Malays. *Ann Hum Biol* 2013; 40(1): 102-6.

Apalasamy YD, Rampal S, Salim A, *et al*. Polymorphisms of the resistin gene and their association with obesity and resistin levels in Malaysian Malays. *Biochem Genet* 2015; 53(4-6): 120-31.

Austin J, Marks D. Hormonal regulators of appetite. *Int J Pediatr Endocrinol* 2009; 2009: 141753.

Barrett JC, Clayton DG, Concannon P, *et al*. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet* 2009; 41(6): 703-7.

Bi P, Kuang S. Notch signaling as a novel regulator of metabolism. *Trends Endocrinol Metab* 2015; 26(5): 248-55.

Buret A, Gall DG, Olson ME, Hardin JA. The role of the epidermal growth factor receptor in microbial infections of the gastrointestinal tract. *Microbes Infect* 1999; 1(13): 1139-44.

Cao S, Pan Y, Tang J, *et al*. *EGFR*-mediated activation of adipose tissue macrophages promotes obesity and insulin resistance. *Nat Commun* 2022; 13(1): 4684.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015; 4: 7.

Chen J, Yue F, Kim KH, *et al*. FAM210A mediates an inter-organelle crosstalk essential for protein synthesis and muscle growth in mouse. *bioRxiv* [Preprint] 2023 Aug 5: 2023.08.03.551853.

Chua H, Fan S, Say Y. Prevalence of melanocortin receptor 4 (MC4R) V103I gene variant and its association with obesity among the Kampar Health Clinic cohort, Perak, Malaysia. *Med J Malaysia* 2012; 67(2): 234-5.

Clark JA, Doelle SM, Halpern MD, *et al*. Intestinal barrier failure during experimental necrotizing enterocolitis: protective effect of EGF treatment. *Am J Physiol Gastrointest Liver Physiol* 2006; 291(5): G938-49.

Comuzzie AG, Cole SA, Laston SL, *et al*. Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population. *PloS One* 2012; 7(12): e51954.

Craig SJ, Kenney AM, Lin J, *et al*. Constructing a polygenic risk score for childhood obesity using functional data analysis. *Econom Stat* 2023; 25: 66-86.

Damiano S, Morano A, Ucci V, *et al*. Dual oxidase 2 generated reactive oxygen species selectively mediate the induction of mucins by epidermal growth factor in enterocytes. *Int J Biochem Cell Biol* 2015; 60: 8-18.

Daniel WW. Biostatistics: a foundation for analysis in the health sciences. 7th ed. New York, NY: John Wiley & Sons 1999.

Dasgupta A. Genetic aspects of alcohol metabolism and drinking behavior. In: Dasgupta A, editor. Alcohol and its Biomarkers. Amsterdam, Netherlands. Elsevier; 2015. p. 37-63.

Després JP, Lemieux I. Abdominal obesity and metabolic syndrome. *Nature* 2006; 444(7121): 881-7.

Dvorak B. Milk epidermal growth factor and gut protection. *J Pediatr* 2010; 156(2 Suppl): S31-5.

Elliott S N, Wallace J, McKnight W, *et al*. Bacterial colonization and healing of gastric ulcers: the effects of epidermal growth factor. *Am J Physiol Gastrointest Liver Physiol* 2000; 278(1): G105-12.

Fadahunsi N, Petersen J, Metz S, *et al*. Targeting postsynaptic glutamate receptor scaffolding proteins PSD-95 and PICK1 for obesity treatment. *Sci Adv* 2024; 10(9): eadg2636.

Fadista J, Manning AK, Florez JC, Groop L. The (in) famous GWAS P-value threshold revisited and updated for low-frequency variants. *Eur J Hum Genet* 2016; 24(8): 1202-5.

Fan SH, Say YH. Leptin and leptin receptor gene polymorphisms and their association with plasma leptin levels and obesity in a multi-ethnic Malaysian suburban population. *J Physiol Anthropol* 2014; 33(1): 15.

Fetterly TL, Catalfio AM, Ferrario CR. Effects of junk-food on food-motivated behavior and NAc glutamate plasticity; insights into the mechanism of NAc calcium-permeable AMPA receptor recruitment. *bioRxiv* [Preprint] 2023: 2023.05.16.540977.

Fox CS, Heard-Costa N, Cupples LA, Dupuis J, Vasan RS, Atwood LD. Genome-wide association to body mass index and waist circumference: the Framingham Heart Study 100K project. *BMC Med Genet* 2007; 8 Suppl 1 (Suppl 1): S18.

Goldstein JI, Crenshaw A, Carey J, *et al*. zCall: a rare variant caller for array-based genotyping: genetics and population analysis. *Bioinformatics* 2012; 28(19): 2543-5.

Guruharsha K, Kankel MW, Artavanis-Tsakonas S. The Notch signalling system: recent insights into the complexity of a conserved pathway. *Nat Rev Genet* 2012; 13(9): 654-66.

Harun NS, Wan Jusoh AF, Rohin MAK, Yahaya R, Nik Him NAS, Zahary MN. The contribution of *FTO* rs9939609 and *RETN* rs1862513 polymorphisms in predisposing resettled indigenous (Orang Asli) Temiar to metabolic syndrome. *Egypt J Med Hum Genet* 2023; 24: 44.

Heard-Costa NL, Zillikens MC, Monda KL, *et al*. NRXN3 is a novel locus for waist circumference: a genome-wide association study from the CHARGE Consortium. *PLoS Genet* 2009; 5(6): e1000539.

Heid IM, Jackson AU, Randall JC, *et al*. Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat Genet* 2010; 42(11): 949-60.

Henry BA, Clarke IJ. Adipose tissue hormones and the regulation of food intake. *J Neuroendocrinol* 2008; 20(6): 842-9.

Herold C, Hooli BV, Mullin K, *et al*. Family-based association analyses of imputed genotypes reveal genome-wide significant association of Alzheimer's disease with *OSBPL6*, *PTPRG*, and *PDCL3*. *Mol Psychiatry* 2016; 21(11): 1608-12.

Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009; 5(6): e1000529.

Institute for Public Health. Fact Sheet, National Health and Morbidity Survey 2023: Non-Communicable Diseases & Healthcare Demand, 2024 [cited 2024 Jul 10]. Available from: URL: https://iku.gov.my/images/nhms2023/fact-sheet-nhms-2023.pdf

Knoop KA, Coughlin PE, Floyd AN, *et al*. Maternal activation of the *EGFR* prevents translocation of gut-residing pathogenic *Escherichia coli* in a model of late-onset neonatal sepsis. *Proc Natl Acad Sci USA* 2020; 117(14): 7941-9.

Legge SE, Hamshere ML, Ripke S, *et al*. Genome-wide common and rare variant analysis provides novel insights into clozapine-associated neutropenia. *Mol Psychiatry* 2017; 22(10): 1502-8.

Lei T, Bi Y, Gao MJ, *et al*. HES1 inhibits adipogenesis of porcine mesenchymal stem cells via transcriptional repression of FAD24. *Domest Anim Endocrinol* 2013; 45(1): 28-32.

Lim KG. A review of adult obesity research in Malaysia. *Med J Malaysia* 2016; 71(Suppl 1): 1-19.

Lim ZM, Chie QT, Teh LK. Influence of dopamine receptor gene on eating behaviour and obesity in Malaysia. *Meta Gene* 2020; 25: 100736.

Liu J, Li S, Li X, *et al*. Genome-wide association study followed by trans-ancestry meta-analysis identify 17 new risk loci for schizophrenia. *BMC Med* 2021; 19(1): 177.

Liu M, He L. Dietary cysteine and methionine promote peroxisome elevation and fat loss by induction of *CG33474* expression in Drosophila adipose tissue. *Cell Mol Life Sci* 2024; 81: 190.

Margulies CE, Ladurner AG. PARP-1 flips the epigenetic switch on obesity. *Mol Cell* 2020; 79(6): 874-5.

Mohamad Nor NS, Ambak R, Mohd Zaki N, *et al*. An update on obesity research pattern among adults in Malaysia: a scoping review. *BMC Womens Health* 2018; 18(Suppl 1): 114.

Morales LD, Cromack DT, Tripathy D, *et al*. Further evidence supporting

a potential role for ADH1B in obesity. *Sci Rep* 2021; 11(1): 1932.

Morrison DJ, Preston T. Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism. *Gut Microbes* 2016; 7(3): 189-200.

Nagaraj S, Tey NP, Ng CW, Lee KH, Pala J. Counting ethnicity in Malaysia: the complexity of measuring diversity. In: Simon P, Piche V, Gagnon AA, editors. Social statistics and ethnic diversity. Paris, France: Springer Open Press; 2015. p. 143-73.

Namkoong H, Edahiro R, Takano T, *et al*. DOCK2 is involved in the host genetics and biology of severe COVID-19. *Nature* 2022; 609(7928): 754-60.

Need AC, Attix DK, McEvoy JM, *et al*. A genome-wide study of common SNPs and CNVs in cognitive performance in the CANTAB. *Hum Mol Genet* 2009; 18(23): 4650-61.

Palmer ND, Goodarzi MO, Langefeld CD, *et al*. Genetic variants associated with quantitative glucose homeostasis traits translate to type 2 diabetes in Mexican Americans: the GUARDIAN (Genetics Underlying Diabetes in Hispanics) Consortium. *Diabetes* 2015; 64(5): 1853-66.

Peikin SR. Role of cholecystokinin in the control of food intake. *Gastroenterol Clin North Am* 1989; 18(4): 757-75.

Pilia G, Chen WM, Scuteri A, *et al*. Heritability of cardiovascular and personality traits in 6,148 Sardinians. *PLoS Genet* 2006; 2(8): e132.

Polimanti R, Gelernter J. ADH1B: from alcoholism, natural selection, and cancer to the human phenome. *Am J Med Genet B Neuropsychiatr Genet* 2018; 177(2): 113-25.

Pourhoseingholi MA, Vahedi M, Rahimzadeh M. Sample size calculation in medical studies. *Gastroenterol Hepatol Bed Bench* 2013; 6(1): 14-7.

Province M. NHLBI Family Heart Study (FamHS-Visit1 and FamHS-Visit2) dbGaP Study Accession: phs000221.v1.p1, 2011 [cited 2024 Aug 20]. Available from: URL: https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000221.v1.p1

Purnell JQ. Definitions, classification, and epidemiology of obesity, 2023 [cited 2024 May 08]. vailable from:

URL: https://www.ncbi.nlm.nih.gov/books/NBK279167/

Ren J, Peng C, Chen H, *et al*. Asian screening array and next-generation sequencing based panels applied to preimplantation genetic testing for monogenic disorders preclinical workup in 294 families: a retrospective analysis. *Prenat Diagn* 2024; 44(11): 1344-53.

Sancho R, Blake SM, Tendeng C, Clurman BE, Lewis J, Behrens A. Fbw7 repression by hes5 creates a feedback loop that modulates Notch-mediated intestinal and neural stem cell fate decisions. *PLoS Biol* 2013; 11(6): e1001586.

Sani H, Teh LK, Noorizhab F, Salleh MZ, Kasim S. Genetic analysis of a young adult presented with acute myocardial infarction with no traditional risk factors: a novel case report. *Mal J Med Health Sci* 2022; 18(1): 349-52.

Sasaki T, Matsui S, Kitamura T. Control of appetite and food preference by NMDA receptor and its co-agonist D-serine. *Int J Mol Sci* 2016; 17(7): 1081.

Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014; 511(7510): 421-7.

Silventoinen K, Jelenkovic A, Sund R, *et al*. Differences in genetic and environmental variation in adult BMI by sex, age, time period, and region: an individual-based pooled analysis of 40 twin cohorts. *Am J Clinl Nutr* 2017; 106(2): 457-66.

Silventoinen K, Konttinen H. Obesity and eating behavior from the perspective of twin and genetic research. *Neurosci Biobehav Rev* 2020; 109: 150-65.

Surakka I, Isaacs A, Karssen LC, *et al*. A genome-wide screen for interactions reveals a new locus on 4p15 modifying the effect of waist-to-hip ratio on total cholesterol. *PLoS Genet* 2011; 7(10): e1002333.

Svecla M, Da Dalt L, Moregola A, *et al*. *ASGR1* deficiency diverts lipids toward adipose tissue but results in liver damage during obesity. *Cardiovas Diabetol* 2024; 23(1): 42.

Tang X, Liu H, Yang S, Li Z, Zhong J, Fang R. Epidermal growth factor and intestinal barrier function. *Mediators Inflamm* 2016; 2016: 1927348.

The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* 2015; 526(7571): 68-74.

Topalidou I, Cattin-Ortolá J, Hummer B, Asensio CS, Ailion M. EIPR1 controls dense-core vesicle cargo retention and EARP complex localization in insulin-secreting cells. *Mol Biol Cell* 2020; 31(1), 59-79.

Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *J Open Source Software* 2018; 3(25): 731.

Verma A, Huffman JE, Rodriguez A, *et al*. Diversity and scale: genetic architecture of 2068 traits in the VA Million Veteran Program. *Science*. 2024; 385(6706): eadj1182.

Verma G, Bowen A, Gheibi S, *et al*. Ribosomal biogenesis regulator DIMT1 controls β-cell protein synthesis, mitochondrial function, and insulin secretion. *J Biol Chem* 2022; 298(3): 101692.

Wan Mohamed Radzi CWJ, Salarzadeh Jenatabadi H, Alanzi ARA, Mokhtar MI, Mamat MZ, Abdullah NA. Analysis of obesity among Malaysian university students: A combination study with the application of Bayesian structural equation modelling and Pearson correlation. *Int J Environ Res Public Health* 2019; 16(3): 492.

Wang D, Xu S, Lin Y, *et al*. Recombinant porcine epidermal growth factor-secreting Lactococcus lactis promotes the growth performance of early-weaned piglets. *BMC Vet Res* 2014; 10: 171.

Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010; 38(16): e164.

Wang KS, Tonarelli S, Luo X, *et al*. Polymorphisms within *ASTN2* gene are associated with age at onset of Alzheimer's disease. *J Neural Transm* 2015; 122(5): 701-8.

Willer CJ, Schmidt EM, Sengupta S, *et al*. Discovery and refinement of loci associated with lipid levels. *Nat Genet* 2013; 45(11): 1274-83.

Wood AR, Esko T, Yang J, *et al*. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genetics* 2014; 46(11): 1173-86.

World Health Organization (WHO).

Obesity and overweight, 2024 [cited 2024 May 08]. Available from: URL: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

World Health Organization Regional Office for the Western Pacific (WHO WPRO). The Asia-Pacific perspective: redefining obesity and its treatment, 2000 [cited 2024 May 08]. Available from: URL: https://iris.who.int/bitstream/handle/10665/206936/0957708211_eng.pdf?sequence=1&isAllowed=y

Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data using the sequence kernel association test. *Am J Hum Genet* 2011; 89(1): 82-93.

Yeung AY, Tadi P. Physiology, obesity neurohormonal appetite and satiety control, 2023 [cited 2024 May 08]. Available from: URL: https://www.ncbi.nlm.nih.gov/books/NBK555906/