Université de Lille

Doctoral School of Biology and Health

Academic Year 2022/2023

Dissecting the genetic factors contributing to shared susceptibility to obesity and cancer

Thesis defended in fulfilment of the requirements for the degree of Doctor of Philosophy of the University of Lille, France

By **Jared Gichohi Maina**

Defended on 17th February 2023 before the jury composed of:

Examiner	**Prof Karoline Kuchenbäcker**
Examiner (Jury President)	**Prof Eric Sijbrands**
Reviewer	**Dr Christian Dina**
Thesis Co-supervisor	**Prof Inga Prokopenko**
Thesis Director and Co-supervisor	**Prof Philippe Froguel**

Table of Contents

ACKNOWLEDGEMENTS

I would like to express my appreciation to the jury members: Prof Karoline Kuchenbäcker, Prof Eric Sijbrands and Dr Christian Dina for accepting to review my work. To Eric and Christian, thank you for being my thesis progress assessors over the last three years.

I am sincerely grateful to Prof Philippe Froguel for providing the funding that enabled me to actualise this PhD, and for his academic supervision. I am particularly grateful for his guidance in preparing my academic papers, his advice to me on being an independent thinker in science, and how to properly interpret and present scientific findings. I also thank him for supporting me financially for all conferences I had the honour of attending, and for my academic visit to University of Surrey, UK

I am greatly indebted to Prof Inga Prokopenko for her supervision, mentorship, and support throughout the last 5 years for my Masters and PhD. Thank you for guiding me through the PhD during a global pandemic. Also, thank you for hosting me at the University of Surrey for 5 months during my final year of PhD

I would also like to thank all the members, past and present, in both of my co-supervisors' groups. Thank you to the UMR 1283/8199 in EGID for making me feel welcome in a new country with a different culture from mine, and for your assistance in preparing my thesis. I particularly thank Dr Amelié Bonnefond and the Team 2 including the bioinformatics (Lijiao, Mathilde, Dr Mickaël) and biostatistics (Dr Mehdi, Alaa, Souhila) teams, the open space students (including Lucas, Vincent, Dr Alex, Maxime, Cloé), Dr Martine, Dr Sarah, Lise, Nawel, Justine and many more. I would like to thank Dr Amna Khamis for her contribution in my paper, her hospitality and for being a role model to me as a young Black scientist. Thanks also to Marlene who helped teach me French and invited me and others for Christmas Eve dinners so we would not be alone during the festive period

I am equally grateful to the Prokopenko group members at the University of Surrey for their support since I started my MSc at the lab. Special thanks to Ivona, Dr Zhanna, Dr Ayse, Dr Harmen, Dr Anna, Dr Yevheniya for their assistance with coding and presentations. Thanks to Dr Marika Kaakinen for all your support and career advice when things got tough. Heartfelt gratitude and admiration to Igor Pupko and Liudmila Zudina who we did our MSc and PhD at the same time-we are almost there.

To my football friends in Lille (Kamara, Abdel, Balde, Seku and more) who provide an opportunity for me to enjoy a community of brotherhood every weekend, I appreciate you

I am extremely grateful to my UK family (Bruno, Regina, Ariah, Amari, Steve, Mary, Montel, Melanie, Morgan, Jackie, Mike, Thimba, Eve, Kisila, Carol), who ensured I felt at home away from home. Thank you for hosting me during my visit to UK and for constantly encouraging and cheering me on

To my friends in Kenya (Irungu, Manu, Shabz, Njeri, Ndalo, Kyule, Mary, Penina, Jackline, Dr Norman, Dr Kuria, Dr Kevin, Dr Munyua, Dr Kaigunye, Dr Ruhiu, Lorna, Chege, Muts) and Canada (Mash and Keshi), thank you for being my support system and all-round solid friends. I appreciate our friendship. Thanks to Mwari for helping me get past the last 6 months of my PhD

I equally grateful to my family including my grandmother, aunts, uncles and cousins. Thank you Wangeci for being the sister I never had. For your endless support and for being a role model

Last but not least, I pass my most sincere gratitude to my dear mum and dad for being my rock. I appreciate all the audio and video calls we have shared since I left the comfort of home. I may have missed you physically, but I felt your presence with me daily. Thank you also for your prayers, I know I got through difficult moments thanks to them

To my darling mother, thank you for being the reason why I never give up

To God Be The Glory

ABSTRACT

Obesity is a common, complex condition that poses serious health problems worldwide. It is also a known critical risk factor for some non-communicable diseases including cancers. Different anthropometric measures such as body mass index (BMI) and waist-to-hip ratio (WHR) have been used to assess obesity. The latter is an index for central or abdominal obesity while the former represents total or overall obesity. Epidemiological studies provide evidence that central and overall obesity measures may relate to cancer risk differently. The exact physiological mechanisms that enable the obesity and cancer co-morbidity remain unclear. However, certain factors such as insulin-like growth factors, hyperglycaemia, dysregulated lipid profile and adipokine factors have been hypothesised. Genome-wide association studies (GWAS) have identified numerous common genetic variations for obesity and cancer phenotypes. However, these variations provide only modest clues as to the underlying comorbidity. Nevertheless, output from GWAS can be applied to statistical methods such as polygenic scores and Mendelian randomization that aid in the unravelling of shared determinants. In this PhD project, I assessed the impact of overall and central obesity on the risk of cancers including overall breast, post-menopausal breast, prostate, colorectal, lung and pancreatic cancers. I defined the genetic correlation between BMI/WHRadjBMI and cancers using the UK Biobank dataset. I then used established BMI and WHRadjBMI genome-wide *loci* to create obesity polygenic scores which were then tested for association with cancer phenotypes in the UK biobank. Further, using established genetic variants associated with these phenotypes, I performed MR between the two obesity phenotypes and three cancers (breast, prostate and colorectal) to investigate the causal relationships between them.

RÉSUMÉ

L'obésité est une affection courante et complexe qui pose de graves problèmes de santé dans le monde entier. Elle est également un facteur de risque critique connu pour certaines maladies non transmissibles, dont les cancers. Différentes mesures anthropométriques telles que l'indice de masse corporelle (IMC) et le rapport taille-hanche (RTH) ont été utilisées pour évaluer l'obésité. Ce dernier est un indice de l'obésité centrale ou abdominale, tandis que le premier représente l'obésité totale ou globale. Des études épidémiologiques fournissent des preuves que les mesures de l'obésité centrale et de l'obésité globale peuvent avoir un rapport différent avec le risque de cancer. Les mécanismes physiologiques exacts qui permettent la comorbidité entre l'obésité et le cancer restent flous. Cependant, certains facteurs tels que les facteurs de croissance analogues à l'insuline, l'hyperglycémie, la dérégulation du profil lipidique et les facteurs adipokines ont fait l'objet d'hypothèses. Les études d'association pangénomique (GWAS) ont identifié de nombreuses variations génétiques communes pour les phénotypes de l'obésité et du cancer. Cependant, ces variations ne fournissent que de modestes indices sur la comorbidité sous-jacente. Néanmoins, les résultats des études d'association pangénomique peuvent être appliqués à des méthodes statistiques telles que les scores polygéniques et la randomisation Mendélienne (MR) qui aident à démêler les déterminants communs.

Dans ce projet de doctorat, j'ai évalué l'impact de l'obésité globale et centrale sur le risque de cancers, notamment le cancer du sein, le cancer du sein post-ménopausique, le cancer de la prostate, le cancer colorectal, le cancer du poumon et le cancer du pancréas. J'ai défini la corrélation génétique entre l'IMC/l'RTH et les cancers en utilisant l'ensemble des données de la UK Biobank. J'ai ensuite utilisé des loci génomiques établis pour l'IMC et l'RTH afin de créer des scores polygéniques

d'obésité dont l'association avec les phénotypes de cancer a ensuite été testée dans la UK Biobank. En outre, à l'aide de variantes génétiques établies associées à ces phénotypes, j'ai effectué une MR entre les deux phénotypes d'obésité et trois cancers (sein, prostate et colorectal) afin d'étudier les relations causales entre eux.

ABBREVIATIONS

AGE – Advanced Glycation End products

BCAC – Breast Cancer Association Consortium

BMI – Body Mass Index

CPRD – Clinical Practice Research Datalink

GIANT – Genetic Investigation of ANthropometric Traits

GLUT1 – Glucose Transporter 1

GLUT4 – Glucose Transporter 4

GWAS – Genome-Wide Association Studies

HDL – High Density Lipoprotein

IGF-IIR – Insulin-like Growth Factor 2 Receptor

IGF-IR – Insulin-like Growth Factor 1 Receptor

IGFBP – Insulin-like Growth Factor Binding Proteins

IR – Insulin Receptor

LDL – Low Density Lipoprotein

LDSC – Linkage Disequilibrium Score

MAF – Minor Allele Frequency

MR – Mendelian randomization

ObR – Leptin Receptor

PGS – Polygenic Score

RAGE – Receptor for Advance Glycation End products

SNP – Single Nucleotide Polymorphism

T2D – Type 2 Diabetes

UKBB – UK Biobank

WC – Waist Circumference

WHR – Waist-to-Hip Ratio

WHRadjBMI – BMI adjusted WHR

1. INTRODUCTION

1.1 Obesity and cancer

Overweight and obesity are common and complex conditions defined by excessive fat accumulation in adipose tissue that pose a threat to health.

Globally, obesity continues to become a health concern affecting not just developing countries, but also in low- and middle-income countries. Worldwide, an estimated 13% of adults in 2016 were obese and furthermore, 2.8 million deaths yearly are attributed to being overweight and obese[1]. Additionally, excess body weight has been established as a risk factor for several non-communicable diseases including cancers[1].

Cancer refers to a disease characterised by abnormal and uncontrolled cell growth that has potential of spreading to other parts of the body. According to a recent international cancer research study, the top five common cancer types in the world are female breast, lung, colorectal and prostate cancers[2]. An estimated 19.3 million new cancer cases were reported in 2020 with this number projected to exceed 28 million by 2040[2]. Moreover, cancer is the second leading cause of mortality worldwide, after cardiovascular disease, with nearly 10 million deaths attributed to cancer as of 2020[3].

1.2 Measures of obesity

Since its development in the mid-1800s, the body mass index (BMI) is the most common anthropometric measure use in clinical and research settings to indirectly assess adiposity. It is computed by dividing someone's weight in kilograms by the square of their height in meters (kg/m^2). Based on the World Health Organization

(WHO) guidelines, BMI is used to define four main weight categories[1]. Specifically, normal healthy weight includes BMI between 18.5 and 24.9 kg/m^2, while BMI less than 18.5 kg/m^2 is considered underweight. Individuals with BMI greater than or equal to 25 kg/m^2, but below 30 kg/m^2 are considered overweight. BMI greater than or equal to 30 kg/m^2 defines the obese category.

Despite being a routine measure of adiposity, BMI falls short of being a perfect measure for several reasons. For instance, BMI may not accurately define obesity since it does not distinguish between lean and fat mass[4]. Additionally, individuals who may be metabolically unhealthy can be classified in the normal weight category[5]. Adipose tissue distribution, which is a significant risk factor in type 2 diabetes (T2D), cardiovascular disease, and cancer, is also not captured using BMI. Therefore, other anthropometric measures that assess adipose tissue distribution and improve clinical evaluation of metabolic health have been developed including waist circumference (WC) and the waist-to-hip (WHR) ratio (unitless measure). WHR is defined by dividing someone's WC, measured in cm, to their hip circumference, in cm. According to the WHO, a healthy WHR is 0.8 or lower for women and 0.95 or lower for men[6]. A WHR of 0.86 and greater is considered a high health risk for women, while for men, a WHR equal to or greater than 1.0 poses high health risk[6].

While BMI is considered an index for overall/total adiposity, WC and WHR assess central/abdominal/visceral adiposity. Central adiposity correlates to insulin resistance, dyslipidaemia, hypertension which comprise the metabolic syndrome[7–9]. It thus follows that overall and central adiposity measures may relate to disease risk/prevalence differently, with cancer being the disease of interest for my research.

1.3 Epidemiological associations

The relationship between cancer and obesity has been a growing topic of research over the last three decades. In fact, recent global estimates on obesity and cancer risk have indicated that among adults aged 30 and above, approximately 3.6% of all new cancer cases can be linked to high BMI[10].

From multiple studies examining the relationship between body weight and cancer incidence and mortality, it appears that the link between the two is gender-, site-, age- and menopause status-specific[11–13].

For instance, in large prospective study among 900,053 cancer-free adults (404,576 men and 495,477 women) at baseline in the United States of America (USA), the authors defined the relationship between obesity and cancer mortality following a 16-years follow-up period[11]. More specifically, they tested for epidemiological association between overweight and obesity (measured using BMI) and the risk of death caused by overall cancer at cancer-specific sites in the body, highlighting the following. 1) For both men and women with BMI > 40 kg/m^2, the overall mortality due to all cancers was 52% and 62% higher, respectively, than their counterparts of normal BMI range (23 kg/m^2 – 29 kg/m^2)[11]. 2) Additionally, high BMI was associated with a higher risk of death due to cancer of the colon, rectum, liver, oesophagus, gall bladder, kidney and pancreas in both men and women[11]. 3) The association between high BMI and cancer mortality was gender specific for specific cancer types[11]. For men with a BMI higher than 35 kg/m^2, the authors observed an increased risk of death due to cancers of the prostate and stomach, compared to men within normal BMI range. Similarly, women with BMI higher than 40 kg/m^2 had significant risk of death due to cancers of the ovary, cervix, uterus, and breast. Overall, this study demonstrated, by leveraging on large

scale data, that overweight and obesity was associated with greater risk of death from all cancers in both men and women.

While the study above focused on the relationship between obesity and the risk of death by cancer, others have assessed the relationship between obesity and cancer incidence[12,13].

In a landmark systematic review and meta-analysis of prospective observational studies, Renehan et al. evaluated the relationship between incremental increase in BMI and the risk of cancer incidence for both men and women[12]. In total, they analysed data from 141 articles spanning 221 datasets comprising 282,137 incident cancer cases (154,333 men and 127,804 women). They reported that for every 5 kg/m^2 increase in BMI among men, there was significant increase in risk of cancers of the colon, rectum, thyroid, kidney as well as oesophageal adenocarcinoma, non-Hodgkin's lymphoma, and leukaemia[12]. In contrast, they report a significant decrease incidence of lung cancer and squamous cell carcinoma of the oesophagus associated with every 5 kg/m^2 BMI increase[12]. In women, similar BMI increments were associated with increased incidence of endometrial, renal, thyroid, post-menopausal breast, pancreatic and colon cancers as well as oesophageal adenocarcinoma and leukaemia[12]. Increase in BMI was however associated with a decreased risk of lung and premenopausal breast cancers and squamous cell carcinoma of the oesophagus. Additionally, the authors highlighted several points based on their analyses. 1) For post-menopausal breast cancer, the direct association observed with increased BMI was consistent in studies that included post-menopausal women only and those that included both pre- and post-menopausal breast cancer[12]. 2) The association between increased BMI and cancer differed between the sexes for some cancers[12]. For instance, in colon cancer, the associations with increased BMI were stronger in men than in women. However,

for rectal cancer, the associations with increased BMI were stronger in women than men. The association with increased BMI and pancreatic cancer appeared similar in both men and women[12]. 3) Despite the association between increased BMI and most cancers being consistent across different populations, for some cancer sites the risk estimates varied from one population to the other[12]. Case in point, the authors show that despite North America, European and Australian populations having an inverse association between increased BMI and premenopausal cancer, in Asia-Pacific populations the association was positive[12].

Finally, in a more recent study based on routinely collected primary care records, the authors investigated the relationship between BMI and site-specific cancers in the United Kingdom (UK)[13]. The UK Clinical Practice Research Datalink (CPRD) captures a wide range of computerised primary care data from general practitioners in the UK. Data available in the CPRD include hospital admissions and referrals, primary and secondary diagnosis, information regarding lifestyle factors (e.g., smoking status) and body measurements such as height and BMI. In this cohort study, the authors present results for 22 cancers among 5.24 million individuals with BMI data and highlight several findings. 1) Higher BMI was associated with an increased risk of uterine, gallbladder, kidney, cervical, thyroid, liver, colon, ovarian, post-menopausal breast cancers, and leukaemia while inverse associations were shown between high BMI and lung, oral cavity, premenopausal breast and prostate cancers[13]. 2) For colon and liver cancers, the associations with BMI were stronger in men than women[13]. 3) There was a positive association between BMI and both pre- and post-menopausal breast cancers at BMI levels less than 22 kg/m^2. However, above this BMI cut-off, the risk of premenopausal breast reduces[13]. 4) A similar pattern was seen for prostate cancer in men where the risk associated with BMI peaked at 24 kg/m^2, after which the risk of

prostate cancer reduces markedly[13]. 5) Low BMI was associated with higher risk for lung, oral cavity, and stomach cancers but only among current and former smokers[13].

1.3.1 Limitations of observational studies

Despite their usefulness in highlighting the associations between obesity and cancer, several limitations of epidemiological studies need to be considered.

Results from epidemiological studies often suffer from bias and confounding by factors that are either inaccurately or completely accounted for in the study design. A classic confounder that has emerged in almost all studies is smoking. Several studies have reported the inverse association between BMI and lung, oral cavity and stomach cancers[11–14]. This association, however, only holds among current and former smokers, and is not seen in those who have no history of smoking. Moreover, similar apparent confounding by smoking has been reported in oesophageal cancer[12]. High BMI is shown to be associated with higher risk of oesophageal adenocarcinoma but is inversely associated with squamous cell carcinoma of the oesophagus which is more associated with smoking. Indeed, it has been shown that for the same sex and age, smokers tend to have lower BMI than their non-smoking counterparts[15]. Therefore, the interpretation of such observational findings, as well as the study design, needs careful consideration of such factors.

Epidemiological studies assessing the relationship between obesity and cancer have focused mostly on overall obesity. As such, there are far fewer studies assessing the relationship between measures of central/abdominal obesity and cancers such as WC and WHR. BMI is shown to be an imperfect measure of obesity and it follows that other anthropometric measures such as those assessing central adiposity need addressing.

In fact, for some cancers such as of the prostate, central obesity appears to be a better predictor of cancer risk than overall BMI[16,17]. Additionally, central obesity and other components of the metabolic syndrome have been shown to be elevate the risk of pancreatic, colon and breast cancers[18–20]. More studies are therefore needed to quantify the relationship between central obesity and cancers.

1.4 Mechanisms linking obesity and cancer

Several mechanisms have been suggested to play a role in the manifestation of the obesity-cancer co-morbidity.

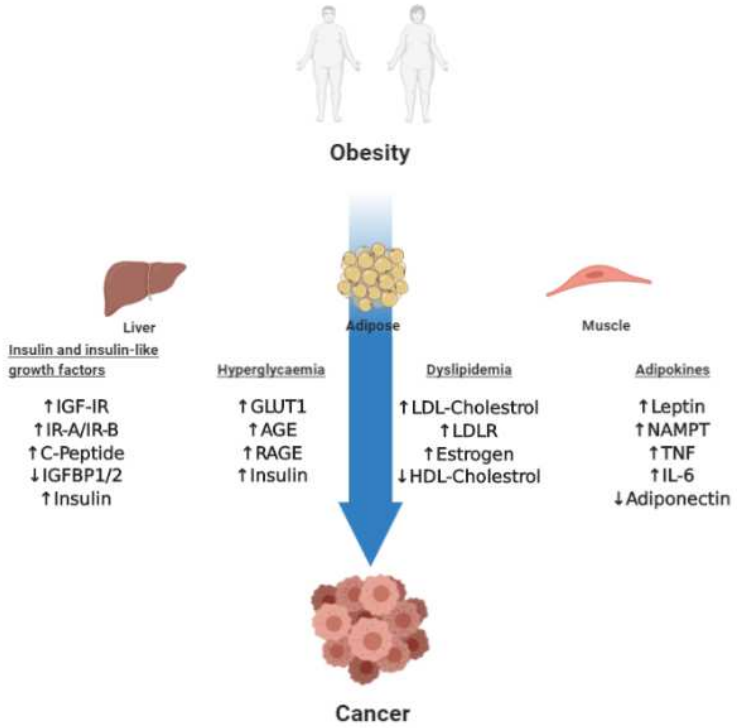


Figure 1. Mechanisms linking obesity and cancer. The liver, adipose and muscle tissues play a role in the link between obesity and cancer. The mechanisms involved include insulin and insulin-like growth factors, hyperglycaemia, dyslipidaemia and adipokines.

(Source: Adapted from Gallagher and LeRoith. Physiological Reviews (2015) 95(3) 727-748)

1.4.1 Insulin and insulin-like growth factors

The insulin-like growth factor system comprises of the insulin receptor (IR), insulin-like growth factor 1 and 2 receptors (IGF-IR/IGF-IIR) and their ligands: insulin, IGF-I, IGF-II and insulin-like growth factor binding proteins (IGFBP)[21]. Circulating hyperinsulinemia, leading to insulin resistance has been associated with an increase in cancer[22,23]. Overexpression of IGF-IR has been shown in breast, colorectal, liver and prostate cancers[24] with a loss of tumour suppressor genes BRCA1, p53 and PTEN potentially driving the increased cancer risk[25,26]. Hyperinsulinemia driven by IR overexpression on tumour cells may also lead to tumour growth and progression in breast, colon, lung and prostate cancers[21,23]. The IR has two isoforms: IR-A and IR-B. IR-A lacks exon 11 of the *IR* gene and is mainly expressed in cancer cells increasing their affinity for IGF-II and insulin, providing a possible link between the cancer-promoting effects of hyperinsulinemia seen in individuals with obesity[21]. Dysregulated signalling in tumour cells often leads to differential expression of splice factors (e.g., SRSF3) which leads to increased IR-A/IR-B ratio responsible for the effects of hyperinsulinemia on tumour development[21,27]. C-peptide levels, a more stable marker of insulin secretion, have been associated with increased incidences of breast and colorectal cancer[28,29] but have not been associated with prostate cancer[30–32].

1.4.2 Hyperglycaemia

Cancer cells preferentially use glycolysis for energy production over oxidative phosphorylation; a hallmark of cancer cells[33]. Metabolic tissues (skeletal and adipose) use the glucose transporter 4 (GLUT4) to take up glucose into their cells. However,

most cancer cells use the GLUT1 with increased affinity for glucose[34]. This promotes aerobic glycolysis in those cells which provides the precursors needed for lipid, amino acid and nucleotide synthesis[33]. Increase in HbA1c levels, a marker for circulating glucose levels, has been associated with a higher risk for breast and colorectal cancer but no correlation has been observed with prostate cancer[35]. Circulating hyperglycaemia also leads to production of advanced glycation end products (AGEs) and their receptors (RAGEs)[36]. AGEs are formed when sugars such as glucose non-enzymatically react with the free amino groups on proteins, lipids, and nucleic acids[36]. Individuals with obesity and T2D have higher levels of AGEs and RAGEs. Oxidative stress and inflammation which arise from the interaction of RAGEs and their ligands lead to promoter tumour growth, angiogenesis, and metastases[37].

1.4.3 Dyslipidaemia

Obesity is characterised by elevated levels of low-density lipoprotein (LDL) – cholesterol and low levels of high-density lipoprotein (HDL) – cholesterol. Elevated levels of total cholesterol, triacylglycerides (TAGs) and low levels of HDL-cholesterol have been associated with up to 20% increase in cancer risk[38]. In addition, polymorphisms in genes associated with hyperlipidaemia (*APOE, APOA-1*) have been associated with an increased breast cancer risk[39]. Cholesterol plays a chief role in cancer growth and progression through increased PI3K/AKT signalling as shown in vitro in breast, colon and prostate cancer cell lines which leads to increased cell proliferation[40–42]. Cholesterol is also a precursor for progesterone, oestrogen, and androgen. Studies have shown that human prostate cancers are able to synthesise their own androgens, including testosterone, from cholesterol[43,44].

1.4.4 Adipokines

Adipose tissue factors (adipokines), inflammatory cytokines and enzymes produced by adipose tissue are abnormally regulated in obesity and T2D promoting tumour growth and metastases. The adipose tissue presents a vital organ in tumour development and progression in many organs as it not only surrounds many organs (e.g. heart, kidney) but is also abundant in organs where cancer develops, such as breast. The adipose tissue provides a local environment that enables cancer cells proliferation. Various adipokines and cytokines are relevant to cancer including leptin, adiponectin, resistin, TNF-α and interleukin 6 (IL-6)[45–49]. Leptin is a pro-inflammatory adipokine that is a regulator of appetite[50] that binds the leptin receptor (ObR). Higher ObR expression is observed in breast tumours[51] and is associated with poor prognosis. Binding of the ObR by leptin activates key intracellular pathways that promote tumour growth and metastases. These pathways include those involved in cell growth and survival (PI3K/AKt, cyclin D1), inflammation response (NF-κβ, COX-2), angiogenesis (STAT4, VEGF) and differentiation (Notch, Wn)[46,51–54]. Adiponectin (an anti-inflammatory adipokine) plasma protein levels have been shown to be low in individuals who are obese and is associated with increased cancer risk[47,55]. The protective role of adiponectin signalling in cancer progression is mediated through phosphorylation of the AMPK which antagonises leptin signalling[55]. Resistin is another pro-inflammatory adipokine associated with insulin resistance and is elevated in obesity and T2D. Resistin mediates the effects of insulin resistance (described in hyperglycaemia above) by activating the suppressor of cytokine signalling 3 (SOCS3) that interferes with insulin signalling[48]. Resistin is highly expressed in prostate cancer and promotes its

proliferation via the P13K/AKt signalling pathways[56]. TNF-α and IL-6 are pro-inflammatory cytokines that are overexpressed in obesity. The pro-inflammatory environment created by such cytokines promotes insulin resistance by blocking adipocyte insulin action[49]. The ensuing insulin resistance can promote tumour development as illustrated earlier. Inflammatory cytokines also promote cancer development via activation of NFκβ and Stat3 signalling pathways involved in angiogenesis giving the cancer cells metastatic properties[57].

1.5 Genome-wide association studies

Genome-wide association studies (GWAS) are instrumental in dissecting the associations between common genetic variation (single nucleotide polymorphisms [SNPs] with a minor allele frequency [MAF] > 5%) and diseases or traits of interest. GWAS help unravel specific positions on a chromosome where a particular DNA variant or other genetic marker associated with a disease or trait is located. The identification of these positions on a chromosome, referred to as *loci* (singular *locus*), has enabled the successful elucidation of the genetic architecture of complex traits and diseases (https://www.ebi.ac.uk/gwas/). Since the first hallmark GWAS was conducted in the early 2000s[58], there have been significant advances in GWAS. Notably, there has been an increase in the study sample sizes involved and the number of common SNPs amenable for association analysis. Equally, the downstream application of GWAS output has seen remarkable improvements.

1.5.1 GWAS of obesity phenotypes

The largest-to-date GWAS of obesity phenotypes have been realised in-part through the Genetic Investigation of Anthropometric Traits (GIANT) consortium (GIANT consortium - Giant Consortium (broadinstitute.org)).

1.5.1.1 BMI GWAS

In the most recent and largest GWAS of BMI to-date[59], the authors meta-analysed previous GWAS of BMI by GIANT consortium[60] and UK biobank (UKBB) BMI GWAS. Altogether, there were 681,275 participants in this meta-analysis. Leveraging on this sample size, there were 670 genome-wide significant *loci* (P<5X10^{-8}) associated with BMI (**Figure 1**). The proportion of phenotypic variance in BMI attributable to common SNPs (SNP heritability was 22.4% (standard error (SE)=0.037).

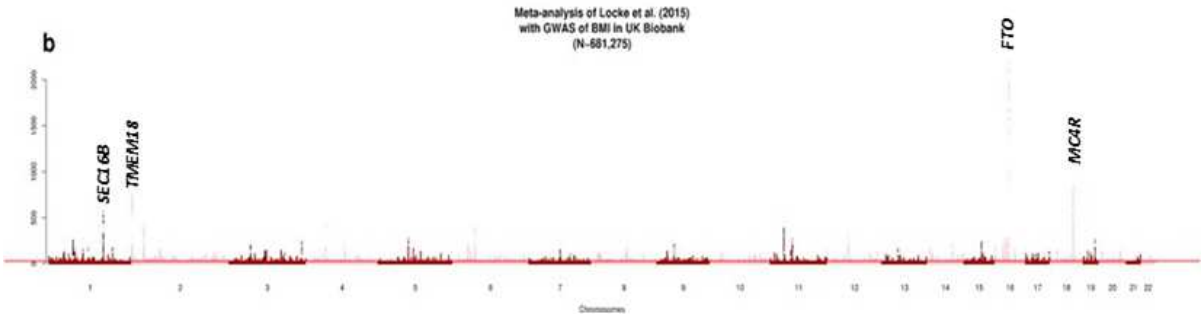


Figure 2. Manhattan plot of BMI GWAS meta-analysis performed by GIANT consortium (Source: Yengo et al. Human Molecular Genetics (2018) 27:20)

1.5.1.2 WHR/WHRadjBMI GWAS

Similarly, as with BMI GWAS, the largest WHR and BMI adjusted WHR (WHRadjBMI) GWAS was a meta-analysis performed by the GIANT consortium[61]. They meta-analysed studies included previous WHR/WHRadjBMI GWAS[62] and UKBB GWAS on WHR/WHRadjBMI. In total, there were 697,734 and 694,649 study participants in the meta-analysis for WHR and WHRadjBMI respectively. There were 316 and 346 genome-wide significant *loci* associated with WHR and WHRadjBMI respectively (**Figure 2**). The SNP heritability of WHR and WHRadjBMI was 19.4% (SE=0.002) and 17.4% (0.002) respectively.

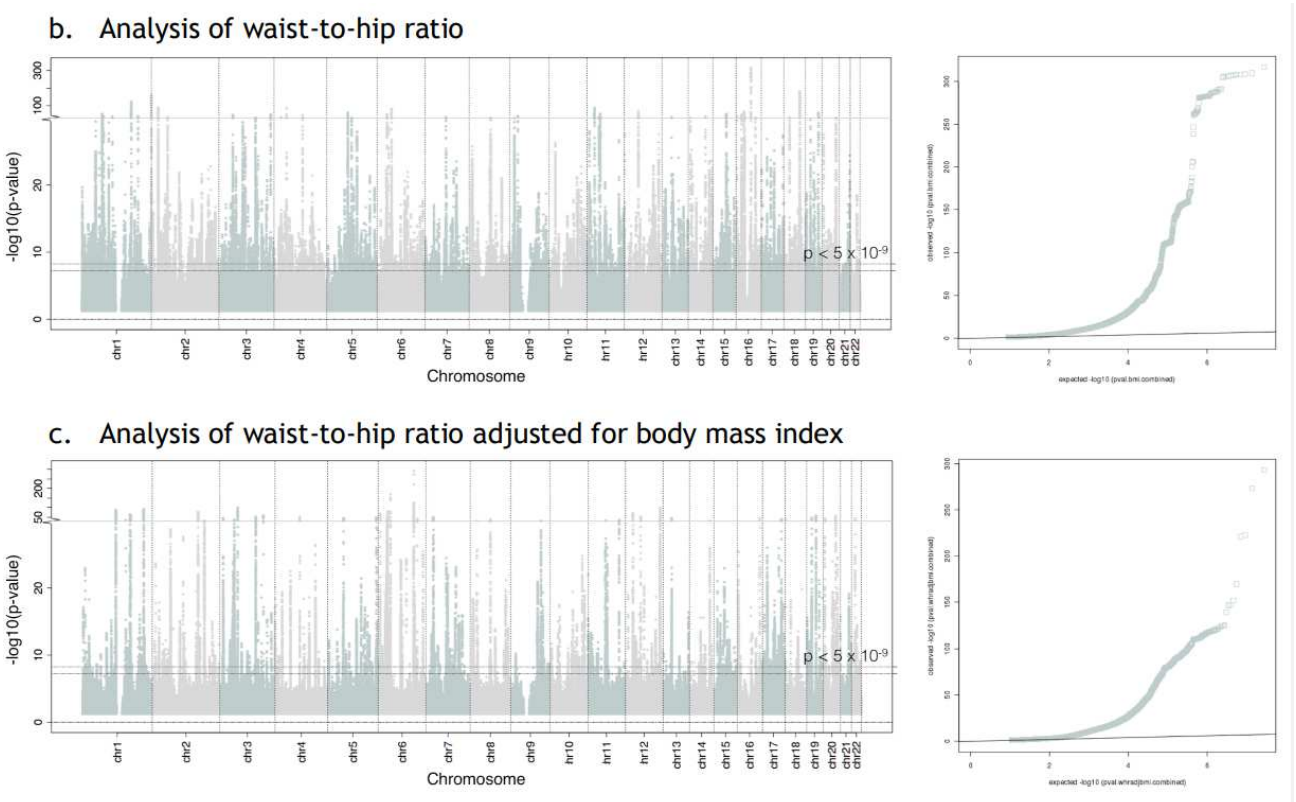


Figure 3. Manhattan plot of WHR and WHRadjBMI GWAS meta-analysis conducted by GIANT consortium

(Source: Pulit et al. Human Molecular Genetics (2019) 28:1)

1.5.2 Cancer GWAS

In the same way as with obesity GWAS, concerted efforts through working groups and consortia led to the discovery of common genetic variation associated with different cancers.

1.5.2.1 Breast cancer

According to the WHO, the most common cancer in the word in terms of new cases was breast cancer[63]. The most recent and largest GWAS of breast cancer was achieved through a meta-analysis of 82 breast cancer studies across 20 countries under the Breast Cancer Association Consortium (BCAC)[64]. The total sample size in the meta-analyses included 133,384 cases and 113,789 controls (N=247,173) of European ancestry. This meta-analysis brought the total number of genome-wide significant *loci* associated with breast cancer to 201.

1.5.2.2 Prostate cancer

Prostate cancer is the most common cancers among men. The largest GWAS to date of prostate cancer is a meta-analysis composed of 52 studies[65]. In total, there were 79,148 cases and 61,106 controls (N=140,254) of European ancestry. From this effort, the resultant number of genome-wide significant *loci* for prostate cancer was 248.

1.5.2.3 Colorectal cancer

Recent global data on cancer suggest that colorectal cancer is the second leading cause of cancer deaths. The most recent GWAS of colorectal cancer comprises a meta-analysis of 16 studies[66]. This study had 34,627 cases and 71,379 (N=106,006) controls of European ancestry. Currently, there are 137 genome-wide significant *loci* associated with colorectal cancer.

1.5.2.4 Pancreatic cancer

Pancreatic cancer is a leading cause of cancer-related mortality worldwide. In fact, in America, it ranks third after lung and colon cancers in terms of cancer-related deaths. The largest GWAS of pancreatic cancer comprises of 9,040 cases and 12,946 controls (N=21,536) of European ancestry[67]. The two consortia involved in this meta-analysis were the Pancreatic Cancer Cohort Consortium (PanScan) and the Pancreatic Cancer Case-Control Consortium (PanC4). Currently, there are 22 genome-wide significant *loci* for pancreatic cancer.

1.5.2.5 Lung cancer

Globally, lung cancer ranks first among the leading cause of cancer-related deaths[63]. The largest GWAS to date of lung cancer included 27,065 study participants of European ancestry (14,803 cases and 12,262 controls)[68]. This meta-analysis identified 18 susceptibility *loci* associated with lung cancer.

Table 1. Summary of the cancer GWAS studies to date

Cancer	Cases	Controls	Total	*Number of associated Loci*	PubMed ID
Breast	133,384	113,789	247,173	201	32424353
Prostate	79,148	61,106	140,254	248	29892016
Colorectal	34,627	71,379	106,006	137	31089142
Pancreatic	9,040	12,946	21,536	22	29422604
Lung	14,803	12,262	27,065	18	28604730

Legend: *loci*=number of genome-wide significant loci

1.5.3 Limitations of GWAS

GWAS have been pivotal in broadening our understanding of complex diseases over the last decade. However, several limitations have hampered the utility of GWAS in understanding the pathophysiology underlying most complex, polygenic phenotypes.

GWAS study design focuses mostly on SNP of common allele frequency (MAF>5%). The majority of common SNPs tend to have moderate to small effects sizes on a phenotype.

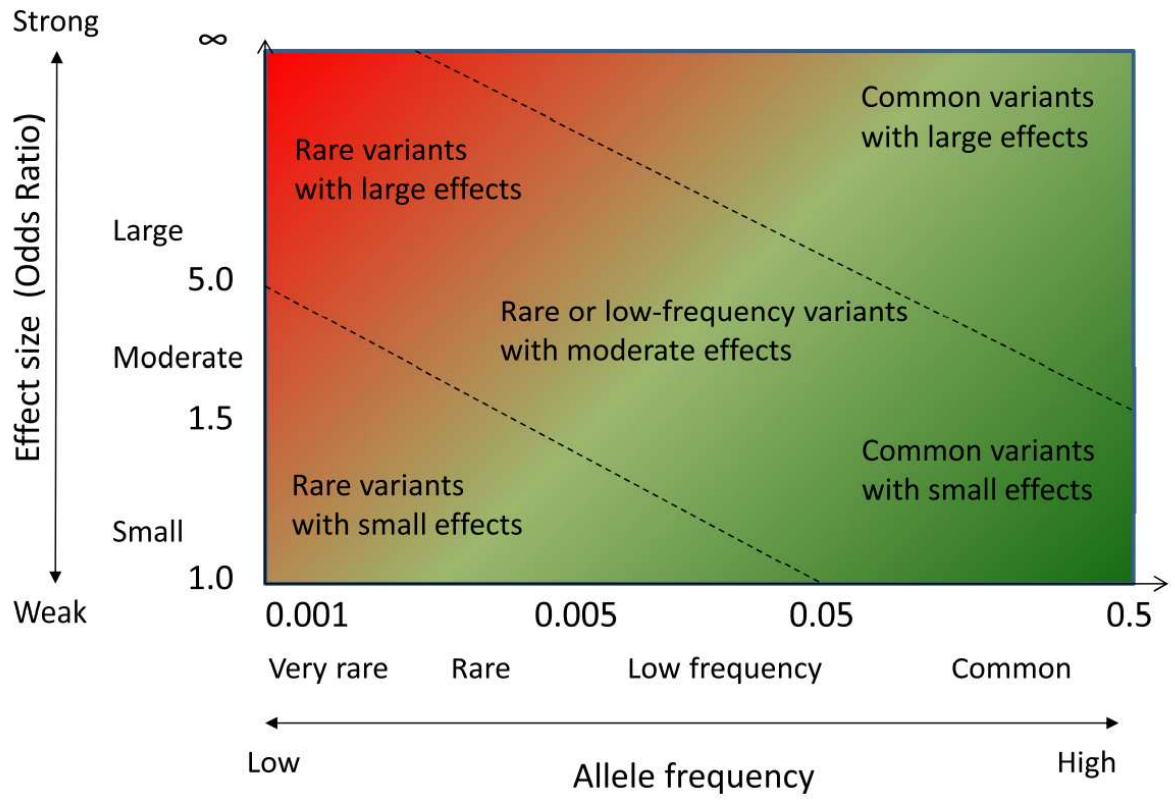


Figure 4. Graph highlighting the relationship between effect size estimates and the minor allele frequency

(Source: Roten et al. BMC Pregnancy and Childbirth (2015) 15:319)

Consequently, individual associations from typical GWAS often have modest effect sizes, while attaining to the strict significance thresholds set up for multiple testing

correction. The proportion of phenotypic variance explained by genetic factors is referred to as heritability[69,70]. Narrow-sense (h^2) and broad sense (H^2) heritability, usually defined from pedigree studies, refer to the phenotypic variance explained by additive and total (additive and non-additive) genetic effects respectively[69,70]. SNP heritability (h^2_{SNP}) on the other hand refers to the proportion of variance explained by genome-wide significant *loci* from GWAS[71].

Since the advent of GWAS, there has emerged the so-called "issue of missing heritability", where h^2_{SNP} estimates are usually much less than h^2 estimates[71,72]. One explanation suggested to account for the "missing heritability" is the lack of coverage of rare and low-frequency variation in genotyping arrays, as seen in most GWAS of the past decade[60,62,73]. Other explanations proposed include the existence of gene-by-gene, or gene-by-environment interactions[71–73].

The results from individual GWAS studies offer little in elucidating potential shared pathophysiology between related phenotypes. For any two related phenotypes, such as obesity and cancer, their individual GWAS results provide association results independent of each other. The genetic correlation and/or heritabilities between such traits is not taken into consideration in typical GWAS pipelines. However, there are methods that have been developed to jointly analyse phenotypes in GWAS[74–76] and are described in **Section 1.5.4.3.**

1.5.4 Application of GWAS outputs

The above-mentioned limitations notwithstanding, output from GWAS can be incorporated in downstream analyses that enhance the utility of GWAS.

1.5.4.1 Polygenic (risk) scores (PGS)

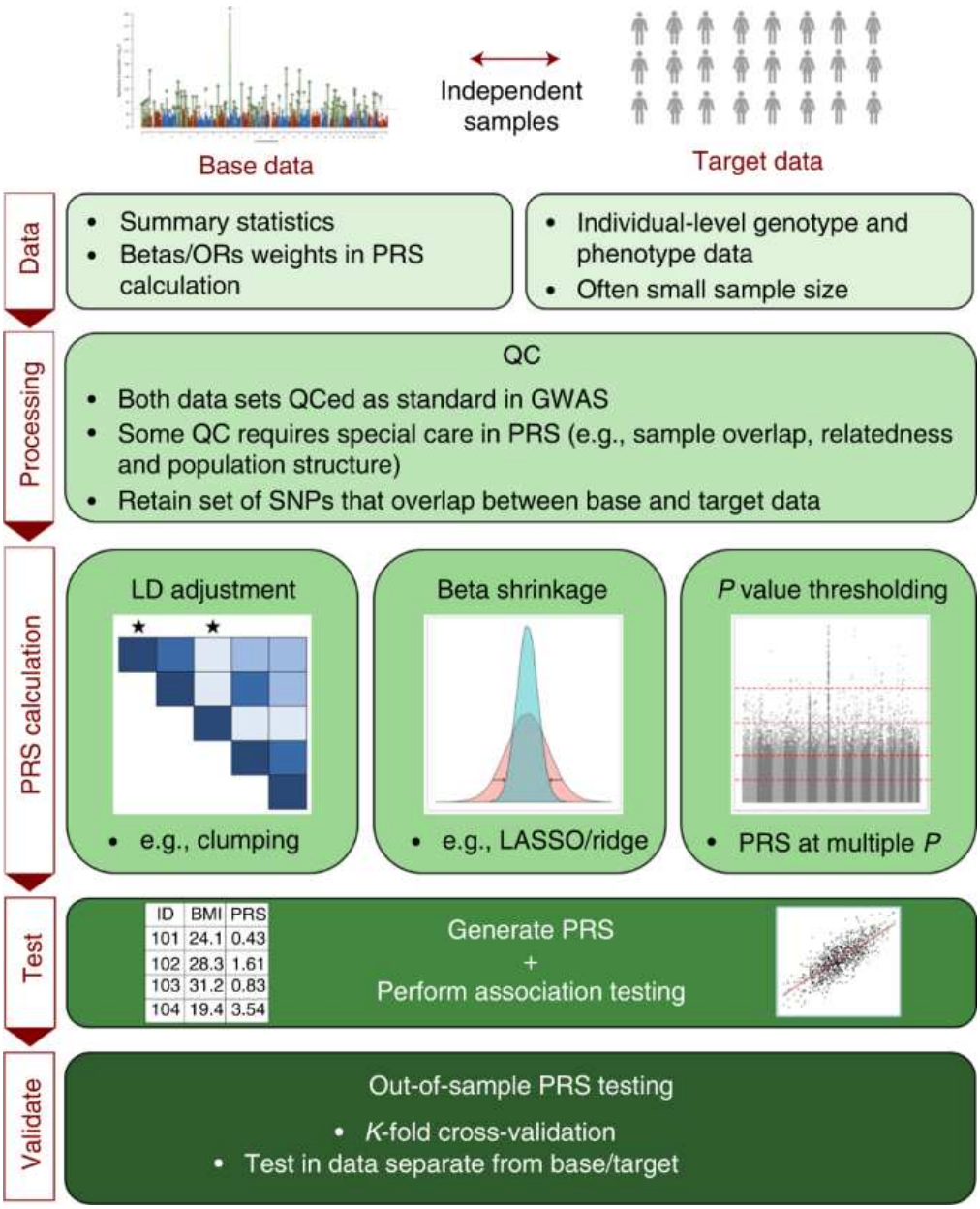


Figure 5. Polygenic scores analyses overview
(Source: Choi et al. Nature Protocols (2020) 15 :2759-2772)

As highlighted previously, GWAS identified numerous genetic variations associated with complex, polygenic phenotypes. However, independently these variants have modest effect sizes thus limiting their utility in predictive analyses. Statistical genetics methods such as polygenic scores (PGS, continuous phenotypes) or polygenic risk scores (PRS, binary phenotypes) have been developed to combine the effects of multiple variants across the genome to improve their predictive power[77,78].

A polygenic score (PGS) refers to the weighted sum of (genome-wide) risk variants associated with a particular phenotype. The variants are weighted by their effect sizes and are derived from the most informative GWAS, usually the largest. Summary statistics from GWAS (effect sizes and their *p*-values), through which the PGS are based on, constitute what is referred to as the base data. On the other hand, target data refers to the genotype-phenotype data for the individuals used to calculate the PGS. It is important to ensure that the base and target data are independent with no sample overlap. The independence of datasets reduces the inflation of the association between the PRS and phenotypes of interest. At the same time, the predictive ability of PGS also depends on the ancestral similarity between base and target datasets[79].

Both the base and target data must undergo further quality control steps[78]. The base pair positions in both base and target data should be from the same genome build. Additionally, strand ambiguous SNPs which cannot be resolved using allele frequencies, and duplicated SNPs should be excluded from the analysis. Strand-flipping of mismatching alleles between the base and target data is performed as part of most PGS software pipelines. Otherwise, unresolved mismatching SNPs should be excluded from the analyses.

PGS calculation can be done using various platforms including *Plink*[80] and dedicated PRS software such as LDpred[81] and PRSice-2[82]. Once constructed, PGS can be used

to test for association with phenotypes of interest, disease status prediction among other uses[77,78].

1.5.4.2 Mendelian randomization

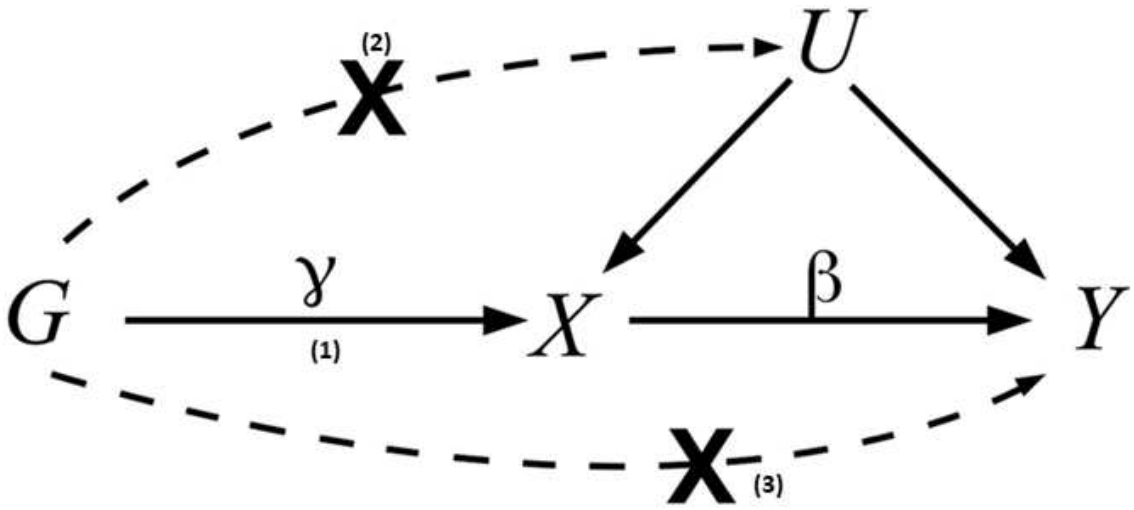


Figure 6. Mendelian randomization framework and assumptions. *G* represents the genetic variants (SNPs), *X* is the exposure, *Y* is the outcome, and *U* represents confounders. ɣ is the SNP-exposure association. β is the causal effect estimate of the exposure on the outcome.

(Source: Adapted from Bowden and Holmes. Research Synthesis Methods (2019) 10(4)486-496)

Another application of GWAS output has been in Mendelian randomization (MR) studies. Genetic variants associated with the exposure constitute the "instrument"; their distribution in the populations is random, given the random nature of inheritance patterns and fixation of alleles at the point of conception. In MR analyses, genetic variants (typically SNPs from GWAS) are used as instrumental variables[83] (IVs) (*G*) to assess the causal relationship between a risk factor (exposure, *X*) and a health outcome of interest *Y* (**Figure 5**)[84].

Three core IV assumptions exist (**Figure 5**)[84,85]:

1. *G* should be associated with the exposure.
2. *G* should be independent of confounders of the exposure-outcome association
3. *G* is associated with the outcome only through the exposure

In most MR studies, the relationship between an exposure and an outcome, plus the reverse is investigated. This gives rise to bi-directional MR studies. Leveraging on GWAS summary statistics of both the exposure and the outcome, researchers are able to perform two-sample bi-directional MR through software such as the *TwoSampleMR* R package[86].

In a recent MR study, the authors investigated the relationship between two obesity related traits (BMI and WHR) and breast, colorectal, ovarian, prostate and lung cancers[87]. They used cancer GWAS summary statistics from the Genetic Associations and Mechanisms in Oncology (GAME-ON) Consortium which constituted 51,537 cases and 61,600 controls across the cancers analysed. 77 and 14 SNPs of BMI and WHR respectively derived from published GWAS[60,88] were used as instrument variables in the one-sample MR study. They reported a statistically significant inverse relationship between BMI and both overall and oestrogen-receptor (ER)- negative breast cancer. Additionally, BMI was causal for ovarian, lung and colorectal cancers. WHR MR tests were not significant for any cancer tests. However, there were an inverse association between WHR and overall breast cancer that was marginally outside significance threshold. The reverse direction, cancers to obesity phenotypes, was however not investigated. Given the limited number of instrument variables for BMI and WHR, as well as the cancer sample sizes, the statistical power was limited in this analysis. Therefore, larger sample sizes and more instrument variables would boost the findings of such analysis.

1.5.4.3 Multi-phenotype GWAS

Conventional GWAS analyse diseases and phenotypes independently. Therefore, the association results from standard GWAS offer little in explaining underlying genetic determinants between related traits.

By jointly taking into account information from related traits, multi-phenotype GWAS approaches help improve the power for loci discovery, improve the accuracy of effect size estimates and provide potential indicators of multi-phenotype effects such as pleiotropy. Several tools exist to perform multi-phenotype analyses of GWAS using either individual level or summary level data[74–76]. GWAS summary statistics of related traits can thus be jointly analysed to unravel underlying genetic co-morbid determinants.

1.5.4.4 Genetic correlation

The proportion of phenotypic variance between two phenotypes that is attributable to genetic causes is referred to their genetic correlation (r_G). Genetic correlation estimates range from 0 to 1 with 0 signifying no genetic correlation and 1 suggesting complete genetic correlation.

Tools such as the linkage disequilibrium score (LDSC) regression tool have enabled the efficient computation of genetic correlation estimates between phenotypes[89] using GWAS summary statistics. Genetic correlation between phenotypes may be the result of linkage disequilibrium, biological pleiotropy or underlying confounding[90].

2 PROBLEM STATEMENT

There is growing evidence from observational studies of the link between obesity and risk of cancer incidence and mortality. Several mechanisms that potentially contribute to the emergence of the two diseases have over the years been postulated. However, our understanding of the co-morbidity remains limited.

As sample sizes in GWAS increase, numerous SNPs have been identified for both obesity and cancer phenotypes. However, individually these GWAS contribute modestly to explaining the shared genetic determinants between obesity and cancer. Various tools that leverage on GWAS output have been developed including PGS and MR. However, existing studies have been limited in statistical power due to limited number of published variants and low sample sizes at the time these studies were conducted.

As most GWAS studies make their summary statistics publicly available, and the emergence of large biobanks such as the UK biobank, researchers can design more powerful studies leveraging on improved statistical power.

3 HYPOTHESIS AND AIMS

We hypothesize that there are shared genetic determinants between obesity and cancer that can be elucidated using polygenic scores and Mendelian randomization analyses applied to large scale genetic data.

The present project includes the following aims:

1. To define the genetic correlation between overall (BMI) and central (WHRadjBMI) obesity and cancers in the UK Biobank resource. These cancers included overall and post-menopausal breast, prostate, colorectal, pancreatic and lung cancers
2. To construct BMI and WHRadjBMI polygenic scores from the largest GWAS of these phenotypes and test for their association with the above-mentioned cancers defined in the UK Biobank
3. To assess the causal relationships between the two obesity phenotypes (BMI and WHRadjBMI) and breast, prostate, pancreatic and colorectal cancers to perform a two-sample bi-directional Mendelian randomization approach

4 FIRST ARTICLE

“Abdominal obesity is a more important causal risk factor for pancreatic cancer than overall obesity”

(Brief communication article: Accepted by the European Journal of Human Genetics)

Abdominal obesity is a more important causal risk factor for pancreatic cancer than overall obesity

Jared G Maina[1,2], Vincent Pascat[1,2,3], Liudmila Zudina[3], Anna Ulrich[5], Igor Pupko[3], Amélie Bonnefond[1,2,5], Zhanna Balkhiyarova[3,4,5], Marika Kaakinen[3,4,5], Philippe Froguel[1,2,5], Inga Prokopenko[1,2,3,4]

1. INSERM UMR 1283, CNRS UMR 8199, European Genomic Institute for Diabetes (EGID), Institut Pasteur de Lille, F-59000 Lille, France
2. University of Lille, Lille University Hospital, Lille, F-59000, France
3. Section of Statistical Multi-Omics, Department of Clinical and Experimental Medicine, School of Biosciences and Medicine, University of Surrey, Guildford, United Kingdom
4. People-Centred Artificial Intelligence Institute, University of Surrey, Guildford, United Kingdom
5. Department of Metabolism, Digestion and Reproduction, Imperial College London, London, United Kingdom

Corresponding authors:

Prof Philippe Froguel (p.froguel@ic.ac.uk)

Prof Inga Prokopenko (i.prokopenko@surrey.ac.uk)

Word count: (1,590/1,500)

Ref count (22/20)

Tables/Figures (1 table, 2 figures) (Max 3 tables and figures combined)

ABSTRACT

Obesity and type 2 diabetes (T2D) are associated with increased risk of pancreatic cancer. Here we assessed the relationship between pancreatic cancer and two distinct measures of obesity, namely total adiposity, using BMI, versus abdominal adiposity, using BMI adjusted waist-to-hip ratio (WHRadjBMI) by utilising polygenic scores (PGS) and Mendelian randomization (MR) analyses. We constructed z-score weighted PGS for BMI and WHRadjBMI using publicly available data and tested for their association with pancreatic cancer defined in UK biobank (UKBB). Using publicly available summary statistics we then performed bi-directional MR analyses between the two obesity traits and pancreatic cancer. PGS_{BMI} was significantly (multiple testing-corrected) associated with pancreatic cancer (OR[95%CI]=1.0804[1.025-1.14], *P*=0.0037). The significance of association declined after T2D adjustment (OR[95%CI]=1.073[1.018-1.13], *P*=0.00904). $PGS_{WHRadjBMI}$ association with pancreatic cancer was at the margin of statistical significance (OR[95%CI]=1.047[0.99-1.104], *P*=0.086). T2D adjustment effectively lost any suggestive association of $PGS_{WHRadjBMI}$ with pancreatic cancer (OR[95%CI]=1.039[0.99-1.097], *P*=0.14). MR analyses showed a nominally significant causal effect of WHRadjBMI on pancreatic cancer (OR[95%CI]=1.00095[1.00011-1.0018], *P*=0.027) but not for BMI on pancreatic cancer. Overall, we show that abdominal adiposity measured using WHRadjBMI, may be a more important causal risk factor for pancreatic cancer compared to total adiposity, with T2D being a potential driver of this relationship.

KEY WORDS: Pancreatic cancer, obesity, polygenic scores, type 2 diabetes, Mendelian randomization

INTRODUCTION

Pancreatic cancer is rare form of cancer, associated with poor prognosis and low survival rates(1). Furthermore, epidemiological evidence from observational studies suggests obesity and type 2 diabetes (T2D) are major risk factors for pancreatic cancer(2,3). Body mass index (BMI) and waist-to-hip ratio (WHR) are two common metrics used to assess total and abdominal adiposity. However, despite being a routine measure of adiposity in clinical and research settings, BMI is an imperfect measure of metabolic health. Alternatively, WHR represents abdominal adiposity which has a stronger correlation to the metabolic syndrome compared to total adiposity(4). To date, only 22 genome-wide significant signals are established in genome-wide association studies (GWAS) for pancreatic cancer(5). In contrast, more than 600 and 300 signals have been reported for BMI and WHR, respectively(6,7). These individual associations from GWAS, however, do not explain the shared co-morbidity between obesity and pancreatic cancer. Nevertheless, genomic loci identified in GWAS could be implemented in methods such as polygenic scores (PGS)(8) and Mendelian randomization (MR)(9). PGS can be used to define the shared genetic component between epidemiologically related phenotypes, while MR uses genetic variants as instruments to assess causality in relationships between phenotypes. In the present study, the impact of total and abdominal adiposity on pancreatic cancer risk was examined through PGS analyses, using publicly available GWAS of obesity traits data and information about pancreatic cancer within UK biobank. Moreover, using established genetic variants, we conducted a bi-directional MR between two adiposity traits and pancreatic cancer to assess the causal relationships between them.

MATERIALS AND METHODS

UK Biobank

The UK Biobank (UKBB) resource (www.ukbiobank.ac.uk) was used to define adiposity and cancer phenotypes for this study. We used the BMI data collected at the time of recruitment (UKBB field 21001). WHR data was computed by dividing waist circumference (UKBB field 48) by hip circumference (UKBB field 49) measured at baseline. BMI and WHR data were available for 457,270 individuals (**Supplementary Figure 1**). For pancreatic cancer, we used a combination of hospital admissions data, the tenth revision of the International Classification of Disease (ICD-10) codes and self-report data. Individuals with an ICD-10 code (code C25) and who self-reported to have a pancreatic cancer diagnosis (code 1026) were set as cases, while individuals with no cancer diagnosis were set as controls. In total, there were 1,416 cases and 455,854 controls (n=457,270) for pancreatic cancer. To limit confounding by ancestry, only individuals of European ancestry were included in our analyses (**Supplementary Methods, Supplementary Figure 1**).

UKBB GWAS

We performed single phenotype GWAS in UKBB using the BOLT-LMM software(10). BOLT-LMM applies a linear mixed model while age, sex, genotyping array and six principal components (PCs) were used as covariates for pancreatic cancer and BMI. BMI was an extra covariate in WHR GWAS to obtain WHRadjBMI analyses. The statistical threshold for genome-wide significant SNPs used was $P<5\times10^{-8}$.

Genetic correlation estimation

To estimate the genetic correlation (rG) between adiposity phenotypes (BMI/WHRadjBMI), T2D (**Supplementary Methods**) and pancreatic cancer in UKBB,

we used the linkage disequilibrium (LD) score (LDSC) regression approach and tool(11).

Polygenic scores

To construct BMI and WHRadjBMI PGS, we used risk increasing alleles at 567 and 274 SNPs respectively. The SNP list was obtained from recent large scale GWAS meta-analyses by GIANT consortium(6,7). However, as the target data for PGS analysis was the UKBB, which was part of the GIANT meta-analyses, we used weights from the study which did not include UKBB in the meta-analyses(12,13) (**Supplementary Figure 2**). We used the PLINK software(14) to generate the PGS. We used sex, age, genotyping array and six PCs as covariates in the regression model. As a sensitivity analyses, we ran a regression model with T2D as an extra covariate.

Mendelian Randomization

To assess causality between the two adiposity measures and pancreatic cancer, we performed bi-directional MR using the *TwoSampleMR* R package(15). We obtained the genetic instrument for BMI (566 SNPs) and WHRadjBMI (278 SNPs) from the GIANT consortium(6,7). The genetic instruments for pancreatic cancer (16 SNPs) were obtained from Klein et al(5). The causal effect estimate was derived from the inverse-variance weighted (IVW) method(16). The MR-Egger, simple mode, weighted mode and weighted median tests were used as sensitivity analyses(17). We excluded palindromic SNPs from the exposure-outcome pairs and matched alleles between summary statistics as part of the *TwoSampleMR* pipeline. Outliers were removed after inspection of scatter plots and leave-one-out results. Heterogeneity among the genetic instruments was evaluated using the Cochran's Q test.

RESULTS

UKBB GWAS and genetic correlation estimates

In UKBB GWAS, we identified 998, 1,014 and 4 significant independent SNPs at 901, 718, 4 loci for BMI, WHRadjBMI and pancreatic cancer respectively (**Figure 1**). The four *loci* identified for pancreatic cancer were *TERT*, *ABO, KLF* and *ZFP1* (**Figure 1C**) in line with recently published GWAS of pancreatic cancer(5). None of the obesity signals were shared with pancreatic cancer in the UKBB. However, 3 of the 22 established pancreatic cancer loci by Klein et al(5) were shared with WHRadjBMI in UKBB and had same direction of effect. These were *NR5A2*, *ETAA1* and *ZNRF3*. Conversely, only *ETAA1* from Klein et al(5) was shared with BMI in the UKBB. Additionally, there was positive genetic correlation between both obesity measures and pancreatic cancer, but the estimates did not meet statistical significance (rG_{BMI}=0.472, *P*=0.479, $rG_{WHRadjBMI}$=0.098, *P*=0.671) (**Supplementary table 1**). Similarly, the genetic correlation between T2D and pancreatic cancer in the UKBB was underpowered and did not meet statistical significance (rG=-0.0139, *P*=0.961) (**Supplementary Table 1**).

Effects of obesity variants on pancreatic cancer via polygenic scores

We identified a significant (Bonferroni multiple testing corrected *P*=0.05/2 tests=0.025) direct association between BMI PGS and pancreatic cancer (OR[95%CI]=1.0804[1.025-1.14], *P*=0.0037). We also identified a direct association between WHRadjBMI PGS and pancreatic cancer, however, this association was not statistically significant (OR[95%CI]=1.047[0.99-1.104], *P*=0.086) (**Table 1**). To determine if the association between adiposity PGS and pancreatic cancer was driven by T2D, we adjusted for T2D in the association tests. After T2D adjustment, the significance of the association for both BMI and WHRadjBMI PGS declined suggesting

that T2D could be acting via adiposity in pancreatic cancer risk (OR_{BMI_PGS}[95%CI]=1.073[1.018-1.13],*P*=0.00904); $OR_{WHRadjBMI\_PGS}$[95%CI]=1.039[0.99-1.097], *P*=0.14). Notably, the decline in association after T2D adjustment was more for WHRadjBMI PGS than BMI PGS (**Table 1**).

Causality results using Mendelian randomization

We report a causal effect of WHRadjBMI on pancreatic cancer at nominal significance (OR[95%CI]=1.00095[1.00011-1.0018], *P*=0.027) based on the IVW method, indicating a weak but positive causal effect estimate (**Figure 2**). However, none of the other MR tests for this direction were significant. The Cochran's Q test indicated absence of heterogeneity among the genetic instruments (Q_{IVW}=258.08, *P*=0.787). On the contrary, we have not identified any causal effect (Bonferroni *P*=0.05/4 tests=0.0125) of BMI on pancreatic cancer in either of the MR tests performed (**Supplementary Table 2**). There was no evidence of a causal effect from pancreatic cancer to WHRadjBMI (OR_{IVW}(*P*)=0.143(0.604). The results from pancreatic cancer to BMI were less informative with large standard errors despite nominal significance in some of the sensitivity MR tests ($OR_{WeightedMedian}$[95%CI]=58.105[3.997-844.69], *P*=0.003) (**Supplementary Table 2**).

DISCUSSION

In this study, using large scale datasets and a multi-method approach, we show that abdominal obesity assessed using WHRadjBMI is a causal risk factor for pancreatic cancer, in line with epidemiological evidence(18).

The mechanisms underlying the obesity-pancreatic cancer co-morbidity remain unclear. However, several factors such as inflammation, insulin resistance and hyperinsulinemia are potential mechanisms linking obesity to cancers including that of the pancreas(3,19). Notably, majority of these factors are hallmarks of the metabolic syndrome which correlate with abdominal obesity(20). Therefore, it is not surprising that our Mendelian randomization results show that WHRadjBMI rather than BMI is a more important causal risk factor for pancreatic cancer. Furthermore, the metabolic syndrome is considered a predictor of T2D(21). In our polygenic score analyses, we show that after adjusting for T2D status, the significance of the association declined modestly for PGS_{BMI} while any evidence of association in $PGS_{WHRadjBMI}$ on pancreatic cancer risk was effectively lost. Taken together, our polygenic analyses and Mendelian randomization suggest that the metabolic syndrome proxied by abdominal obesity may be a causal risk factor for pancreatic cancer. Additionally, obesity associated T2D(22) may be a potential cause driver of the metabolic syndrome underlying pancreatic cancer progression in obesity(3).

Several limitations in our present studies should be considered. Pancreatic cancer is a rare form of cancer characterised by low sample sizes as compared to other more common cancers. Consequently, there is less power in GWAS to identify genetic loci amenable for statistical analyses. Additionally, the causal effect identified in MR is only nominally significant and therefore interpretation of our findings should consider this. Future work will focus on validating our results in larger datasets, especially for

pancreatic cancer to improve statistical power of the analyses. Moreover, further analyses to properly control for T2D would be needed due to the complex relationship between obesity and T2D, more so in Mendelian randomization. Additional analyses to include components of the metabolic dysfunction such as fasting glucose levels will be part of future direction of this effort.

In conclusion, we show that abdominal adiposity measured through WHRadjBMI, may be a more important risk factor for pancreatic cancer, compared to total adiposity. Our results highlight the relationship between the metabolic syndrome component and a higher risk for pancreatic cancer, with T2D being a potential driver of this association. Furthermore, we demonstrate the importance and therefore encourage the assessment of diverse measures of obesity in clinical practice and research in the context of pancreatic cancer risk. Additionally, health care providers should emphasize the need for patients to monitor their visceral weight gain and not just overall weight gain to minimise the risk for pancreatic cancer.

DATA AVAILABILITY

The datasets generated and/or analysed during the current study are available from the corresponding author on reasonable request.

REFERENCES

1. Rahib L, Smith BD, Aizenberg R, Rosenzweig AB, Fleshman JM, Matrisian LM. Projecting cancer incidence and deaths to 2030: The unexpected burden of thyroid, liver, and pancreas cancers in the united states. Cancer Res [Internet]. 2014 Jun 1 [cited 2022 Jul 28];74(11):2913–21. Available from: https://aacrjournals.org/cancerres/article/74/11/2913/592763/Projecting-Cancer-Incidence-and-Deaths-to-2030-The

2. Xu M, Jung X, Hines OJ, Eibl G, Chen Y. Obesity and Pancreatic Cancer: Overview of Epidemiology and Potential Prevention by Weight Loss. Pancreas [Internet]. 2018 Feb 1 [cited 2022 Jul 3];47(2):158–62. Available from: https://journals.lww.com/pancreasjournal/Fulltext/2018/02000/Obesity_and_Pancreatic_Cancer__Overview_of.3.aspx

3. Froguel Philippe, Bonnefond Amelie. Does Type 2 diabetes increase the risk of developing cancer? Diabetes Management [Internet]. 2013 Nov [cited 2022 Aug 9];3(6):439–41. Available from: https://www.openaccessjournals.com/articles/does-type-2-diabetes-increase-the-risk-of-developing-cancer.pdf

4. Alberti KGMM, Eckel RH, Grundy SM, Zimmet PZ, Cleeman JI, Donato KA, et al. Harmonizing the Metabolic Syndrome. Circulation [Internet]. 2009 Oct 20 [cited 2022 Jul 5];120(16):1640–5. Available from: https://www.ahajournals.org/doi/abs/10.1161/CIRCULATIONAHA.109.192644

5. Klein AP, Wolpin BM, Risch HA, Stolzenberg-Solomon RZ, Mocci E, Zhang M, et al. Genome-wide meta-analysis identifies five new susceptibility loci for pancreatic cancer. Nature Communications 2018 9:1 [Internet]. 2018 Feb 8 [cited 2021 Oct 6];9(1):1–11. Available from: https://www.nature.com/articles/s41467-018-02942-5

6. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. Hum Mol Genet [Internet]. 2018 Oct 15 [cited 2019 Apr 28];27(20):3641–9. Available from: https://academic.oup.com/hmg/article/27/20/3641/5067845

7. Pulit SL, Stoneman C, Morris AP, Wood AR, Glastonbury CA, Tyrrell J, et al. Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. Hum Mol Genet [Internet]. 2019 Jan 1 [cited 2021 Oct 6];28(1):166–74. Available from: https://academic.oup.com/hmg/article/28/1/166/5098227

8. Lewis CM, Vassos E. Polygenic risk scores: From research tools to clinical instruments. Genome Med [Internet]. 2020 May 18 [cited 2022 Jul 3];12(1):1–11. Available from: https://genomemedicine.biomedcentral.com/articles/10.1186/s13073-020-00742-5

9. Lawlor DA, Harbord RM, Sterne JAC, Timpson N, Smith GD. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. Stat Med [Internet]. 2008 Apr 15 [cited 2022 Feb 18];27(8):1133–63. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/sim.3034

10. Loh PR, Tucker G, Bulik-Sullivan BK, Vilhjálmsson BJ, Finucane HK, Salem RM, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nature Genetics

2015 47:3 [Internet]. 2015 Feb 2 [cited 2022 Jul 28];47(3):284–90. Available from: https://www.nature.com/articles/ng.3190

11. Bulik-Sullivan B, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nature Genetics 2015 47:3 [Internet]. 2015 Feb 2 [cited 2022 Jul 3];47(3):291–5. Available from: https://www.nature.com/articles/ng.3211

12. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature [Internet]. 2015 Feb 12 [cited 2018 Nov 1];518(7538):197–206. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25673413

13. Shungin D, Winkler T, Croteau-Chonka DC, Ferreira T, Locke AE, Mägi R, et al. New genetic loci link adipose and insulin biology to body fat distribution. Nature 2015 518:7538 [Internet]. 2015 Feb 11 [cited 2022 Jul 3];518(7538):187–96. Available from: https://www.nature.com/articles/nature14132

14. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. The American Journal of Human Genetics. 2007;

15. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-base platform supports systematic causal inference across the human phenome. Elife. 2018 May 30;7.

16. Burgess S, Butterworth A, Thompson SG. Mendelian Randomization Analysis With Multiple Genetic Variants Using Summarized Data. Genet Epidemiol [Internet]. 2013 Nov 1 [cited 2022 Jul 3];37(7):658–65. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/gepi.21758

17. Bowden J, Holmes M v. Meta-analysis and Mendelian randomization: A review. Res Synth Methods [Internet]. 2019 Dec 1 [cited 2022 Dec 14];10(4):486–96. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/jrsm.1346

18. Xia B, He Q, Pan Y, Gao F, Liu A, Tang Y, et al. Metabolic syndrome and risk of pancreatic cancer: A population-based prospective cohort study. Int J Cancer [Internet]. 2020 Dec 15 [cited 2022 Jul 28];147(12):3384–93. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/ijc.33172

19. Scully T, Ettela A, LeRoith D, Gallagher EJ. Obesity, Type 2 Diabetes, and Cancer Risk. Front Oncol. 2021 Feb 2;10:3196.

20. Lalia AZ, Dasari S, Johnson ML, Robinson MM, Konopka AR, DIstelmaier K, et al. Predictors of Whole-Body Insulin Sensitivity Across Ages and Adiposity in Adult Humans. J Clin Endocrinol Metab [Internet]. 2016 Feb 1 [cited 2022 Jul 1];101(2):626–34. Available from: https://academic.oup.com/jcem/article/101/2/626/2810955

21. Shin JA, Lee JH, Lim SY, Ha HS, Kwon HS, Park YM, et al. Metabolic syndrome as a predictor of type 2 diabetes, and its clinical interpretations and usefulness. J Diabetes Investig [Internet]. 2013 Jul [cited 2022 Jul 28];4(4):334–43. Available from: https://pubmed.ncbi.nlm.nih.gov/24843675/

22. Fall T, Hägg S, Mägi R, Ploner A, Fischer K, Horikoshi M, et al. The Role of Adiposity in Cardiometabolic Traits: A Mendelian Randomization Analysis. PLoS Med [Internet]. 2013

[cited 2022 Aug 9];10(6):e1001474. Available from:
https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001474

286 ACKNOWLEDGEMENTS

287 We would like to thank the Genetic Investigation of Anthropometric Traits (GIANT)
288 consortium for enabling access to their GWAS summary statistics which we have used
289 in our PGS and MR analyses. We would also like to appreciate the participants and
290 staff of the UKBB for their valuable contributions. This work was conducted under the
291 UKBB application number 37685.

ETHICAL APPROVING AND FUNDING

This research has been conducted using the UK Biobank resource application number 37685. This research was in part funded by the World Cancer Research Fund (WCRF UK) and World Cancer Research Fund International (2017/1641), the Diabetes UK (BDA number: 20/0006307), the European Union's Horizon 2020 research and innovation programme (LONGITOOLS, H2020-SCI-2019-874739), Agence Nationale de la Recherche (PreciDIAB, ANR-18-IBHU-0001), by the European Union through the "Fonds Européen de Dévelopment Regional" (FEDER), by the "Conseil Régional des Hauts-de-France" (Hauts-de-France Regional Council) and by the "Métropole Européenne de Lille" (MEL, European Metropolis of Lille)

COMPETING INTEREST

The authors report no competing interests.

FIGURE LEGENDS

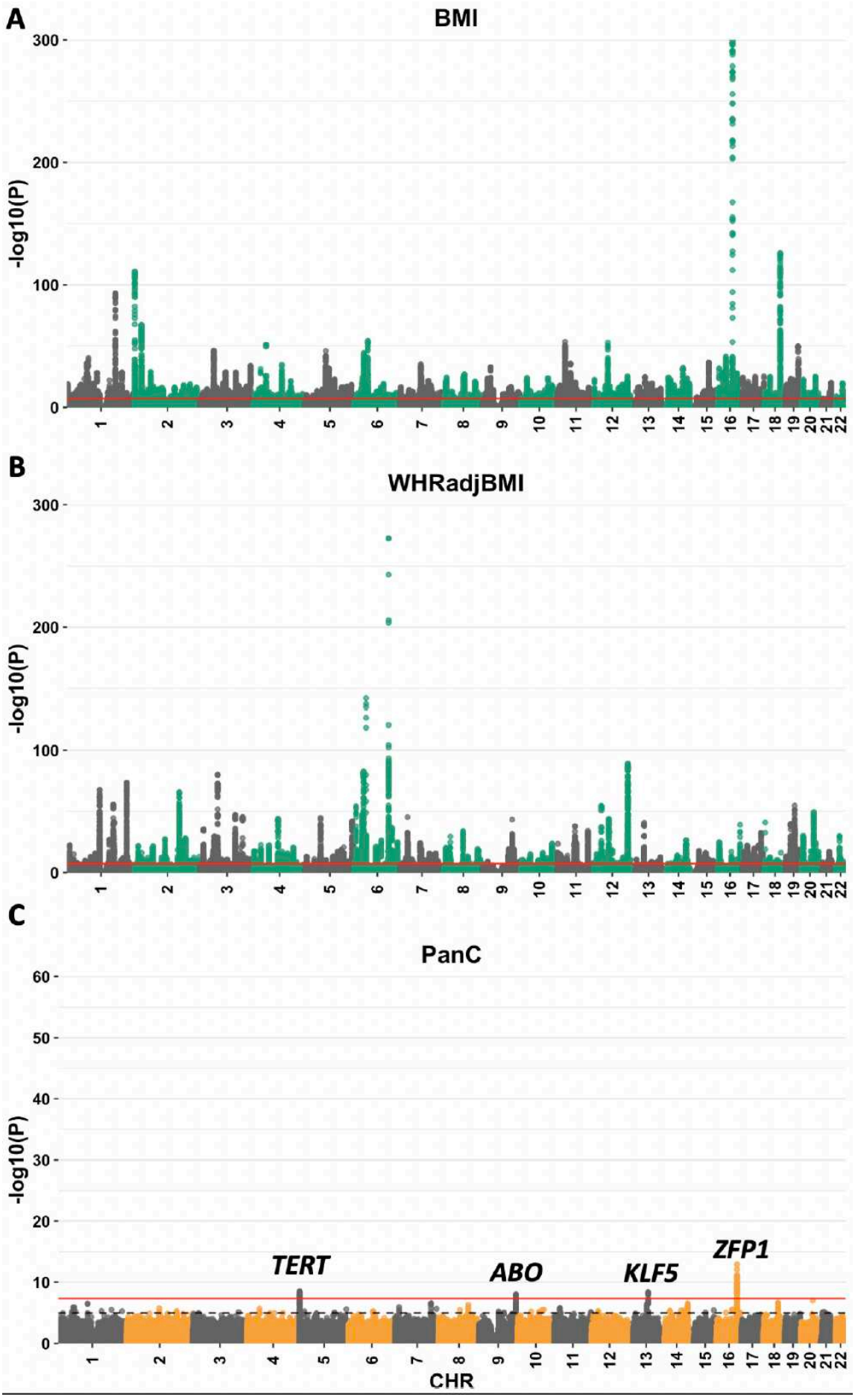
Figure 1. Manhattan plots of (A) BMI, (B) WHRadjBMI and (C) pancreatic cancer GWAS in UK Biobank. The red horizontal line shows genome-wide significance threshold ($P<5x10^{-8}$). The dashed grey line shows suggestive significance threshold ($P<1x10^{-5}$)

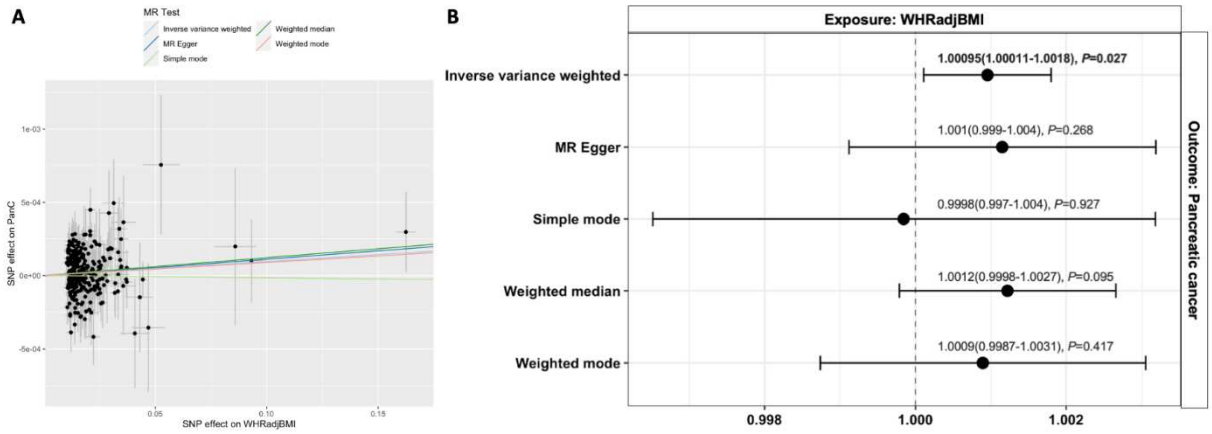
Figure 2. (A) Scatter and (B) forest plots for the WHRadjBMI to pancreatic cancer MR test. The scatter plot includes the intercepts of the various MR methods used while the odds ratio plot shows the MR effect estimate for each MR method used

Table 1. Association between adiposity polygenic scores and pancreatic cancer

	Unadjusted model		T2D adjusted model	
Adiposity trait	**OR (95%CI)**	***P***	**OR (95%CI)**	***P***
BMI	1.0804 (1.025-1.14)	0.0037	1.073 (1.018-1.13)	0.00904
WHRadjBMI	1.047 (0.99-1.104)	0.086	1.039 (0.99-1.097)	0.14

Legend: OR(95%CI)=Odds ratio of association and the lower and upper 95% confidence intervals (CI)


A
BMI
-log10(P)
300
200
100
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
B
WHRadjBMI
-log10(P)
300
200
100
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
C
PanC
-log10(P)
60
50
40
30
20
10
0
TERT
ABO
KLF5
ZFP1
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
CHR


A
MR Test
Inverse variance weighted
MR Egger
Simple mode
Weighted median
Weighted mode
SNP effect on PanC
SNP effect on WHRadjBMI
B
Exposure: WHRadjBMI
Outcome: Pancreatic cancer
Inverse variance weighted
MR Egger
Simple mode
Weighted median
Weighted mode
1.00095(1.00011-1.0018), P=0.027
1.001(0.999-1.004), P=0.268
0.9998(0.997-1.004), P=0.927
1.0012(0.9998-1.0027), P=0.095
1.0009(0.9987-1.0031), P=0.417
0.998
1.000
1.002

4.1 Supplementary data

Pancreatic cancer definition in UKBB

Pancreatic cancer in UKBB was defined using a combination of the tenth revision of the International Classification of Disease (ICD-10) codes and self-report data. Additionally, hospital admissions data, recently made available to researchers by UKBB, were used to supplement the number of cases. Individuals with an ICD-10 code (C25) and those who self-reported to have a pancreatic cancer diagnosis (code 1026) were set as cases. In total, there were 629 cases and 458,987 controls of European ancestry for pancreatic cancers after exclusions (**Supplementary Figure 1**). 1,340 European pancreatic cancer cases were defined from hospital admissions data (ICD-10 code C25). 544 of these cases were shared with the 629 cases defined using ICD-10 and self-report data only. 85 cases (self-reported) from the 629 cases defined earlier were added to the 1340 hospital admissions cases. 796 controls which had case status in the hospital admissions data were excluded from controls. In total, after all exclusions were applied, there were 1,416 cases and 455,854 controls of European ancestry for pancreatic cancer (**Supplementary Figure 1**).

Type 2 diabetes definition in UKBB

To determine the role of type 2 diabetes (T2D) in the relationship between obesity and pancreatic cancer, we sought to first define the genetic correlation between T2D and pancreatic cancer. Secondly, we included T2D as an additional covariate in our polygenic scores (PGS) analyses. A T2D case in UKBB was defined if a participant self-reported a diabetes diagnosis made by a doctor, were on insulin medication one-year post-diagnosis and were at least 40 years old by the time the diagnosis was made.

T2D controls included individuals who did not meet the case criterion. From both cases and controls, we excluded individuals with gestational diabetes (UKBB field 4041, code=1), individuals on insulin medication within the first year of diagnosis (UKBB field 2986) and individuals who were younger than 40 years old at the time of diagnosis (UKBB field 2976). In total, we had 19,344 cases and 463,641 controls of European ancestry.

Genetic correlation estimation in UKBB

We used the LDSC regression tool[89] to estimate the genetic correlation between BMI, WHRadjBMI and pancreatic cancer in UKBB. UKBB GWAS summary statistics were filtered based on the following parameters: imputation score > 0.9, minor allele frequency (MAF) > 0.01 and 0.1 ≥ P-value > 0. Strand ambiguous, duplicated SNPs and variants that did not represent SNPs (e.g., indels) were removed. The Bonferroni corrected significance threshold to determine significant genetic correlation estimates was set as P<0.025 (0.05/2, the number of genetic correlation tests done in our analyses; one for BMI and one for WHRadjBMI). Nominal significance threshold was set at 0.05 ≥ P > 0.025.

Table 2. Genetic correlation between adiposity measures, type 2 diabetes and pancreatic cancer in UKBB

	BMI			WHRadjBMI			Type 2 diabetes		
Cancer	**rG (SE)**	**Z score**	**P**	**rG (SE)**	**Z score**	**P**	**rG (SE)**	**Z Score**	**P**
Pancreatic	0.472 (0.667)	0.708	0.479	0.098 (0.230)	0.425	0.671	-0.0139 (0.287)	-0.0484	0.961

Legend: rG(SE)=genetic correlation estimate and the standard error, Z score=rG/SE

Polygenic score analyses

The SNP lists for BMI and WHRadjBMI were obtained from GIANT consortium's meta-analyses[59,61]. The meta-analyses included previous GIANT studies[60,62] and UKBB. Since we used the UKBB as the target data (testing cohort for our PGS), we use the weights from the studies that did not include UKBB in the meta-analyses. The workflow for PGS analyses is shown in **Figure 7**.

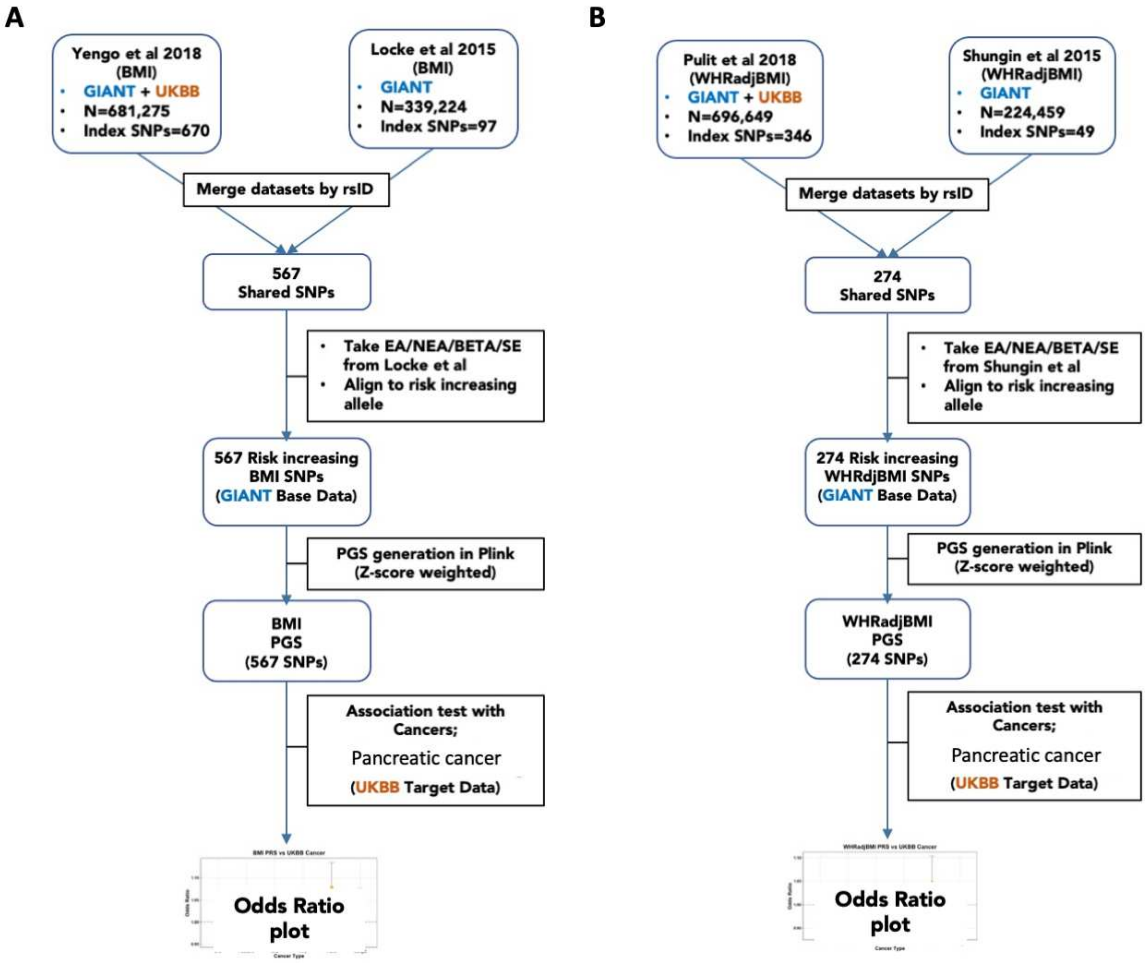


Figure 7. (A) BMI and (B) WHRadjBMI polygenic score analyses pipeline. A two-sample approach was used to construct our PGS base data. The SNPs used for the PGS were from GIANT's latest BMI and WHRadjBMI meta-analyses. Since the meta-analyses comprised of the UK Biobank (our target data), we used weights from the non-UK Biobank study in GIANT's meta-analyses.

Mendelian randomization (MR)

We assessed the causal relationships between BMI, WHRadjBMI and pancreatic cancer using bi-directional MR. The *TwoSampleMR* R package[86] was used for this analysis. We tested the effect of obesity (BMI and WHRadjBMI) as an exposure for pancreatic cancer (outcome), and the reverse direction with pancreatic cancer as a risk factor for obesity (BMI and WHRadjBMI) using summary statistics from independent datasets. The genetic instruments for BMI (670 SNPs) and WHRadjBMI (346 SNPs) were obtained from GIANT consortium[59,61]. Additionally, the pancreatic cancer genetic instruments (22 SNPs) were obtained from a recent large-scale meta-analysis by Klein et al[67].

Table 3. Detailed results of the Mendelian randomization results between adiposity phenotypes and pancreatic cancer

Exposure	Outcome	MR Method	NSNPs	OR(95%CI)	P value	Q statistic (P value)
BMI	PanC	MR Egger	566	0.999 (0.997-1.001)	0.389	500.41 (0.974)
BMI	PanC	Weighted median	566	1.000 (0.999-1.002)	0.561	NA
BMI	PanC	Inverse variance weighted	566	1.001 (1.000-1.001)	0.090	502.99 (0.971)
BMI	PanC	Simple mode	566	1.001 (0.997-1.004)	0.714	NA
BMI	PanC	Weighted mode	566	1.000 (0.998-1.002)	0.802	NA
WHRadjBMI	PanC	MR Egger	278	1.001 (0.999-1.0032)	0.268	258.035 (0.774)
WHRadjBMI	PanC	Weighted median	278	1.0012 (0.9998-1.0027)	0.095	NA
WHRadjBMI	PanC	Inverse variance weighted	278	1.00095 (1.00011-1.0018)	0.027	258.078 (0.787)
WHRadjBMI	PanC	Simple mode	278	0.9998 (0.997-1.0032)	0.927	NA
WHRadjBMI	PanC	Weighted mode	278	1.0009 (0.9987-1.0031)	0.417	NA
PanC	BMI	MR Egger	16	0.444 (0.000-11023.55)	0.877	99.368 (6.27x10-15)
PanC	BMI	Weighted median	16	58.105 (3.997-844.69)	0.003	NA
PanC	BMI	Inverse variance weighted	16	58.526 (0.301-11367.20)	0.130	108.025 (3.86x10-16)
PanC	BMI	Simple mode	16	70.019 (3.66-1341.18)	0.013	NA
PanC	BMI	Weighted mode	16	91.921 (7.73-1092.50)	0.003	NA
PanC	WHRadjBMI	MR Egger	16	21.142 (0.00-64574354.85)	0.695	32.171 (3.79x10E-03)
PanC	WHRadjBMI	Weighted median	16	0.070 (0.000-74.001)	0.454	NA
PanC	WHRadjBMI	Inverse variance weighted	16	0.143 (0.000-222.403)	0.604	33.487 (4.018x10-03)
PanC	WHRadjBMI	Simple mode	16	1.057 (0.000-11211.818)	0.991	NA
PanC	WHRadjBMI	Weighted mode	16	0.137 (0.000-99.21)	0.563	NA

Legend: PanC=pancreatic cancer, NSNPs=number of SNPs/genetic instruments used to estimate causality, OR(95%CI) =Odds ratio and the lower and upper 95% confidence intervals (CI).

4.2 Insights

In this article prepared in brief communication format for EJHG, I presented the results dissecting the relationship between two obesity phenotypes and pancreatic cancer. I showed that central/abdominal obesity is potentially a more important causal risk factor pancreatic cancer than overall obesity. Additionally, we show that after adjusting for T2D in our polygenic scores analyses, the association between central obesity and pancreatic cancer was lost. This suggests that T2D could be the driver of the association between the metabolic syndrome and pancreatic cancer.

The findings presented in this article provide evidence of the need to stratify obesity into discrete categories when assessing the contribution of obesity in the risk of pancreatic cancer. Both in research and clinical contexts.

Several limitations that have been highlighted in this study should also be taken into perspective when interpreting our findings. Future studies will focus on utilising larger sample sizes for pancreatic cancer in order to improve statistical power. Additionally, as more GWAS studies consortia make their summary statistics public, the instruments variables for pancreatic cancer will get stronger and provide an opportunity to validate our findings.

Overall, the study presented suggests that central obesity independent of BMI is associated with the risk of pancreatic cancer with T2D possibly driving this association. Therefore, there is need to maintain a healthy weight and minimising the risk of T2D as the two factors may predispose an individual to pancreatic cancer. Future work leveraging on larger samples sizes for pancreatic cancer is needed to confirm our findings.

5 SECOND ARTICLE

“Genetic relationships and causality between overall and central adiposity and breast, prostate, lung, and colorectal cancer” (Under review at Obesity journal. Preprint link https://doi.org/10.1101/2022.12.19.22283607)

Genetic relationships and causality between overall and central adiposity and breast, prostate, lung and colorectal cancer

Jared G Maina[1,2], Vincent Pascat[1,2,3], Liudmila Zudina[3], Anna Ulrich[4], Igor Pupko[3], Ayse Demirkan[3,5], Amélie Bonnefond[1,2,4], Zhanna Balkhiyarova[3,4,5], Marika Kaakinen[3,4,5], Philippe Froguel[1,2,4], Inga Prokopenko[1,2,3,5]

1. INSERM UMR 1283, CNRS UMR 8199, European Genomic Institute for Diabetes (EGID), Institut Pasteur de Lille, F-59000 Lille, France
2. University of Lille, Lille University Hospital, Lille, F-59000, France
3. Section of Statistical Multi-Omics, Department of Clinical and Experimental Medicine, University of Surrey, Guildford, United Kingdom
4. Department of Metabolism Digestion and Reproduction, Imperial College London, London United Kingdom
5. People-Centred Artificial Intelligence Institute, University of Surrey, Guildford, United Kingdom

Corresponding authors:

Prof. Inga Prokopenko

Section of Statistical Multi-Omics,

Department of Clinical and Experimental Medicine, University of Surrey

Leggett Building, Daphne Jackson Road, Manor Campus,

Guildford, Surrey, UK, GU2 7WG

Email: i.prokopenko@surrey.ac.uk

Prof. Philippe Froguel

Genomics of Common Diseases,

Department of Metabolism, Digestion and Reproduction

Hammersmith Hospital Campus, Burlington-Danes Building, Du Cane Road

London, UK, W12 0NN

Email: p.froguel.imperial.ac.uk

RUNNING TITLE: Overall and central obesity in cancer risk (44/50 characters including spaces):

WORD COUNT: (3181/4000)

KEYWORDS: (4/5) Obesity, Cancer, Mendelian randomization, polygenic scores

FUNDING: This research has been conducted using the UK Biobank Resource under Application Number 37685. This research was in part funded by the World Cancer Research Fund (WCRF UK) and World Cancer Research Fund International (2017/1641), the Diabetes UK (BDA number: 20/0006307), the European Union's Horizon 2020 research and innovation programme (LONGITOOLS, H2020-SCI-2019-874739), Agence Nationale de la Recherche (PreciDIAB, ANR-18-IBHU-00001), by the European Union through the 'Fonds Européen de Dévelopment Regional" (FEDER), by the "Counseil Régional des Hauts-de-France" (Hauts-de-France Regional Council) and by the "Métropole Européenne de Lille" (MEL, European Metropolis of Lille).

DISCLOSURE: The authors declared no conflict of interest

STUDY IMPORTANCE (*up to 3 bullet points answers for each of the questions below*)

What is already known about this subject?

- Observational studies suggest obesity is associated with higher risk of certain cancers and at the same time is protective of other cancers. The direction of association is in part influenced by the anthropometric trait used to assess obesity.
- Higher BMI appears protective from prostate, breast and lung cancers but is a risk factor for post-menopausal breast, pancreatic and colorectal cancers.

What are the new findings in your manuscript?

- We implement Mendelian randomization approach using large scale datasets and show a protective causal effect of higher BMI from prostate cancer but suggest that higher WHRadjBMI is causal for prostate cancer.
- We also show nominal evidence of WHRadjBMI being causally protective from breast cancer.

How might your results change the direction of research or the focus of clinical practice?

- We demonstrate the importance of partitioning obesity into discrete types depending on the area of fat deposition rather than using an overall measure.
- Our results show that diverse measures of obesity relate differently to cancer risk. In fact, even for the same type of cancer, overall and central obesity measures may impact in opposite direction in terms of risk to cancer.

ABSTRACT

OBJECTIVE. Diverse measures of obesity relate to cancer risk differently. Here we assess the relationship between overall and central adiposity and cancer.

METHODS. We constructed z-score weighted polygenic scores (PGS) for two obesity-related phenotypes; body mass index (BMI) and BMI adjusted waist-to-hip ratio (WHRadjBMI) and tested for their association with five cancers in the UK Biobank: overall breast (BrC), post-menopausal breast (PostBrC), prostate (PrC), colorectal (CrC) and lung (LungC) cancer. We utilised publicly available data to perform bi-directional Mendelian randomization (MR) between BMI/WHRadjBMI and BrC, PrC and CrC.

RESULTS. PGS_{BMI} had significant multiple testing-corrected inverse association with PrC (OR[95%CI]=0.97[0.95-0.99], P=0.0012) but $PGS_{WHRadjBMI}$ was not associated with PrC. PGS_{BMI} was associated with PostBrC (OR[95%CI]=0.97[0.96-0.99], P=0.00203) while $PGS_{WHRadjBMI}$ had nominal association with BrC. PGS_{BMI} had nominal positive association with LungC. MR analyses showed significant multiple testing-corrected protective causal effect of BMI on PrC (OR[95%CI]=0.993[0.988-0.998], $P=4.19\times10^{-3}$). WHRadjBMI had a nominal causal effect on higher PrC risk (OR[95%CI]=1.022[1.0067-1.038], P=0.0053). We also report nominal causal protective effect of WHRadjBMI on breast cancer (OR[95%CI]=0.99[0.98-0.997], P=0.0068). Neither PGS nor MR analyses were significant for CrC.

CONCLUSIONS. Higher overall adiposity appears protective from PrC while higher central adiposity is a potential risk factor for PrC but protective from BrC.

INTRODUCTION

Two common anthropometric measures used to define obesity are body mass index (BMI) and waist-to-hip circumference ratio (WHR). They represent the body's overall and central abdominal adiposity, respectively. Despite being a routine measure of adiposity, BMI does not accurately capture well body composition as it does not distinguish lean mass from fat mass(1). Additionally, individuals who are within normal BMI ranges may be metabolically unhealthy(2), and others with elevated BMI present normal metabolic parameters. Other measures of adiposity have hence been used to improve clinical evaluations of metabolic health, including WHR. High WHR is associated with insulin resistance and contributes to the definition of the metabolic syndrome(3). Epidemiological evidence suggests BMI is positively associated with increased cancer risk of post-menopausal breast cancer(4), pancreatic cancer(5), colorectal cancer(6), while inversely associated with prostate cancer(7), pre-menopausal breast cancer and lung and oral cavity cancers(4). The relationship between cancer incidence, its mortality, and WHR is relatively understudied compared to that with BMI. Despite this, WHR appears to be a better predictor of cancer risk compared to BMI(8,9). However, correlation observed in epidemiological studies does not infer causality. Moreover, observational studies are prone to bias by unadjusted confounders such as tobacco use in lung cancer studies(4). To date, genome-wide association studies (GWAS) have identified hundreds of genomic loci, associated with BMI and body fat(10,11) and many loci with WHR adjusted for BMI (WHRadjBMI)(11) with a relatively small overlap between the two sets of associated loci. GWAS have also found many loci associated with cancer phenotypes(12–15). However, one limitation of GWAS is that these loci identified individually for adiposity and cancer phenotypes offer little evidence of the shared pathophysiology between

them. Despite the above-mentioned limitation of GWAS, genomic loci identified in GWAS can be implemented in methods such as polygenic scores (PGS)(16) and Mendelian randomization (MR)(17). PGS refer to the weighted (by effect size) sum of risk variants identified from GWAS that an individual has for a particular phenotype(16). PGS are useful for risk prediction analyses as well as association testing. In MR analyses, genetic variants identified in GWAS, referred to as instrument variables, are used to estimate the causal effect of a risk factor (exposure) on an outcome of interest(17).

In this study, we aimed to assess the impact of overall and central adiposity on cancer risk. First, we defined the genetic correlation between BMI/WHRadjBMI and cancers using the UK Biobank (UKBB) dataset. Additionally, we used established BMI and WHRadjBMI genome-wide loci(10,11) (P<5x10^{-8}) to create PGS which were then tested for association with five cancer phenotypes in UKBB including overall breast, post-menopausal breast, colorectal, prostate and lung cancer. Further, using established genetic variants, associated with these phenotypes, we performed MR between the two adiposity traits and three cancers (breast, prostate and colorectal) to investigate the causal relationships between them. The MR analyses did not include post-menopausal breast and lung cancer due to the unavailability of summary statistics.

METHODS

UK Biobank

We used the UKBB resource (www.ukbiobank.ac.uk) to define adiposity and cancer phenotypes. The UKBB includes approximately 500,000 individuals, aged between 40-69 years recruited from 22 centres across the United Kingdom. Phenotypic data collected from recruited participants includes biological samples, physical measurements and responses from a questionnaire administered at recruitment. Genetic data is available for 488,377 individuals who were genotyped using the UKBB BiLEVE array (n=49,979) and the UK Biobank Axiom Array (n=438,398)(18). 457,270 individuals of European ancestry had weight (kg), height (m), waist and hip circumference (cm) measurements which were used to define the BMI (weight/[height]^2 and WHR (waist/hip) phenotypes. Cancer phenotypes were defined using the criteria described in **Table S1**. There were 18,676 overall breast cancer, 13,355 postmenopausal breast cancer, 11,825 prostate cancer, 8,201 colorectal cancer and 4,237 lung cancer cases (**Table S2**).

Genetic correlation and SNP heritability estimation

We used the LD Score (LDSC) regression approach and tool(19) to estimate the genetic correlation (rG) between adiposity phenotypes and cancer in the UKBB. The proportion of genetic variance explained by genome-wide SNPs (h^2_{SNP}) for each of our UKBB phenotypes was also computed using the LDSC tool. UKBB GWAS summary statistics were filtered based on the following parameters: imputation score > 0.9, minor allele frequency (MAF) > 0.01 and $0.1 \geq P > 0$. Ambiguous strand, duplicated SNPs and variants that did not represent SNPs (e.g., indels) were removed. The Bonferroni corrected significance threshold to determine significant genetic correlation estimates was set as P<0.005 (0.05/10, the number of genetic

correlation tests done in our analyses; five for BMI and five for WHRadjBMI). Nominal significance threshold was set at $0.05 \geq P > 0.005$.

Polygenic scores

We used the adiposity-increasing alleles at genome-wide significant variants from recent large scale meta-analyses of BMI (670 SNPs) and WHRadjBMI (346 SNPs) to construct PGSs. The SNP sets for BMI and WHRadjBMI PGSs base data were derived from the latest adiposity GWAS meta-analyses from the GIANT consortium(10,11). These studies meta-analysed UKBB with previous BMI and WHRadjBMI studies(20,21). Since the target data (UKBB) was part of the adiposity meta-analyses which formed the PGS base data in this analysis, we used variant effect size estimates from earlier studies which has not included UKBB in the meta-analyses for the weighted PGS(20,21). In total, there 567 and 274 SNPs for BMI and WHRadjBMI PGS analyses (**Figure 1**). The PGSs were generated using PLINK version 1.90b4.1(22). The PGSs were tested for association with five cancers in UKBB using logistic regression models in RStudio(23). For sex-specific cancers (breast and prostate), we included age, batch array and six principal components (PCs) as covariates in the regression model. For all the other cancers, sex was included as an additional covariate. Additionally, we performed BMI PGS association tests by splitting UKBB dataset into four BMI categories for each of the cancers (underweight [BMI < 18.5kg/m^2], normal weight [25kg/m^2 ≥ BMI > 18.5kg/m^2], overweight [30kg/m^2 ≥ BMI > 25kg/m^2], obese [BMI > 30kg/m^2]). The Bonferroni corrected significance threshold (Bonferroni^{PGS}) to determine significant associations was set as $P<0.005$ (0.05/10, the number of association tests done in our analyses; five for BMI and five for WHRadjBMI). Nominal significance threshold was set at $0.05 \geq P > 0.005$.

Mendelian Randomization

To investigate the causality between BMI/WHRadjBMI and cancer, we performed bi-directional MR using the *TwoSampleMR* R package version 0.5.6(24). We tested the effect of adiposity (BMI and WHRadjBMI) as an exposure for cancer outcomes, and the reverse direction with cancers as risk factors to adiposity (BMI and WHRadjBMI) using summary statistics from independent datasets. Independent, genome-wide significant SNPs ($P<5\times10^{-8}$) were used as genetic instruments for MR in our analyses. We obtained the genetic instruments for BMI (670 SNPs) and WHRadjBMI (346 SNPs) from the GIANT consortium(10,11). Furthermore, the cancer genetic instruments were derived from recent large-scale GWAS of breast (201 SNPs)(12), colorectal (137 SNPs)(14) and prostate (248 SNPs)(13) cancers using genome-wide significant SNPs. The number of SNPs available for each MR test are summarised in **Figure S1**. Moreover, the causal effect estimate was obtained using the inverse-variance weighted (MR-IVW) method(25), which combines the ratio estimates of individual variants using a random-effect meta-analysis. Sensitivity MR analyses were performed using MR Egger, weighted median, weighted mode and simple mode methods(26). Exclusion of palindromic SNPs from the exposure-outcome pairs, as well as allele matching was performed as part of the *TwoSampleMR* pipeline. Additional quality control steps included removal of outliers after inspection of scatter plots and leave-one-out results also performed using the MR package. We also performed the Cochran's Q test as part of the MR pipeline to assess heterogeneity on our instruments. The Bonferroni corrected significance threshold (Bonferroni^{MR}) to determine significant associations was set as $P<0.0042$ (0.05/12, the number of MR tests done in our analyses; **Figure S1**). Nominal significance threshold was set at $0.05 \geq P > 0.00442$.

RESULTS

Definition of overlap of the loci between adiposity and cancer phenotypes in the UKBB

We assessed the overlap of genome-wide significant loci between the cancers and adiposity phenotypes using the UKBB GWAS summary statistics (**Supplementary data**). There was no loci overlap between adiposity phenotypes and either colorectal and lung cancers. However, we observed overlap between the sex-specific cancers (BrC, PostBrC and PrC) and both BMI and WHRadjBMI. There were five shared loci between BMI and BrC in UKBB (**Figure 2a**): *FTO, EBF1*, *ERBB4*, *TBX3/MED13L* and *CASC16*. In contrast, there was no overlap between PostBrC and BMI. Furthermore, three loci were shared between BMI and PrC (**Figure 2a**): *TMEM17/EHBP1*, *JAZF1* and *MIR4686/ASCL2*. In relation to WHRadjBMI, eight loci were shared between BrC (**Figure 2b**): *ZMIZ1, NRIP1/USP25, EBF1, ESR1, RAD51B, CASC21/CASC8, CCND1* and *FGFR2*. Five loci were shared between PostBrC and WHRadjBMI (*EBF1, ESR1, RAD51B, CCND1, CASC21/CASC8*). Further, 11 loci were shared between PrC and WHRadjBMI (**Figure 2b**): *THADA, HLA-DQB1-AS1, SLC22A3, JAZF1, PRNCR1/CASC19, CASC8, MYEOV, CASC17, FGFR2, CCND1, CASC21/CASC8*.

Genetic correlation and SNP heritability estimates

The UKBB genetic correlation analyses identified nominally significant inverse genetic correlation between BMI and prostate cancer (rG=-0.076, *P*=0.0075). Both BMI and WHRadjBMI had a nominally significant positive genetic correlation with lung cancer in UKBB (rG_{BMI}=0.18, *P*=0.0014, $rG_{WHRadjBMI}$=0.16, *P*=0.0065). Nominal inverse genetic correlation was also observed between BMI and post-menopausal breast cancer (rG=-0.0803, *P*=0.014), whereas positive genetic correlation was

observed between WHRadjBMI and colorectal cancer (rG=0.103, *P*=0.017) (**Figure 3a, Table S3**).

The observed SNP heritability estimates (h^2_{SNP}) for the BMI and WHRadjBMI were 24.59% and 13.43% respectively. Estimates for cancer ranged from 0.36% (lung) to 4.41% (prostate) (**Figure 3b** and **Table S4**).

Association analyses with polygenic scores

BMI PGS (567 SNPs) had a significant, after correction for multiple testing, inverse association with prostate (OR[95%CI]=0.97[0.95-0.99], *P*=0.0012) and post-menopausal breast (OR[95%CI]=0.97[0.96-0.99], *P*=0.00203) cancers (**Figure 4, Table 1**). We also identified nominal associations with overall breast cancer (OR[95%CI]=0.98[0.96-0.99], *P*=0.0086) and lung cancer (OR[95%CI]=1.044[1.013-1.076], *P*=0.0057).

WHRadjBMI PGS (274 SNPs) was not significantly associated with any of the cancers. However, a nominal association was identified with overall breast cancer (OR[95%CI]=0.98[0.97-0.99], *P*=0.021). WHRadjBMI PGS had a trend towards positive association with prostate cancer (OR[95%CI]=1.018[0.99-1.037, *P*=0.074) and colorectal cancer (OR[95%CI]=1.021[0.999-1.044], *P*=0.062) (**Figure 4, Table 1**). The direct relationship detected through suggestive association between WHRadjBMI and prostate cancer is noteworthy as it is contrary to the BMI PGS results. Additionally, we calculated the associations between BMI PGS and cancer while stratifying the data by BMI categories. On average, the strength of associations for BMI PGS was higher for overweight and obese individuals (BMI higher than 25 kg/m2) (**Table S5**). We then used smoking status as a proxy for tobacco use in the lung cancer association tests, but we did not identify any significant association among current and previous smokers (**Table S6**). However, there was a nominal

inverse association between WHRadjBMI PGS and lung cancer risk among individuals who had never smoked (OR[95%CI]=0.92[0.85-1.00], *P*=0.046).

Causality using Mendelian randomization

In this study we investigated the bi-directional causal relationship between two adiposity phenotypes and three cancers. Our results demonstrate a significant protective causal effect of BMI on prostate cancer (OR_{IVW}[95%CI]=0.993[0.988-0.998], *P*=4.19x10^{-3}, 574 SNPs) (**Table 2**). Additionally, the sensitivity analysis done using the weighted median (W.Med) MR method agreed with the MR-IVW results ($OR_{W.Med}$[95%CI]=0.993[0.985-0.9996], *P*=0.039, 574 SNPs). The scatter plot and odds ratio plots are shown in **Figure 5a**. The MR Egger intercept test suggested no evidence of pleiotropy among the selected SNPs in the BMI to prostate cancer test (Egger intercept =-46.47, *P*=0.713). The Cochran's Q test indicated significant heterogeneity in this association (Q_{IVW}=863.64, *P*=3.93×10^{-14}) (**Table S7**). However, in the reverse direction from prostate cancer to BMI, there was no evidence of causality. All other BMI to cancer direction tests (and their reverse) did not show evidence of causality (**Figure S2 and S3, Table S7**).

In contrast, the WHRadjBMI to cancer causal test suggested a nominal causal risk effect of WHRadjBMI on prostate cancer risk (**Table 2**) based on the two MR tests used as sensitivity analyses ($OR_{MR\text{-}Egger}$[95%CI]=1.016[1.00018-1.032], *P*=0.048, 284 SNPs; $OR_{W.Mode}$[95%CI]=1.022[1.0067-1.038, *P*=0.0053, 284 SNPs (**Figure 5b, Table 2**). Additionally, there was a nominal causal protective effect of WHRadjBMI on overall breast cancer based on the MR-IVW test (OR_{IVW}[95%CI]=0.99[0.98-0.997], *P*=0.0068, 284 SNPs) (**Table 2, Table S7**). However, the sensitivity tests were insignificant. None of the other WHRadjBMI to cancer direction tests (and their reverse) were significant (**Figure S4 and S5, Table S7**).

DISCUSSION

In this study, we provide evidence that higher overall adiposity, measured with BMI, may confer men a protective advantage against prostate cancer, using genetic correlation analyses, polygenic scores, and Mendelian randomization. Conversely, higher abdominal adiposity, determined with WHRadjBMI, may be a risk factor for prostate cancer in men. We also report a nominal protective causal effect of central obesity on breast cancer. Additionally, we show nominal genetic correlation between higher abdominal adiposity and lung cancer. Still, among individuals with no history of tobacco use, higher abdominal adiposity appears to be protective of lung cancer.

The association between higher BMI and lower prostate cancer risk is in line with observational studies(7,27). These results are not surprising since it is well-documented that type 2 diabetes (T2D) appears to be protective of prostate cancer(28,29), with obesity being a likely risk factor(30) for T2D. The mechanisms behind the inverse association between prostate cancer and high BMI are yet to be fully defined but several factors need to be considered when interpreting this finding. First, the impact of height on BMI must be put into perspective while highlighting the inverse association between high BMI and prostate cancer risk. BMI is defined by dividing weight by the square of height and as such, tall men may present with lower BMI despite having high body weight. Furthermore, height is an established risk factor for prostate cancer with elevated height linked to increased growth factors such as the insulin-like growth factor 1 (IGF-1)(31,32). Additionally, it has been suggested that the low serum testosterone levels typical among obese men(33) may be responsible as it has been shown that elevated free testosterone levels in men are associated with an increased risk of prostate cancer(34). Moreover, the association between higher BMI and prostate cancer may be influenced by tumour

characteristics. For instance, higher BMI has been shown to increase the risk of aggressive prostate cancer and mortality(27,35,36), while higher BMI has been associated with reduced risk of overall prostate cancer and non-aggressive prostate cancer tumours(27,36).

Conversely, we also show a nominal positive association between higher abdominal adiposity and prostate cancer, in line with epidemiological evidence(9). Abdominal adiposity is associated with markers of metabolic dysfunction, such as insulin resistance and impaired glucose metabolism(3). These results therefore suggest that while total adiposity appears to be protective of prostate cancer, WHRadjBMI is a better predictor of prostate cancer risk, similar to epidemiological observations(37). However, the impact of body height once again warrants a consideration while interpreting this observation. By adjusting for BMI, the effect of height which is an established risk factor for prostate cancer(31) is excluded and could potentially explain the opposite results got for BMI and WHRadjBMI in respect to prostate cancer.

Furthermore, our results indicate that abdominal adiposity, independent of BMI, relates to cancer risk in an opposite manner for sex-specific cancers. Particularly, higher abdominal adiposity in men is a risk factor for prostate cancer, and in contrast, higher abdominal adiposity in women appears to offer a protective advantage against overall breast cancer. However, the latter is contrary to what is seen from observational studies(38). In this study, we considered cross-sectional adiposity measures, which do not reflect the obesity trajectories in women's lifetime, a factor that might influence the exposure to diverse hormones. Other factors such as adiposity at age of menarche could have prolonged effects into adulthood and could in part explain the negative association between central adiposity and breast

cancer. Indeed, higher BMI in early childhood and adolescence have been associated with decreased risk of breast cancer(39). Further work to unravel how early-life adiposity and sex hormones influence risk in an opposite direction for sex-specific cancers is needed.

In addition, the observation from genetic correlation and polygenic scores that higher BMI PGS directly relates to increased lung cancer risk is contrary to what is observed in epidemiological studies(4,40,41). Nevertheless, the significance of this positive association declined once we adjusted for smoking status as a proxy for tobacco use that is associated with lower BMI. Interestingly, this data shows that among individuals who reported to have no history of smoking, higher WHR independent of BMI was protective of lung cancer risk. Moreover, previous findings suggest that higher abdominal adiposity is a risk factor for lung cancer among current smokers(40,41). The mechanism behind this observation is unclear. However, it has been shown that preclinical lung cancer at baseline among current smokers may be associated with higher central adiposity(40).

Despite the lack of statistically significant findings in either PGS and MR analyses for colorectal cancer, there was a trend towards positive association between WHRadjBMI and colorectal cancer in PGS analyses whereas with BMI there was no evidence of association. This is consistent with epidemiological evidence suggesting that abdominal obesity is a stronger predictor for colorectal cancer than overall body weight(42,43). Insulin resistance and consequent hyperinsulinemia, and dyslipidemia associated with metabolic dysfunction have been describes as potential pathophysiology underlying the association with abdominal obesity and colorectal cancer risk(44,45).

This study has some limitations to be considered. Cancer subtypes based on tumour aggressiveness, or biomarker specificity are not considered based on the cancer definitions utilised in this study. A range of cancer subgroups would be characterised by different properties that could relate in different ways to adiposity measures. Future work on this analysis will work on partitioning adiposity GWAS variants into groups based on their apparent mechanistic functions. This would aide in explaining underlying co-morbidity between adiposity and cancer with specific biological pathways such as insulin resistance, inflammation, lipid metabolism, and immunity, in mind. Additionally, the low sample sizes for cancers in our analyses undermine the statistical power for cancer GWAS compared to that of adiposity phenotypes(10–15). Consequently, the precision of the risk estimates derived from Mendelian randomization results was affected and as such validation using larger datasets for cancer phenotypes is required. Finally, due to the lack of publicly available data for certain cancers such as lung and post-menopausal breast cancers, we did not assess causality between these cancers and adiposity.

Overall, we highlight the importance of assessing different adiposity measures in the context of cancer risk by employing analysis of two routine yet different measures of adiposity, BMI and WHRadjBMI, dissecting the relationships between them. We conclude that metabolic dysfunction, through WHRadjBMI, rather than just overall adiposity may be more informative of cancer risk for prostate cancer. However, the impact of height may play a role in the complex relationship between obesity and prostate cancer. Additionally, the differences in risk conferred by central adiposity may occur in opposite direction for sex-specific cancers as seen with prostate and breast cancer. Validation using larger cancer-focussed datasets would be required to confirm our findings.

ACKNOWLEDGEMENTS

We would like to thank the Genetic Investigation of Anthropometric Traits (GIANT) consortium for enabling access to their GWAS summary statistics which we have used in our PGS and MR analyses. We would also like to appreciate the participants and staff of the UK Biobank for their valuable contribution.

REFERENCES

1. Roche AF, Siervogel RM, Chumlea WC, Webb P. Grading body fatness from limited anthropometric data. Am J Clin Nutr [Internet]. 1981 Dec 1 [cited 2022 Jul 1];34(12):2831–8. Available from: https://academic.oup.com/ajcn/article/34/12/2831/4693045
2. Alberti KGMM, Eckel RH, Grundy SM, Zimmet PZ, Cleeman JI, Donato KA, et al. Harmonizing the Metabolic Syndrome. Circulation [Internet]. 2009 Oct 20 [cited 2022 Jul 5];120(16):1640–5. Available from: https://www.ahajournals.org/doi/abs/10.1161/CIRCULATIONAHA.109.192644
3. Lalia AZ, Dasari S, Johnson ML, Robinson MM, Konopka AR, DIstelmaier K, et al. Predictors of Whole-Body Insulin Sensitivity Across Ages and Adiposity in Adult Humans. J Clin Endocrinol Metab [Internet]. 2016 Feb 1 [cited 2022 Jul 1];101(2):626–34. Available from: https://academic.oup.com/jcem/article/101/2/626/2810955
4. Bhaskaran K, Douglas I, Forbes H, dos-Santos-Silva I, Leon DA, Smeeth L. Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5·24 million UK adults. The Lancet [Internet]. 2014 Aug 30 [cited 2022 Mar 4];384(9945):755–65. Available from: http://www.thelancet.com/article/S0140673614608928/fulltext
5. Xu M, Jung X, Hines OJ, Eibl G, Chen Y. Obesity and Pancreatic Cancer: Overview of Epidemiology and Potential Prevention by Weight Loss. Pancreas [Internet]. 2018 Feb 1 [cited 2022 Jul 3];47(2):158–62. Available from: https://journals.lww.com/pancreasjournal/Fulltext/2018/02000/Obesity_and_Pancreatic_Cancer__Overview_of.3.aspx
6. Kyrgiou M, Kalliala I, Markozannes G, Gunter MJ, Paraskevaidis E, Gabra H, et al. Adiposity and cancer at major anatomical sites: umbrella review of the literature. BMJ [Internet]. 2017 Feb 28 [cited 2022 Jul 3];356:477. Available from: https://www.bmj.com/content/356/bmj.j477
7. Golabek T, Bukowczan J, Chlosta P, Powroznik J, Dobruch J, Borówka A. Obesity and Prostate Cancer Incidence and Mortality: A Systematic Review of Prospective Cohort Studies. Urol Int [Internet]. 2014 Jan [cited 2020 Aug 10];92(1):7–14. Available from: https://www.karger.com/Article/FullText/351325
8. Barberio AM, Alareeki A, Viner B, Pader J, Vena JE, Arora P, et al. Central body fatness is a stronger predictor of cancer risk than overall body size. Nature Communications 2019 10:1 [Internet]. 2019 Jan 22 [cited 2022 Apr 19];10(1):1–12. Available from: https://www.nature.com/articles/s41467-018-08159-w
9. Boehm K, Sun M, Larcher A, Blanc-Lapierre A, Schiffmann J, Graefen M, et al. Waist circumference, waist-hip ratio, body mass index, and prostate cancer risk: Results from the North-American case-control study Prostate Cancer & Environment Study. Urologic Oncology: Seminars and Original Investigations. 2015 Nov 1;33(11):494.e1-494.e7.
10. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. Hum Mol Genet [Internet]. 2018 Oct 15 [cited 2019 Apr 28];27(20):3641–9. Available from: https://academic.oup.com/hmg/article/27/20/3641/5067845
11. Pulit SL, Stoneman C, Morris AP, Wood AR, Glastonbury CA, Tyrrell J, et al. Meta-analysis of genome-wide association studies for body fat distribution in 694 649

individuals of European ancestry. Hum Mol Genet [Internet]. 2019 Jan 1 [cited 2021 Oct 6];28(1):166–74. Available from: https://academic.oup.com/hmg/article/28/1/166/5098227

12. Zhang H, Ahearn TU, Lecarpentier J, Barnes D, Beesley J, Qi G, et al. Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. Nature Genetics 2020 52:6 [Internet]. 2020 May 18 [cited 2021 Oct 6];52(6):572–81. Available from: https://www.nature.com/articles/s41588-020-0609-2
13. Schumacher FR, al Olama AA, Berndt SI, Benlloch S, Ahmed M, Saunders EJ, et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. Nat Genet [Internet]. 2018 Jul 1 [cited 2020 Jul 8];50(7):928–36. Available from: https://doi.org/10.1038/s41588-018-0142-8
14. Law PJ, Timofeeva M, Fernandez-Rozadilla C, Broderick P, Studd J, Fernandez-Tajes J, et al. Association analyses identify 31 new risk loci for colorectal cancer susceptibility. Nat Commun [Internet]. 2019 Dec 1 [cited 2020 Jul 22];10(1):1–15. Available from: https://doi.org/10.1038/s41467-019-09775-w
15. Klein AP, Wolpin BM, Risch HA, Stolzenberg-Solomon RZ, Mocci E, Zhang M, et al. Genome-wide meta-analysis identifies five new susceptibility loci for pancreatic cancer. Nature Communications 2018 9:1 [Internet]. 2018 Feb 8 [cited 2021 Oct 6];9(1):1–11. Available from: https://www.nature.com/articles/s41467-018-02942-5
16. Lewis CM, Vassos E. Polygenic risk scores: From research tools to clinical instruments. Genome Med [Internet]. 2020 May 18 [cited 2022 Jul 3];12(1):1–11. Available from: https://genomemedicine.biomedcentral.com/articles/10.1186/s13073-020-00742-5
17. Lawlor DA, Harbord RM, Sterne JAC, Timpson N, Smith GD. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. Stat Med [Internet]. 2008 Apr 15 [cited 2022 Feb 18];27(8):1133–63. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/sim.3034
18. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. Genome-wide genetic data on ~500,000 UK Biobank participants Supplementary Material. bioRxiv. 2017;
19. Bulik-Sullivan B, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nature Genetics 2015 47:3 [Internet]. 2015 Feb 2 [cited 2022 Jul 3];47(3):291–5. Available from: https://www.nature.com/articles/ng.3211
20. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature [Internet]. 2015 Feb 12 [cited 2018 Nov 1];518(7538):197–206. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25673413
21. Shungin D, Winkler T, Croteau-Chonka DC, Ferreira T, Locke AE, Mägi R, et al. New genetic loci link adipose and insulin biology to body fat distribution. Nature 2015 518:7538 [Internet]. 2015 Feb 11 [cited 2022 Jul 3];518(7538):187–96. Available from: https://www.nature.com/articles/nature14132
22. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. The American Journal of Human Genetics. 2007;
23. Team Rs. RStudio: Integrated Development Environment for R [Internet]. 2000. Available from: https://www.rstudio.com/

24. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-base platform supports systematic causal inference across the human phenome. Elife. 2018 May 30;7.
25. Burgess S, Butterworth A, Thompson SG. Mendelian Randomization Analysis With Multiple Genetic Variants Using Summarized Data. Genet Epidemiol [Internet]. 2013 Nov 1 [cited 2022 May 29];37(7):658–65. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/gepi.21758
26. Bowden J, Holmes M v. Meta-analysis and Mendelian randomization: A review. Res Synth Methods [Internet]. 2019 Dec 1 [cited 2022 Dec 14];10(4):486–96. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/jrsm.1346
27. Perez-Cornago A, Appleby PN, Pischon T, Tsilidis KK, Tjønneland A, Olsen A, et al. Tall height and obesity are associated with an increased risk of aggressive prostate cancer: Results from the EPIC cohort study. BMC Med [Internet]. 2017 Jul 13 [cited 2022 Jul 10];15(1):1–11. Available from: https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-017-0876-7
28. Waters KM, Henderson BE, Stram DO, Wan P, Kolonel LN, Haiman CA. Association of Diabetes With Prostate Cancer Risk in the Multiethnic Cohort. Am J Epidemiol [Internet]. 2009 Apr 4 [cited 2022 Jul 22];169(8):937. Available from: /pmc/articles/PMC2727229/
29. Tsilidis KK, Kasimis JC, Lopez DS, Ntzani EE, Ioannidis JPA. Type 2 diabetes and cancer: umbrella review of meta-analyses of observational studies. BMJ [Internet]. 2015 Jan 2 [cited 2022 Jul 22];350. Available from: https://www.bmj.com/content/350/bmj.g7607
30. Gong Z, Neuhouser ML, Goodman PJ, Albanes D, Chi C, Hsing AW, et al. Obesity, diabetes, and risk of prostate cancer: results from the prostate cancer prevention trial. Cancer Epidemiol Biomarkers Prev [Internet]. 2006 Oct [cited 2022 Jul 22];15(10):1977–83. Available from: https://pubmed.ncbi.nlm.nih.gov/17035408/
31. Perez-Cornago A, Appleby PN, Pischon T, Tsilidis KK, Tjønneland A, Olsen A, et al. Tall height and obesity are associated with an increased risk of aggressive prostate cancer: Results from the EPIC cohort study. BMC Med [Internet]. 2017 Jul 13 [cited 2022 Nov 21];15(1):1–11. Available from: https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-017-0876-7
32. Renehan AG, Zwahlen M, Minder C, O'Dwyer ST, Shalet SM, Egger M. Insulin-like growth factor (IGF)-I, IGF binding protein-3, and cancer risk: systematic review and meta-regression analysis. Lancet [Internet]. 2004 Apr 24 [cited 2022 Nov 21];363(9418):1346–53. Available from: https://pubmed.ncbi.nlm.nih.gov/15110491/
33. Lima N, Cavaliere H, Knobel M, Halpern A, Medeiros-Neto G. Decreased androgen levels in massively obese men may be associated with impaired function of the gonadostat. Int J Obes Relat Metab Disord [Internet]. 2000 [cited 2022 Jul 10];24(11):1433–7. Available from: https://pubmed.ncbi.nlm.nih.gov/11126339/
34. Platz EA, Leitzmann MF, Rifai N, Kantoff PW, Chen YC, Stampfer MJ, et al. Sex steroid hormones and the androgen receptor gene CAG repeat and subsequent risk of prostate cancer in the prostate-specific antigen era. Cancer Epidemiol Biomarkers Prev [Internet]. 2005 May [cited 2022 Jul 10];14(5):1262–9. Available from: https://pubmed.ncbi.nlm.nih.gov/15894683/

35. Obesity, weight gain and cancer risk - WCRF International [Internet]. [cited 2022 Jul 9]. Available from: https://www.wcrf.org/diet-activity-and-cancer/risk-factors/obesity-weight-gain-and-cancer/
36. Perez-Cornago A, Dunneram Y, Watts EL, Key TJ, Travis RC. Adiposity and risk of prostate cancer death: a prospective analysis in UK Biobank and meta-analysis of published studies. BMC Med [Internet]. 2022 Dec 1 [cited 2022 Jul 10];20(1):1–13. Available from: https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-022-02336-x
37. Tang B, Han CT, Zhang GM, Zhang CZ, Yang WY, Shen Y, et al. Waist-hip Ratio (WHR), a Better Predictor for Prostate Cancer than Body Mass Index (BMI): Results from a Chinese Hospital-based Biopsy Cohort. Sci Rep [Internet]. 2017 Mar 8 [cited 2022 Jul 11];7. Available from: /pmc/articles/PMC5341100/
38. George SM, Bernstein L, Smith AW, Neuhouser ML, Baumgartner KB, Baumgartner RN, et al. Central adiposity after breast cancer diagnosis is related to mortality in the Health, Eating, Activity, and Lifestyle study. Breast Cancer Research and Treatment 2014 146:3 [Internet]. 2014 Jul 24 [cited 2022 Jul 11];146(3):647–55. Available from: https://link.springer.com/article/10.1007/s10549-014-3048-x
39. Baer HJ, Tworoger SS, Hankinson SE, Willett WC. Body fatness at young ages and risk of breast cancer throughout life. Am J Epidemiol [Internet]. 2010 [cited 2022 Dec 7];171(11):1183–94. Available from: https://pubmed.ncbi.nlm.nih.gov/20460303/
40. Kabat GC, Kim M, Hunt JR, Chlebowski RT, Rohan TE. Body Mass Index and Waist Circumference in Relation to Lung Cancer Risk in the Women's Health Initiative. Am J Epidemiol [Internet]. 2008 Jul 15 [cited 2022 Jul 25];168(2):158–69. Available from: https://academic.oup.com/aje/article/168/2/158/101166
41. Dewi NU, Boshuizen HC, Johansson M, Vineis P, Kampman E, Steffen A, et al. Anthropometry and the Risk of Lung Cancer in EPIC. Am J Epidemiol [Internet]. 2016 Jul 15 [cited 2022 Jul 25];184(2):129–39. Available from: https://academic.oup.com/aje/article/184/2/129/2236633
42. Pischon T, Lahmann PH, Boeing H, Friedenreich C, Norat T, Tjønneland A, et al. Body size and risk of colon and rectal cancer in the European Prospective Investigation Into Cancer and Nutrition (EPIC). J Natl Cancer Inst [Internet]. 2006 Jul 5 [cited 2022 Nov 21];98(13):920–31. Available from: https://pubmed.ncbi.nlm.nih.gov/16818856/
43. Dong Y, Zhou J, Zhu Y, Luo L, He T, Hu H, et al. Abdominal obesity and colorectal cancer risk: systematic review and meta-analysis of prospective studies. Biosci Rep [Internet]. 2017 Dec 22 [cited 2022 Nov 18];37(6). Available from: https://pubmed.ncbi.nlm.nih.gov/29026008/
44. Moore LL, Bradlee ML, Singer MR, Splansky GL, Proctor MH, Ellison RC, et al. BMI and waist circumference as predictors of lifetime colon cancer risk in Framingham Study adults. International Journal of Obesity 2004 28:4 [Internet]. 2004 Feb 10 [cited 2022 Nov 21];28(4):559–67. Available from: https://www.nature.com/articles/0802606
45. Renehan AG, Roberts DL, Dive C. Obesity and cancer: pathophysiological and biological mechanisms. Arch Physiol Biochem [Internet]. 2008 Feb [cited 2022 Nov 21];114(1):71–83. Available from: https://pubmed.ncbi.nlm.nih.gov/18465361/

Table 1. Association results between BMI and WHRadjBMI PGS and cancers in UK Biobank

	BMI PGS		WHRadjBMI PGS	
Cancer	**OR(95%CI)**	**P**	**OR(95%CI)**	**P**
BrC	0.98(0.97-0.99)	0.0086	0.98(0.97-0.99)	0.021
PostBrC	0.97(0.96-0.99)	0.00203	0.98(0.97-1.0023)	0.089
PrC	0.97(0.95-0.99)	0.0012	1.018(0.99-1.037)	0.074
CrC	0.99(0.97-1.02)	0.73	1.021(0.99-1.044)	0.062
LungC	1.044(1.013-1.076)	0.0057	0.99(0.97-1.027)	0.82

Legend: BrC=breast cancer, PostBrC=post-menopausal breast cancer, PrC=prostate cancer, CrC=colorectal cancer, LungC=lung cancer, OR(95%CI)=odds ratio and the lower and upper confidence intervals

Table 2. Mendelian randomization results with cancers as the exposure and adiposity measures as the outcome variable using the five different methods

		Inverse variance weighted		**MR Egger**		**Weighted median**		**Simple mode**		**Weighted mode**	
Exposure	Outcome	OR (95% CI)	P	OR (95% CI)	P	OR (95% CI)	P	OR (95% CI)	P	OR (95% CI)	P
BMI	BrC	1.000 (0.995-1.005)	0.897	0.985 (0.971-1.000)	0.051	0.996 (0.988-1.003)	0.266	0.944 (0.968-1.020)	0.634	0.994 (0.977-1.011)	0.462
BMI	PrC	**0.993 (0.988-0.998)**	**0.0042**	0.995 (0.982-1.009)	0.473	**0.993 (0.985-0.999)**	**0.039**	0.984 (0.960-1.009)	0.22	0.995 (0.981-1.009)	0.457
BMI	CrC	1.000 (0.998-1.002)	0.92	1.001 (0.996-1.006)	0.682	1.000 (0.997-1.003)	1	0.999 (0.989-1.008)	0.768	1.000 (0.995-1.005)	0.93
WHRadjBMI	BrC	**0.990 (0.983-0.997)**	**0.0068**	1.000 (0.982-1.017)	0.974	0.991 (0.981-1.002)	0.105	1.003 (0.974-1.033)	0.83	0.993 (0.978-1.008)	0.339
WHRadjBMI	PrC	1.0046 (0.998-1.011)	0.179	**1.016 (1.00018-1.032)**	**0.048**	1.007 (0.999-1.016)	0.094	1.018 (0.990-1.045)	0.209	**1.022 (1.00067-1.038)**	**0.0053**
WHRadjBMI	CrC	1.002 (0.994-1.004)	0.125	1.000 (0.995-1.006)	0.885	1.000 (0.996-1.004)	0.917	0.995 (0.984-1.006)	0.391	1.000 (0.994-1.006)	0.988

Legend: BrC=breast cancer, PostBrC=post-menopausal breast cancer, PrC=prostate cancer, CrC=colorectal cancer, LungC=lung cancer, OR(95%CI)=odds ratio and the lower and upper 95% confidence intervals. Causal estimates with *P*<0.05 are shown in bold.

Figure 1. BMI (A) and WHRadjBMI (B) Polygenic score analyses pipeline
Legend: A two-sample approach was used to construct our PGS base data. The SNPs used for the PGS were from GIANT's latest BMI and WHRadjBMI meta-analyses. Since the meta-analyses comprised of the UK Biobank (our target data), we used weights from the non-UK Biobank study in GIANT's meta-analyses

Figure 2. Venn diagram defining overlapping genome-wide significant loci between adiposity phenotypes and sex-specific cancers based on UK Biobank GWAS summary statistics
Legend: A) Loci overlap between BMI and the sex-specific cancers in the UK Biobank. B) Loci overlap between WHRadjBMI and the sex-specific cancers in the UK Biobank. BrC= overall breast cancer, PostBrC=post-menopausal breast cancer, PrC=prostate cancer.

Figure 3. Genetic correlation and SNP heritability estimates of BMI/WHRadjBMI and cancer in UK Biobank
Legend: A). Genetic correlation estimates between UK Biobank BMI/WHRadjBMI and cancer. Estimates which withstood Bonferroni corrections (P<0.05/10=0.005) are marked with triple asterisks (***), double asterisks (**) for $0.05 \geq P > 0.005$. BrC=overall breast cancer, PostBrC=post-menopausal breast cancer, PrC=prostate cancer, CrC=colorectal cancer, LungC=lung cancer.
B). SNP heritability estimates of BMI/WHRadjBMI and cancer in UK Biobank

Figure 4. The association analysis results between PGS for BMI and WHRadjBMI and cancer in the UK Biobank
Legend: BrC=overall breast cancer, PostBrC=post-menopausal breast cancer, PrC=prostate cancer, CrC=colorectal cancer, LungC=lung cancer. Estimates which withstood Bonferroni corrections (P<0.05/10=0.005) are marked with triple asterisks (***), double asterisks (**) for $0.05 \geq P > 0.005$

Figure 5. Scatter and forest plots from Mendelian Randomization analysis for A) BMI to prostate cancer MR, B) WHRadjBMI to prostate cancer
Legend: The scatter plots on the left also include the intercepts of the various MR methods used. The forest plots show the MR estimates for each of the methods used with the odds ratio, 95% confidence intervals and P values annotated. Causal estimates with P<0.05 are highlighted in bold.

Figure 1.

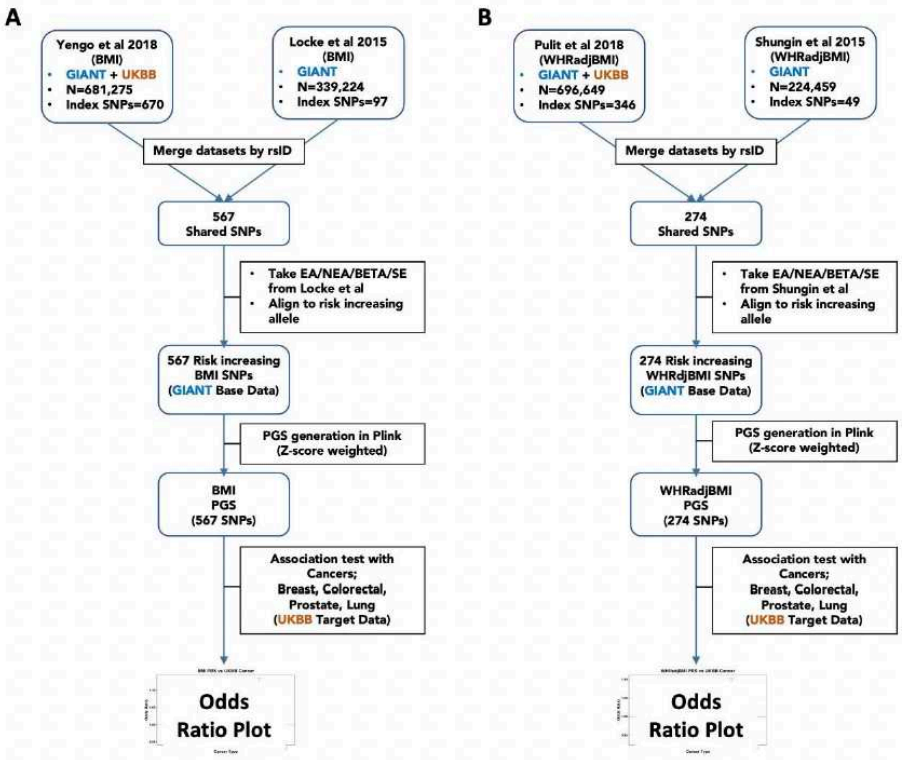


Figure 2.

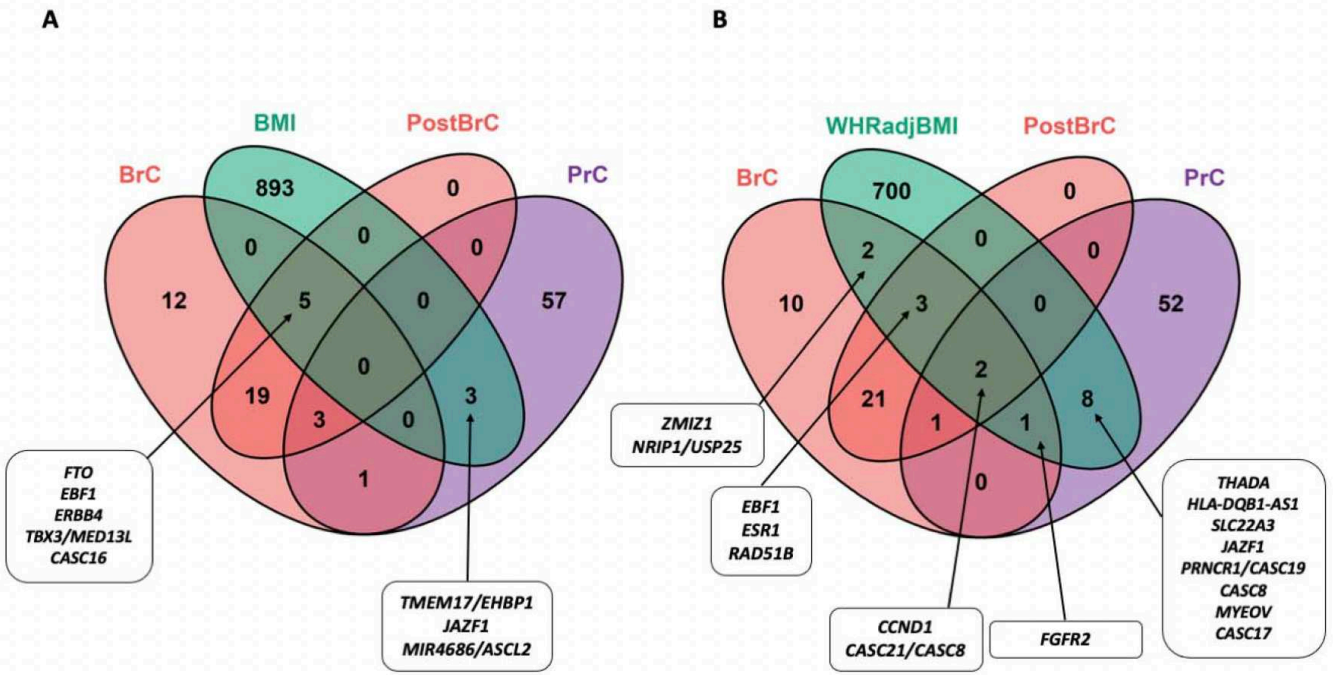


Figure 3.

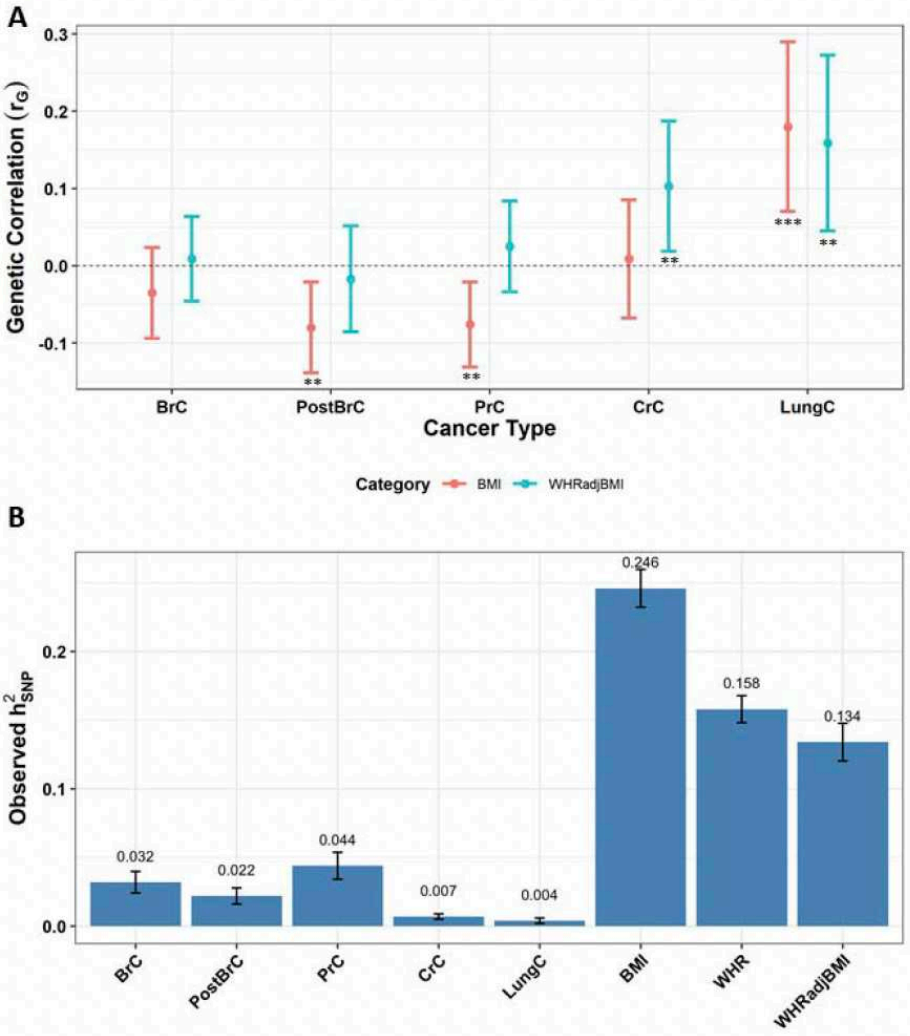


Figure 4.

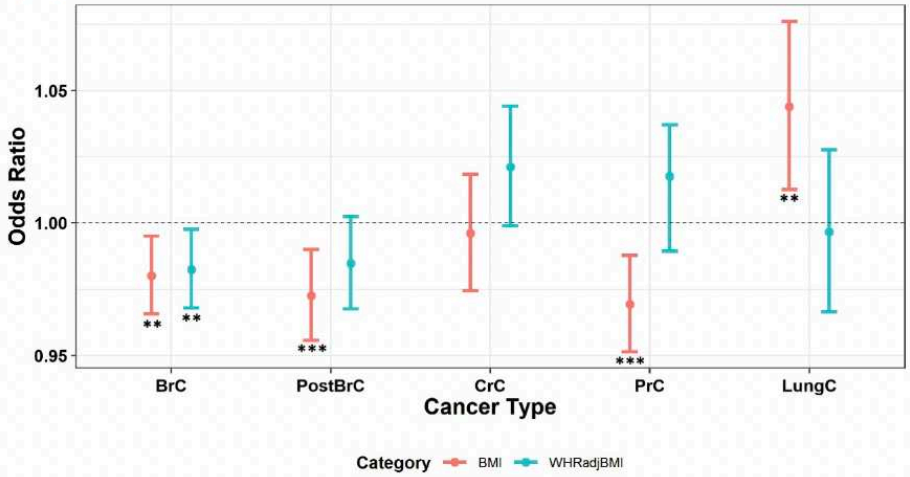
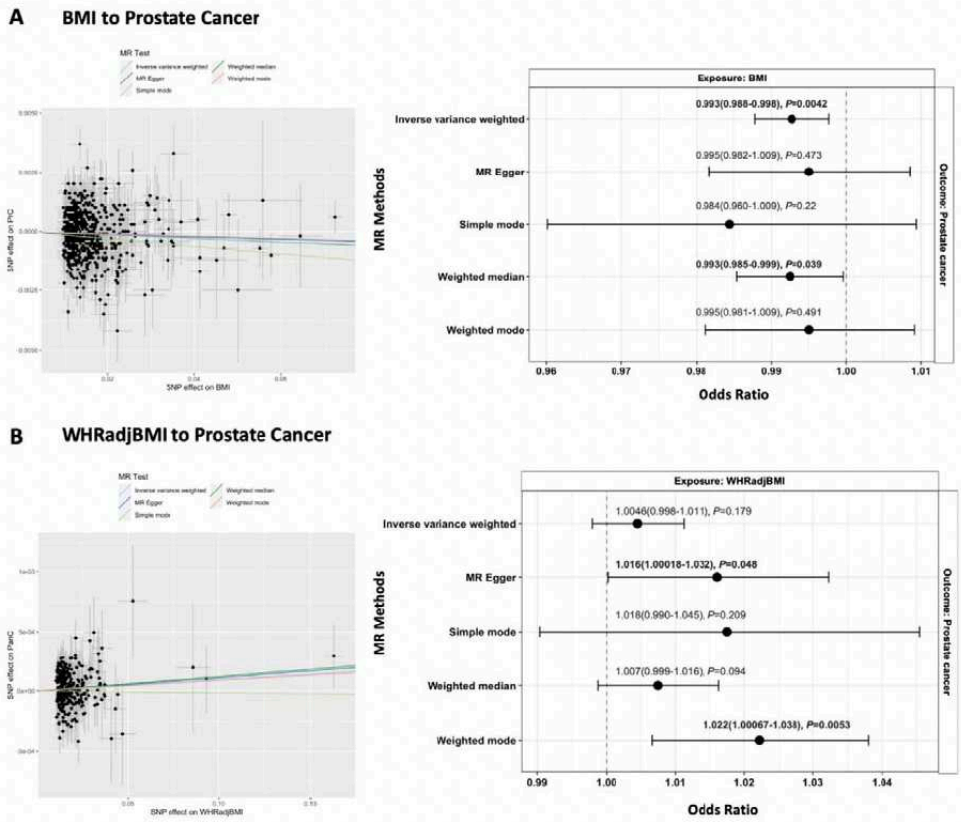


Figure 5.



5.1 Supplementary data

UKBB GWAS

We performed single phenotype GWAS in the UKBB phenotypes using the *BOLT-LMM* version 2.3 software[91] which implements a linear mixed model (LMM) association testing. Consequently, as a result of applying a linear mixed model, related individuals in the UKBB were included in the association analyses. The standard *BOLT-LMM* v2.3 infinitesimal model was used. Among the 487,409 individuals with genetic data, the genetic data was filtered based on MAF > 0.01, imputation score > 0.4, Hardy-Weinberg Equilibrium (HWE) P-value $>1x10^{-6}$ and per SNP variant missingness <0.015. As a result, 471,095 individuals passed these filters. We included age, sex, genotyping array and six principal components (PCs) as covariates in the LMM for BMI, WHRadjBMI, colorectal cancer (CrC) and lung cancer (LungC). For the sex-specific cancer phenotypes breast (BrC), post-menopausal breast (PostBrC) and prostate (PrC) cancers, sex was not included as a covariate. Moreover, BMI was included as an additional covariate in WHR association testing to obtain the WHRadjBMI phenotype. The threshold for statistically significant genome-wide signals (SNPs) was $P<5x10^{-8}$. Manhattan plots for the association results are show in **Supplementary Figures 3-6**.

Table 4. Phenotype definition criteria for cancer phenotypes in UKBB

Cancer	UK Biobank field description	ICD-10 Codes
Overall breast cancer (BrC)	Have been diagnosed with breast cancer AND Breast cancer is the first cancer diagnosed OR Death cause is breast cancer	C50
Post-menopausal breast cancer (Post-BrC)	Have been diagnosed with breast cancer AND Breast cancer is the first cancer diagnosed OR Death cause is breast cancer AND Self-reported menopause status	C50 and X2724 (Menopause status)
Colorectal cancer (CrC)	Have been diagnosed with colon and rectal cancer AND colon and rectal cancers are the first cancers diagnosed OR Death cause is colon and rectal cancers	C18-C21
Prostate cancer (PrC)	Have been diagnosed with prostate cancer AND prostate cancer is the first cancer diagnosed OR Death cause is prostate cancer	C61
Lung cancer (LungC)	Have been diagnosed with lung cancer AND lung cancer is the first cancer diagnosed OR Death cause is lung cancer	C34

Table 5. Genetic correlation estimates between BMI/WHRadjBMI and cancer in UKBB

Cancer	Adiposity trait	Genetic correlation (rG)	SE	P
BrC	BMI	-0.035	0.03	0.236
Post-BrC	BMI	-0.0803	0.03	0.014
PrC	BMI	-0.076	0.028	0.0075
CrC	BMI	0.0089	0.039	0.82
LungC	BMI	0.18	0.056	0.0014
BrC	WHRadjBMI	0.009	0.028	0.75
Post-BrC	WHRadjBMI	-0.017	0.035	0.63
PrC	WHRadjBMI	0.025	0.03	0.408
CrC	WHRadjBMI	0.103	0.043	0.017
LungC	WHRadjBMI	0.159	0.058	0.0065

Table 6. SNP heritability estimates of BMI, WHRadjBMI and cancer in UKBB

Phenotype	h^2_{SNP}	SE
BrC	0.0323	0.0041
Post-BrC	0.0215	0.003
PrC	0.0441	0.005
CrC	0.0072	0.0013
LungC	0.0036	0.0011
BMI	0.2459	0.0072
WHRadjBMI	0.1343	0.0066

Table 7. BMI PRS association with cancer in UKBB by BMI categories

Cancer	BMI Class	Case/Control	OR (95%CI)	P
BrC	Underweight	83/1775	0.94 (0.78-1.13)	0.497
	Normal	4,796/93,114	0.98 (0.96-1.01)	0.178
	Pre-obesity	5,277/86,114	0.98 (0.95-1.00)	0.053
	Obesity	**3,164/54,072**	**0.96 (0.93-0.99)**	**0.012**
Post-BrC	Underweight	124,1,734	0.87 (0.70-1.09)	0.242
	Normal	6,872/91,058	0.97 (0.94-1.00)	0.029
	Pre-obesity	7,188/84,240	0.98 (0.95-1.01)	0.161
	Obesity	**4,436/52,830**	**0.96 (0.92-0.99)**	**0.017**
PrC	Underweight	18.456	0.90 (0.55-1.47)	0.68
	Normal	2,926/48,880	0.98 (0.94-1.02)	0.351
	Pre-obesity	**6,128/97,260**	**0.98 (0.95-1.00)**	**0.066**
	Obesity	2,728/50,611	0.98 (0.94-1.02)	0.412
CrC	Underweight	27/2,305	0.89 (0.60-1.32)	0.569
	Normal	2,440/147,476	0.98 (0.94-1.02)	0.35
	Pre-obesity	3,711/191,068	0.99 (0.96-1.03)	0.65
	Obesity	2,193/108,380	0.97 (0.93-1.02)	0.215
LungC	Underweight	46/2,289	1.13 (0.84-1.50)	0.417
	Normal	1,293/148,528	1.04 (0.99-1.10)	0.136
	Pre-obesity	**1,763/193,131**	**1.08 (1.03-1.14)**	**0.001**
	Obesity	1,105/109,592	1.01 (0.95-1.07)	0.856

Associations with $P<0.05$ are shown in bold

Table 8. Obesity PRS association with lung cancer by smoking status

		BMI PRS		WHRadjBMI PRS	
Smoking status	**Case/Control/N**	**OR (95%CI)**	**P**	**OR (95%CI)**	**P**
Previous smokers	1,970/160,891 (162,861)	1.02 (0.98-1.07)	0.294	0.99 (0.95-1.04)	0.737
Current smokers	1,652/46,242 (47,894)	1.01 (0.96-1.06)	0.839	1.03 (0.98-1.08)	0.281
Never smoked	579/246,273 (246,852)	1.01 (0.93-1.10)	0.759	0.92 (0.85-1.00)	**0.046**
Previous + current smokers	3,622/207,133 (210,755)	1.02 (0.99-1.05)	0.231	1.01 (0.98-1.04)	0.638

Associations with $P<0.05$ are shown in bold

5.2 Insights

In this second article submitted to Obesity journal, I implemented genetic correlation, polygenic scores, and Mendelian randomization approaches to assess how BMI and WHRadjBMI, used as proxies for overall and central obesity respectively, relate to the risk of breast, prostate, colorectal and lung cancers.

The polygenic scores analyses indicated that both central and overall obesity relate to prostate cancer risk in opposite direction. Specifically, overall obesity has an inverse association with prostate cancer risk while central obesity has a direct association with prostate cancer risk. Furthermore, Mendelian randomization corroborated these findings while using published GWAS data from non-overlapping datasets. While the exact mechanisms underlying this observed paradoxical relationship between obesity and prostate cancer, factors such as growth factors (IGF-1), androgens (testosterone) and differences in tumour characteristics may play a role in the manifestation of this co-morbidity. Further work to characterise in detail how central obesity and other components of the metabolic syndrome affect prostate cancer risk are needed. Moreover, the impact of height on prostate cancer needs careful consideration when interpreting these results.

The polygenic scores and Mendelian randomization analyses of breast cancer indicate that central obesity has an inverse association with overall breast cancer risk. Several factors such as sex hormones, menopause status, tumour characteristics, BMI at age of menarche are potentially involved. More work to unravel the exact involvement of these and other factors is required.

Despite nominally significant positive genetic correlation estimates for colorectal cancer and WHRadjBMI, polygenic scores and Mendelian randomization showed no significant results for colorectal cancer. Future work will focus on larger sample sizes for colorectal cancer so as to improve statistical power.

Lung cancer polygenic scores analyses indicated a nominally significant positive association between BMI and lung cancer risk. On the other hand, WHRadjBMI was not significantly associated with lung cancer. Sensitivity analyses, however, suggested that among individuals with no smoking experience, WHRadjBMI may be inversely associated with lung cancer risk. This is despite evidence that smokers, and not their non-smoking counterparts, have lower body fat. Consequently, validation of these findings is needed using larger sample sizes.

Unfortunately, Mendelian randomization analyses could not be performed for lung cancer and post-menopausal breast cancer due to the unavailability of GWAS summary statistics. Therefore, the causality between obesity and these cancers remain unaddressed in this study and thus future studies will aim to fill this gap.

In summary, this article I demonstrates how central and overall obesity have different risk patterns for different cancers. I also illustrate how different statistical genetics tools can assist in disentangling the relationship between obesity and cancer. Researchers and clinical health advisors should thus broaden their definition of obesity in practice to accurately capture the involvement of adiposity in cancer risk.

6 THIRD ARTICLE

“Bi-directional Mendelian randomization and multi-phenotype GWAS show causality and shared pathophysiology between depression and type 2 diabetes” (Preprint link https://doi.org/10.1101/2022.12.06.22283143)

Bi-directional Mendelian randomization and multi-phenotype GWAS show causality and shared pathophysiology between depression and type 2 diabetes

Jared G Maina, MSc[1,2*], Zhanna Balkhiyarova, MD, PhD[3,4,5*], Arie Nouwen, PhD[6], Igor Pupko[3,4], Anna Ulrich, PhD[4], Mathilde Boissel, MSc[1,2], Amélie Bonnefond, PhD[1,2,4], Philippe Froguel, MD PhD[1,2,4], Amna Khamis, PhD[1,2,4], Inga Prokopenko, PhD[1,2,3,5*], Marika Kaakinen, PhD[3,4,5*]

1. INSERM UMR 1283, CNRS UMR 8199, European Genomic Institute for Diabetes (EGID), Institut Pasteur de Lille, F-59000 Lille, France
2. University of Lille, Lille University Hospital, Lille, F-59000, France
3. Department of Clinical & Experimental Medicine, University of Surrey, Guildford, United Kingdom
4. Department of Metabolism, Digestion and Reproduction, Imperial College London, London, UK
5. People-Centred Artificial Intelligence Institute, University of Surrey, Guildford, United Kingdom
6. Department of Psychology, Middlesex University, London, UK

* These authors contributed equally to this research.

Corresponding authors:

Dr Marika Kaakinen,
Section of Statistical Multi-Omics,
Department of Clinical and Experimental Medicine, University of Surrey
Leggett Building, Daphne Jackson Road, Manor Campus,
Guildford, Surrey, UK, GU2 7WG
Phone: +44(0)1483683365
Email:m.kaakinen@surrey.ac.uk

Prof. Inga Prokopenko
Section of Statistical Multi-Omics,
Department of Clinical and Experimental Medicine, University of Surrey
Leggett Building, Daphne Jackson Road, Manor Campus,
Guildford, Surrey, UK, GU2 7WG
Phone: +44(0)1483684648
Email:i.prokopenko@surrey.ac.uk

Dr Amna Khamis
European Genomic Institute for Diabetes - UMR 1283/8199
Pôle Recherche - 1er étage Aile Ouest
1, Place de Verdun, 59045 LILLE CEDEX
Phone: +33(0)3 74 00 8135

Email:amna.khamis@cnrs.fr

Running title (Limit 47 characters including spaces): MR and Multi-trait GWAS of diabetes and depression

Word count (4128/4000)

Number of tables and figures in main text (2 tables, 2 figures)

Abstract (250/250 words)

OBJECTIVE. Depression is a common co-morbidity of type 2 diabetes. However, the causality and underlying mechanisms remain unclear.

RESEARCH DESIGN AND METHODS. We applied bi-directional Mendelian randomization (MR) to assess causality between type 2 diabetes and self-reported depression. Using the UK biobank, we performed 1) GWAS, separately, and 2) multi-phenotype GWAS (MP-GWAS) of type 2 diabetes (cases=19,344, controls=463,641) and depression, using two depression definitions–clinically diagnosed major depressive disorder (MDD, cases=5,262, controls=86,275) and self-reported depressive symptoms (PHQ-9, n=153,079). The FinnGen study was used for replication for MDD (n=23,424) and type 2 diabetes (n=32,469). Based on the results, we analyzed expression quantitative trait loci (eQTL) data from public databases to identify target genes in relevant tissues.

RESULTS. MR demonstrated a significant causal effect of depression on type 2 diabetes (OR=1.18[1.06-1.32], p=0.0024), but not in the reverse direction. GWAS of type 2 diabetes and depressive symptoms did not identify any shared loci between them, whereas MP-GWAS identified seven shared loci mapped to *TCF7L2*, *CDKAL1*, *IGF2BP2*, *SPRY2*, *CCND2-AS1*, *IRS1*, *CDKN2B-AS1*. MDD did not yield genome-wide significant loci in either GWAS or MP-GWAS. We found that most MP-GWAS *loci* had an eQTL, including SNPs implicating the cell cycle gene *CCND2* in pancreatic islets and brain, and key insulin signaling gene *IRS1* in adipose tissue, suggesting a multi-tissue and pleiotropic underlying mechanism.

CONCLUSION. Our study reveals the complexity in the depression-diabetes relationship and our results have important implications for a more efficient prevention of type 2 diabetes from early adulthood when depressive symptoms usually occur.

INTRODUCTION

Type 2 diabetes is a disease characterized by chronic hyperglycemia and depression is a common co-morbidity, potentially due to common shared risk factors, such as lifestyle, early growth environment, psychotropic drugs, and neuro-endocrine dysfunction[1]. Whether the impact of type 2 diabetes on depression is stronger than the reverse remains to be defined. It has indeed been shown that depression, even at sub-clinical levels, increases the risk of incident type 2 diabetes by 25-60%[2,3], whereas others have shown that type 2 diabetes increases the risk of depression by 40-60%[4].

The causality of the associations from observational studies remains unclear due to unmeasured confounding and potential reverse causation. However, this could be circumvented in part through Mendelian randomization, an approach that assesses potential causality between phenotypes using genetic variants as instruments, since genes are allocated randomly at birth and are free of confounding[5]. To date, only one MR study from China has reported a possible causal link from type 2 diabetes to depression[6], yet the reverse direction of causal link was not examined.

Recent large-scale genome-wide association studies (GWAS) for type 2 diabetes and depression have reported 403 and 102 associated genomic *loci* for these diseases, respectively[7,8]. Moreover, analyses based on the GWAS results support a positive genetic correlation (r_G) between them[7,8], suggesting shared genetic background. However, the majority of GWAS investigate each disease independently, without considering the genetic correlation between related phenotypes and their heritabilities.

Therefore, in this study, we addressed the causal relationship between depression and type 2 diabetes by conducting an MR study using summary statistics from recent GWAS of depression[8] and type 2 diabetes[7]. Additionally, we used the UK biobank (UKBB) to perform a multi-phenotype GWAS (MP-GWAS) for type 2 diabetes and depression to identify shared

96 genetic *loci* between the two diseases. For depression, we compared two assessment
97 approaches - clinically diagnosed major depressive disorder (MDD) and depressive
98 symptoms based on self-report.

RESEARCH DESIGN AND METHODS

Mendelian randomization

Summary statistics used

To test for causality between type 2 diabetes and depression, we performed a two-sample bidirectional MR, first using depression as a risk factor and type 2 diabetes as an outcome, then testing type 2 diabetes as a risk factor and depression as an outcome from two non-overlapping datasets (**Supplementary Figure 1**). The single nucleotide polymorphisms (SNPs) used as genetic instruments for type 2 diabetes and self-reported depression were from recent large-scale European GWAS meta-analyses of the two diseases[7,8].

Two-sample bidirectional MR

All MR analyses were conducted using the R software package *TwoSampleMR v0.5.4*[9]. In the type 2 diabetes GWAS summary statistics used, 95 single nucleotide polymorphisms (SNPs, inclusive of 6 proxy variants with a minimum $r^2 \geq 0.8$) out of the 102 independent ($r^2 < 0.01$) depression SNPs were available. We excluded 7 palindromic SNPs (A/T or C/G) with intermediate allele frequencies (minor allele frequency, MAF>45%) to ensure the effects of the SNPs for the two phenotypes were aligned to the same forward strand allele. The genetic instruments for type 2 diabetes included 403 genetic SNPs associated from a recent GWAS[7]. In the depression summary statistics, 358 (inclusive of 4 proxy variants with a minimum $r^2 \geq 0.8$) out of 403 independent ($r^2 < 0.01$) type 2 diabetes SNPs were available for the analysis. To obtain the causal estimate, we applied the inverse variance weighted (IVW) method[5]. We performed sensitivity MR analysis using weighted median (WM)[10], MR-Egger regression[11], the simple mode[12], and the weighted mode[12] methods to evaluate the potential violations of the MR assumptions (**Supplementary Methods**) and confirm the robustness of the two-sample MR results from the IVW approach. F-statistic was used to evaluate the instrument strength, where F>10 indicates the presence of a strong instrument. The F-statistics indicated

a good instrument strength for both type 2 diabetes (F-statistics = 61.26) and depression (F-statistics = 43).We assessed heterogeneity between the causal estimates from each SNP using Cochran's Q-test. The sensitivity of causal inference to any individual genetic variant was tested by leave-one-out analysis. We used the STROBE-MR reporting guideline for MR studies to facilitate the readers' evaluation of our results[13] **(Supplementary Table 1).**

Genome-wide association studies (GWAS) and Multi-phenotype GWAS (MP-GWAS)

Cohorts used

1) UK Biobank (UKBB)

We used data from the UK Biobank (UKBB, www.ukbiobank.ac.uk), which includes over 500,000 individuals from 22 centers across the United Kingdom. Study participants were between 40 and 69 years at recruitment and provided information including body measurements, biological samples, brain imaging data, socio-demographic and lifestyle indicators. Genetic data was available for 488,377 individuals in the UKBB genotyped using the UKBB BiLEVE array (n = 49,979) and the UKBB Axiom Array (n = 438,398)[14]. The genetic data was imputed using the Haplotype Reference Consortium[15], the UK10K[16] and 1000 Genomes Phase 3[17], resulting in approximately 90 million variants available for association testing. This research has been conducted under application number 35327 and all participants gave informed consent during enrolment.

2) FinnGen

We utilized the FinnGen summary statistics for replication of our single- and multi-phenotype association results (www.finngen.fi/fi). The June 2020 data freeze used in our analysis comprised of 135,638 individuals. Summary statistics for 1,801 phenotypes are publicly available for analysis. FinnGen study participants were genotyped using the Illumina and Affymetrix chip arrays (Illumina Inc., San Diego, and Thermo Fisher Scientific, Santa

Clara, CA, USA, https://www.thermofisher.com/). The data was then imputed using the SISu v3 imputation panel (http://sisuproject.fi) resulting in 16,962,023 variants available for association analysis. FinnGen study participants were genotyped using the Illumina and Affymetrix chip arrays (Illumina Inc., San Diego, and Thermo Fisher Scientific, Santa Clara, CA, USA, https://www.thermofisher.com/). Summary statistics provided for the FinnGen data association analyses were generated using the SAIGE software[18].

Phenotype definition

Type 2 diabetes

In the UKBB, type 2 diabetes cases were defined if individuals self-reported to have a diabetes diagnosis by a doctor, were on insulin medication one year after diagnosis, and were at least 40 years old at the time of diagnosis. Individuals not meeting these criteria were classified as controls. For both cases and controls, we excluded individuals with gestational diabetes (Field 4041, code = 1), those younger than 40 years at the time of diagnosis (Field 2976) and individuals on insulin medication within the first year of diagnosis (Field 2986). Sex discordant individuals (genotype vs. reported sex) were also excluded from the analyses. We restricted our analyses to European individuals to limit confounding by ancestry. In total, 19,344 cases and 463,641 controls were available (**Supplementary Table 2**).

The GWAS summary statistics in the FinnGen study were based on the ICD10-coded type 2 diabetes (ICD code E11) on 32,469 individuals (case/control numbers not available).

Depression

We defined depression in two ways: ICD-10 coded major depressive disorder (MDD) based on linked data from hospital records and self-reported depressive symptoms using the Patient Health Questionnaire 9 (PHQ-9)[19].

ICD-coded MDD

Individuals with a primary diagnosis of a depressive episode (ICD code F32) and recurrent depression (ICD code F33) were defined as ICD-coded MDD cases (hereinafter referred to as MDD). Individuals who answered "NO" to the questions "Have you ever seen a general practitioner (GP) for nerves, anxiety, tension or depression?" (Field 2090) and "Have you ever seen a psychiatrist for nerves, anxiety, tension or depression?" and "NO" to either "depressed/down for a whole week" (Field 4598) or "Ever unenthusiastic/disinterested for a whole week" were set as controls. Participants were excluded from the study if they had a diagnosis of bipolar disorder (ICD codes F30, F31), mixed and other personality disorder (F61) and schizophrenia (ICD code F20). Participants on antipsychotic medication (Field 20003) for 58 drugs were also excluded. In total, 5,262 cases and 86,275 controls were used for the MDD phenotype (**Supplementary Table 1**). The proportion of MDD participants who had a type 2 diabetes diagnosis are shown in **Supplementary Table 3**.

FinnGen study data had the ICD-10 codes F32 and F33 available for MDD. The GWAS summary statistics on MDD were based on a total of 23,424 individuals (case/control numbers not available).

Depressive symptoms

In UKBB, self-reported depressive symptoms over the previous two weeks (from the time of study enrolment) were assessed using the (PHQ-9)[19] questionnaire (**Supplementary Table 4**). It has been shown that the PHQ-9 questionnaire is invariant between people with and without diabetes[20] suggesting its interpretation is similar for both diabetes cases and controls. Individuals missing responses for more than three PHQ-9 items were excluded from the analysis. Missing PHQ-9 responses for the remaining individuals were imputed using the ImputeSCOPA software (https://github.com/ImperialStatGen/imputeSCOPA), which implements a random forest approach to impute missing items. The variables sex, age,

education qualification, body mass index (BMI), Townsend Deprivation Index[21] (an area-based measure of deprivation), genotyping array and eight principal components (PCs) were included in the imputation model to improve the predictive accuracy of the imputation (**Supplementary Table 5**). The sum of all nine PHQ-9 items after imputation for everyone was used for quantitative association analysis. PHQ-9 data were available for 153,079 individuals (**Supplementary Table 1**). The proportion of individuals with PHQ-9 data and a type 2 diabetes diagnosis are shown in **Supplementary Tables 3 and 6**. Symptom-based depression phenotypes were unavailable in the FinnGen replication dataset.

GWAS

We performed separate GWAS for type 2 diabetes, MDD and PHQ-9 in UKBB data with BOLT-LMM using a linear mixed model[22]. We adjusted for age, sex, array, BMI and the first 8 PCs. We analyzed common variants (MAF>5%), with imputation scores >0.4, Hardy-Weinberg Equilibrium (HWE) p-value>1×10^{-6} and per SNP variant missingness<0.015. Manhattan plots were constructed using the *ggplot2* R package[23]. All analyses were performed on Human genome build 37. The statistical threshold for genome-wide significant SNPs (signals) used was $p<5\times10^{-8}$.

Multi-phenotype GWAS

We used MTAG (Multi-Trait Analysis of GWAS)[24], which implements a generalized inverse-variance weighted meta-analysis, to increase the power for locus identification, improve SNP effect size estimates for type 2 diabetes and depressive phenotypes, and to identify potential multi-phenotype genetic variant effects. We used the summary statistics of the individual GWAS as inputs of MTAG. In UKBB, two MTAG models were tested - one with type 2 diabetes and MDD and another with type 2 diabetes and total PHQ-9 scores. This was to assess the consequence of using two different depression definition criteria (disease

diagnosis vs. disease symptoms) in an MP-GWAS approach. In FinnGen, only the first approach was applied due to data availability. For each MTAG model tested, MTAG outputs phenotype specific association statistics. To assess the robustness of MP-GWAS results, MTAG computes a maximum false discovery rate (maxFDR) statistic, a theoretical upper bound limit on the FDR for a GWAS[24]. Lower maxFDR values (maxFDR<5%) indicate robust results.

Expression quantitative trait *loci* (eQTL) analyses

To explore and identify target genes of the identified *loci*, we utilized several eQTL databases and datasets of relevant tissues. We extracted eQTL data from the GTEx Portal (https://gtexportal.org) for SNPs identified in our MP-GWAS, focusing on type 2 diabetes and depression relevant tissues (*i.e.*, brain, muscle, liver). In addition, as GTEx does not include data from pancreatic islets, a crucial tissue in type 2 diabetes pathogenicity, we utilized recent eQTL data from Tiger T2D Systems (http://tiger.bsc.es) obtained from > 500 brain-dead organ donor islets[25]. We extracted data for the seven shared SNPs identified in our MP-GWAS from both eQTL studies (using nominal significance, $p< 0.05$).

Furthermore, we also used GTEx Version 7 transcriptome data[26] from European individuals to identify eQTLs using our MP-GWAS summary statistics. We focused on relevant tissues in 1) type 2 diabetes, namely liver, whole pancreas, muscle, adipose subcutaneous, adrenal gland, whole blood, and 2) depression, including putamen basal ganglia, hippocampus, substantia nigra, frontal cortex, amygdala, anterior cingulate cortex. For each tissue, the predicted expression levels were then correlated with type 2 diabetes and PHQ-9 MTAG summary statistics. *P*-values were corrected for multiple testing using Bonferroni correction based on the number of genes tested per tissue (**Supplementary Table 7**). Genes where less than 80% of the SNPs used in the model were found in the GWAS summary statistics were excluded due to low reliability of association results. This analysis focused on type 2 diabetes

and PHQ-9 phenotypes only as the MDD phenotype was underpowered in both GWAS and MP-GWAS.

RESULTS

Mendelian randomization

Our MR analysis revealed that depression was causally and positively associated with type 2 diabetes using the IVW method, with an OR of 1.18 (95%CI = 1.06-1.32; p = 0.0024). This result was consistent with the WM sensitivity analyses, which showed an OR of 1.11 (95%CI = 1.00-1.23, p = 0.043) (**Figure 1 and Supplementary Figure 2A, Supplementary Table 8**). The MR-Egger test showed no evidence of directional pleiotropy (p = 0.51), further confirming the validity of the results. Additionally, the leave-one-out analysis showed no outliers, suggesting that the observed association was not changed after removing any single variant (**Supplementary Figure 3**). The Cochran's Q statistic for heterogeneity was significant for the IVW method (Q=261.62, p=1.26x10^{-20}).

We found no evidence of causality in the reverse direction between type 2 diabetes and depression, in the primary nor sensitivity analysis (IVW: OR = 0.999; CI = 0.99-1.01; p = 0.843) (**Supplementary Figure 2B, Supplementary Table 9**).

Genome wide association study in type 2 diabetes and depression

In order to identify whether type 2 diabetes and depression have a shared genetic etiology, we first performed a GWAS for both phenotypes, separately, using the UKBB. In total, for type 2 diabetes, we used 482,958 individuals (19,344 cases; 463,641 controls), and for depression we used 91,537 (5,262 cases; 86,276 controls) for MDD and 153,079 for PHQ-9. For type 2 diabetes, we identified 92 independent SNPs at 84 *loci*, of which 59 were replicated in the FinnGen with nominal significance (p<0.05) and consistent in direction of effect. For depression, we found three independent SNPs and *loci* for PHQ-9 and no SNPs associated

with the binary depression MDD trait in the UKBB. None of the GWAS SNPs identified in type 2 diabetes were shared with depression (**Supplementary Figure 4, Supplementary Tables 10-12**).

Multi-phenotype GWAS in UKBB

To improve the power to identify shared genetic variants between the two phenotypes, we performed the largest-to-date MP-GWAS of type 2 diabetes and depressive phenotypes in the UKBB using the respective GWAS summary statistics.

The MP-GWAS model with type 2 diabetes and MDD did not identify any significant associations for the binary definition of depression (MDD). For type 2 diabetes, we identified 71 independent signals at 66 *loci* (**Supplementary Figure 5 and 6, Supplementary Table 10**). The maxFDR for the type 2 diabetes and MDD MP-GWAS results was 1.5% and 7.7% respectively (**Supplementary Table 13**), indicating that the MDD results were likely inflated by the higher powered type 2 diabetes GWAS[24]. In FinnGen, the maxFDR were 11.5% and 25.5%, respectively, indicating highly inflated results for both traits.

In contrast, in the MP-GWAS model with type 2 diabetes and PHQ-9, we identified eight independent SNPs for PHQ-9 (compared to only three SNPs detected in GWAS) (**Figure 2, Supplementary Table 14**), suggesting greater power for SNP discovery. Only the *CACNA2D2* locus was reported in both GWAS and MP-GWAS analyses for PHQ-9, although with different, yet highly correlated lead SNPs (*CACNA2D2*, chromosome 2, $SNP_{SP\text{-}GWAS}$ rs35335661, $SNP_{MP\text{-}GWAS}$ rs1467916, r^2=0.99). For type 2 diabetes, MP-GWAS identified 53 SNPs at 50 *loci* (compared to 92 identified in the single-phenotype GWAS), of which 24 *loci* were previously reported[7]. We replicated 37 type 2 diabetes signals in FinnGen cohort (**Supplementary Table 14**).

In total, we found seven SNPs shared between type 2 diabetes and PHQ-9: rs7903146 (*TCF7L2*), rs7766070 (*CDKAL1*), rs1359790 (*SPRY2*), rs16860235 (*IGF2BP2*), rs76895963 (*CCND2-AS1*), rs2972144 (*IRS1*) and rs10811662 (*CDKN2B-AS1*) (**Table 1, Figure 2**). The maxFDR for type 2 diabetes and PHQ-9 after MP-GWAS were 0.98% and 1.8% respectively (**Supplementary Table 13**), indicating the results are robust and not influenced by the sample sizes of either GWAS.

eQTL analyses

To explore whether the seven SNPs shared between type 2 diabetes and self-reported depression had a downstream functional impact, we first extracted eQTL data of the seven identified SNPs using the 1) GTEx Portal in relevant tissues, including muscle, liver and brain, and 2) the Tiger Portal for pancreatic islets. We found that of the seven shared SNPs between type 2 diabetes and depression, six SNPs were associated with the expression of nearby genes in relevant tissues (**Table 2**). This includes rs2972144-G risk allele associated with a decreased expression of *IRS1* and *RP11-395N3.2* in visceral and subcutaneous adipose and *RP11-395N3.1* in subcutaneous adipose tissue. Additionally, the rs76895963-T risk allele was associated with the decreased expression of *CCND2* in pancreatic islets, brain cerebellum, skeletal muscle and subcutaneous adipose tissue, the decreased expression of its antisense *CCND2-AS1* in pancreatic islets, brain cerebellum, basal ganglia and cortex, in addition to *CCND2-AS2* in the cerebellum. For pancreatic islets, we found an increased expression of *NDUFA9* (**Table 2**).

In addition, using our MP-GWAS summary statistics, we tested in GTEx whether type 2 diabetes and PHQ-9 were associated with shared gene expression changes in several target tissues except for pancreatic islets as GTEx does not include data on that. This analysis identified additional target genes associated with both type 2 diabetes and PHQ-9 in eight tissues, consistent in direction of effect for both phenotypes: adipose subcutaneous (*IRS1*,

321 *NCR3LG1*, *RP11-395N3.2*), amygdala (*HSPA1B*), frontal cortex (*CDKAL1*), hypothalamus
322 (*EIF2S2P3*), skeletal muscle (*HLA-DRA, RP11-370C19.2*), substantia nigra (*BETL1*), and
323 whole blood (*EIF2S2P3, HLA-DRB1*) (**Supplementary Figure 7, Supplementary Table 7**
324 **and 15-17**). Altogether, our data demonstrates a functional impact of our *loci* in several
325 related tissues.

CONCLUSIONS

We performed a large comprehensive study to investigate the relationship between type 2 diabetes and depression and found evidence for a causal positive association from depression to type 2 diabetes. We also performed a multi-phenotype GWAS of the two diseases, highlighting seven shared *loci* that target nearby genes in several target tissues. Altogether, our study provides novel insight into the underlying mechanisms linking the two diseases.

Our MR analysis, based on recent large-scale GWAS, provides evidence for causality on the previously reported epidemiological associations from depression to type 2 diabetes[3], but not in the reverse direction. These findings are consistent with the pathophysiology for these diseases, whereby: 1) depression starts in adolescence or early adulthood[27], whereas type 2 diabetes usually develops later[28], 2) poor health habits among individuals with depression, including smoking, physical inactivity and increased caloric intake (and associated overweight), that are known to facilitate the development of type 2 diabetes[3,29], 3) antidepressants frequently induce weight gain leading to type 2 diabetes [29], and 4) the systemic inflammation associated to increased stress hormone levels such as cortisol in the context of depression also favors insulin resistance[30].

While epidemiological studies also show increased risk of depression in people with type 2 diabetes[4], we did not find evidence for causality in this relationship. Our evaluation of instrument strength for the MR analysis suggests that the lack of this causal association is not simply an issue of statistical power. A likely hypothesis is that the epidemiologically observed association is confounded via other factors, which are not easily assessed in epidemiological studies, such as psychosocial factors related to the painful management of a middle-age chronic disease. Diabetes distress, the psychological burden caused by dealing with having diabetes and having to care for it, has indeed been linked to depression[31]. However, a large international longitudinal study of 14 countries has shown that only

depressive symptoms rather than MDD were predicted by diabetes distress one year after diagnosis[32].

In addition, our multi-phenotype GWAS revealed seven shared SNPs between type 2 diabetes and self-reported depressive symptoms consistent in their directions of effect. It is interesting to note that we did not find any shared SNPs associated with the binary MDD clinical diagnosis and type 2 diabetes which we speculate could be due to several reasons: 1) that the association between type 2 diabetes and depression is due to the less strictly defined symptom-based depressive measures, and 2) the PHQ-9 is a continuous definition of depression, and likely improves power compared to the binary MDD definition. Indeed, previous studies showed that depression defined based on self-reported symptoms (minimal phenotyping) rather than strict diagnostic criteria enables greater power for locus discovery in GWAS[33].

Our eQTL analysis provided some clues to the underlying mechanisms linking the two diseases. For instance, we found that the rs76895963-T risk allele was associated with the decreased expression of *CCND2* in the brain and pancreatic islets and insulin target tissues (adipose and skeletal muscle), suggesting a pleiotropic effect of this *locus*. *CCND2* encodes Cyclin D2, which is involved in cell cycle regulation and with a role in pancreatic beta cell proliferation and insulin secretion[34], consistent with our eQTL data showing a decreased expression of the gene in pancreatic islets. Mutations in *CCND2* have been described in individuals with brain malformations[35], confirming a role in maintaining brain growth, and mouse knockout models of *CCND2* have no brain neurogenesis and showed mild depression-like symptoms that were alleviated by anti-depressant chronic fluoxetine treatment[36]. In addition, a recent study showed that neurogenesis in the brain may prevent depressive symptoms. Indeed, these studies demonstrate a potential mechanism whereby *CCND2* down-regulation may halt this process[37]. In addition, the inhibition of *CCND2* in adipose tissue has

been shown to dysregulate adipocyte differentiation[38] and downregulated in obesity, suggesting a key role in maintaining adipocyte regeneration[39]. Therefore, we show that *CCND2* has a multi-system and pleiotropic effect and could potentially mediate this relationship between type 2 diabetes and depression.

In addition, we found that the rs2972144-G risk allele was associated with decreased expression of *IRS1*, which encodes insulin receptor substrate 1, in adipose tissue. *IRS1* is a key signaling molecule necessary for insulin response in insulin target tissues and has been shown to be associated with insulin resistance[40], a condition linked to both the development of type 2 diabetes and depression[41].

Also, our eQTL analyses revealed further insights into the potential mechanisms underlying the relationship between type 2 diabetes and depression. Lower expression of *CDKAL1* in frontal cortex was associated both with type 2 diabetes and depressive symptoms. Although reported for bipolar disorder[42,43], variation in *CDKAL1* has been associated with depressive phenotypes only in one previous study[44]. In addition, although we did not find an eQTL for this *locus* in pancreatic islets in our study, variation at *CDKAL1* has been implicated in type 2 diabetes and shown to reduce insulin secretion[45].

We also found two target genes implicating the HLA region (*HLA-DRA*, *HLA-DRB1*) in the shared pathogenesis of the two diseases, in blood and skeletal muscle. The role of depression treatments, targeting the immune system, is very active[46]. Type 2 diabetes is also known to have an impact on immune system with high blood glucose levels causing an inflammatory response[47]. While trying to underpin the molecular mechanisms within this potential shared pathway, further studies on the type 2 diabetes-depression comorbidity should account also for the role of obesity with its known links with inflammation[48], type 2 diabetes[49] and suggested with depression[50].

Our results highlight a shared genetic effect between the two phenotypes with plausible biological significance that explain how environmental factors (*i.e.*, stress, lifestyle habits and anti-depressant medication) could lead to the underlying co-morbidity. Therefore, we speculate that our results could hold some clinical significance. For instance, the choice of anti-depressant treatment offered to people with depression at risk of type 2 diabetes should favor those that provide better glycemic control such as selective serotonin reuptake inhibitors (SSRIs)[51]. Additionally, people with depression should be encouraged, as part of routine clinical care, to promote positive lifestyle habits such as increased physical activity, adequate sleep, and a proper dietary regime.

This study exhibits several strengths. We report the first study to investigate the causal relationship of type 2 diabetes and depression in both directions and performed the largest-to-date MP-GWAS for the two diseases. Our investigation highlights the ability of multi-phenotype approaches to reveal the shared associations in co-morbid diseases and further confirms the power of large-scale datasets to uncover phenotype associations, including self-reported and continuous measures of diseases and their symptoms. However, there are some limitations to be considered. Our main UKBB analyses on MDD did not yield any genome-wide significant signals in both GWAS and MP-GWAS and a validation in larger datasets is needed. In addition, to effectively probe the credibility of the MTAG MP-GWAS results, replication using another MP-GWAS method is needed as part of future analyses. Finally, we could only replicate the type 2 diabetes-MDD analyses due to unavailability of publicly available depressive symptoms GWAS and in the FinnGen cohort.

In conclusion, the shared *loci* between depressive symptoms and type 2 diabetes support a pleiotropic role in target tissues, providing insight into their pathophysiology and co-morbidity, boosting our understanding of the pathogenesis of these two diseases. Additionally, self-reported depression/depressive symptoms may offer more in deciphering

425 the underlying co-morbidity with type 2 diabetes compared to the strictly defined MDD. The
426 causal effect of depression leading to the development of type 2 diabetes has important
427 implications for a more efficient prevention of type 2 diabetes from early adulthood.

ACKNOWLEDGEMENTS

This research has been conducted using the UK Biobank Resource under Application Number 35327. This research was in part funded by the Diabetes UK (BDA number: 20/0006307), the European Union's Horizon 2020 research and innovation programme (LONGITOOLS, H2020-SC1-2019-874739), Agence Nationale de la Recherche (PreciDIAB, ANR-18-IBHU-0001), by the European Union through the "Fonds européen de développement regional" (FEDER), by the "Conseil Régional des Hauts-de-France" (Hauts-de-France Regional Council) and by the "Métropole Européenne de Lille" (MEL, European Metropolis of Lille).

REFERENCES

1. Holt, R. I. G., Peveler, R. C. & Byrne, C. D. Schizophrenia, the metabolic syndrome and diabetes. *Diabetic Medicine* vol. 21 515–523 Preprint at https://doi.org/10.1111/j.1464-5491.2004.01199.x (2004).
2. Mezuk, B., Eaton, W. W., Albrecht, S. & Golden, S. H. Depression and Type 2 Diabetes Over the LifespanA meta-analysis. *Diabetes Care* **31**, 2383–2390 (2008).
3. Rotella, F. & Mannucci, E. Depression as a Risk Factor for Diabetes. *J Clin Psychiatry* **74**, 31–37 (2013).
4. Nouwen, A. *et al*. Type 2 diabetes mellitus as a risk factor for the onset of depression: a systematic review and meta-analysis. *Diabetologia* **53**, 2480–2486 (2010).
5. Lawlor, D. A., Harbord, R. M., Sterne, J. A. C., Timpson, N. & Smith, G. D. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Stat Med* **27**, 1133–1163 (2008).
6. Xuan, L. *et al*. Type 2 diabetes is causally associated with depression: a Mendelian randomization analysis. *Front Med* **12**, 678–687 (2018).
7. Mahajan, A. *et al*. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet* **50**, 1505–1513 (2018).
8. Howard, D. M. *et al*. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat Neurosci* **22**, 343–352 (2019).
9. Hemani, G. *et al*. The MR-base platform supports systematic causal inference across the human phenome. *Elife* **7**, (2018).
10. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet Epidemiol* **40**, 304–314 (2016).
11. Bowden, J., Smith, G. D. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* **44**, 512–525 (2015).
12. Hartwig, F. P., Smith, G. D. & Bowden, J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol* **46**, 1985–1998 (2017).
13. Skrivankova, V. W. *et al*. Strengthening the reporting of observational studies in epidemiology using mendelian randomisation (STROBE-MR): explanation and elaboration. *BMJ* **375**, (2021).
14. Bycroft, C. *et al*. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
15. McCarthy, S. *et al*. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279–1283 (2016).
16. Huang, J. *et al*. Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nat Commun* (2015) doi:10.1038/ncomms9111.
17. Auton, A. *et al*. A global reference for human genetic variation. *Nature* vol. 526 68–74 Preprint at https://doi.org/10.1038/nature15393 (2015).
18. Zhou, W. *et al*. Scalable generalized linear mixed model for region-based association tests in large biobanks and cohorts. *Nat Genet* **52**, 634–639 (2020).
19. Kroenke, K., Spitzer, R. L. & Williams, J. B. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* **16**, 606–13 (2001).
20. Nouwen, A. *et al*. Measurement invariance testing of the patient health questionnaire-9 (PHQ-9) across people with and without diabetes mellitus from the NHANES, EMHS and UK Biobank datasets. *J Affect Disord* **292**, 311–318 (2021).

21. Townsend, P., Phillimore, P. & Beattie, Alastair. Health and deprivation : inequality and the North. 211 (1988).
22. Loh, P. R. *et al*. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* (2015) doi:10.1038/ng.3190.
23. Wickham, Hadley. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).
24. Turley, P. *et al*. Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat Genet* **50**, 229–237 (2018).
25. Alonso, L. *et al*. TIGER: The gene expression regulatory variation landscape of human pancreatic islets. *Cell Rep* **37**, (2021).
26. Lonsdale, J. *et al*. The Genotype-Tissue Expression (GTEx) project. *Nature Genetics* vol. 45 580–585 Preprint at https://doi.org/10.1038/ng.2653 (2013).
27. Solmi, M. *et al*. Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies. *Molecular Psychiatry 2021* 1–15 (2021) doi:10.1038/s41380-021-01161-7.
28. Bruce, D. G. *et al*. Lifetime depression history and depression risk in type 2 diabetes: A case-control study. *J Diabetes Complications* **30**, 38–42 (2016).
29. Strine, T. W. *et al*. The association of depression and anxiety with obesity and unhealthy behaviors among community-dwelling US adults. *Gen Hosp Psychiatry* **30**, 127–137 (2008).
30. Golden, S. A review of the evidence for a neuroendocrine link between stress, depression and diabetes mellitus. *Curr Diabetes Rev* **3**, 252–259 (2007).
31. Fisher, L. *et al*. Clinical Depression Versus Distress Among Patients With Type 2 DiabetesNot just a question of semantics. *Diabetes Care* **30**, 542–548 (2007).
32. Lloyd, C. E. *et al*. Factors associated with the onset of major depressive disorder in adults with type 2 diabetes living in 12 different countries: results from the INTERPRET-DD prospective study. *Epidemiol Psychiatr Sci* **29**, e134 (2020).
33. Cai, N. *et al*. Minimal phenotyping yields genome-wide association signals of low specificity for major depression. *Nat Genet* **52**, 437–447 (2020).
34. Stamateris, R. E. *et al*. Glucose Induces mouse β-cell proliferation via IRS2, MTOR, and cyclin D2 but Not the insulin receptor. *Diabetes* **65**, 981–995 (2016).
35. Pirozzi, F. *et al*. Proximal variants in CCND2 associated with microcephaly, short stature, and developmental delay: A case series and review of inverse brain growth phenotypes. *Am J Med Genet A* **185**, 2719–2738 (2021).
36. Jedynak, P., Kos, T., Sandi, C., Kaczmarek, L. & Filipkowski, R. K. Mice with ablated adult brain neurogenesis are not impaired in antidepressant response to chronic fluoxetine. *J Psychiatr Res* **56**, 106–111 (2014).
37. Svoboda, E. Brain-cell growth keeps mood disorders at bay. *Nature* **608**, S48–S49 (2022).
38. Hishida, T., Naito, K., Osada, S., Nishizuka, M. & Imagawa, M. Crucial roles of D-type cyclins in the early stage of adipocyte differentiation. *Biochem Biophys Res Commun* **370**, 289–294 (2008).
39. Molina-Ayala, M. A. *et al*. Expression of obesity- and type-2 diabetes-associated genes in omental adipose tissue of individuals with obesity. *Gene* **815**, 146181 (2022).
40. DeFronzo, R. A. & Tripathy, D. Skeletal Muscle Insulin Resistance Is the Primary Defect in Type 2 Diabetes. *Diabetes Care* **32**, S157 (2009).
41. Lyra e Silva, N. de M. *et al*. Insulin resistance as a shared pathogenic mechanism between depression and type 2 diabetes. *Frontiers in Psychiatry* vol. 10 57 Preprint at https://doi.org/10.3389/fpsyt.2019.00057 (2019).

42. Nurnberger, J. I. *et al*. Identification of pathways for bipolar disorder: A meta-analysis. *JAMA Psychiatry* **71**, 657–664 (2014).
43. Swaminathan, S. *et al*. Characteristics of Bipolar I patients grouped by externalizing disorders. *J Affect Disord* **178**, 206–214 (2015).
44. Haljas, K. *et al*. Bivariate Genome-Wide Association Study of Depressive Symptoms With Type 2 Diabetes and Quantitative Glycemic Traits. *Psychosom Med* **80**, 242–251 (2018).
45. Kirchhoff, K. *et al*. Polymorphisms in the TCF7L2, CDKAL1 and SLC30A8 genes are associated with impaired proinsulin conversion. *Diabetologia* **51**, 597–601 (2008).
46. Lee, C. H. & Giuliani, F. The Role of Inflammation in Depression and Fatigue. *Front Immunol* **10**, 1696 (2019).
47. Berbudi, A., Rahmadika, N., Tjahjadi, A. I. & Ruslami, R. Type 2 Diabetes and its Impact on the Immune System. *Curr Diabetes Rev* **16**, 442–449 (2019).
48. Ellulu, M. S., Patimah, I., Khaza'ai, H., Rahmat, A. & Abed, Y. Obesity and inflammation: the linking mechanism and the complications. *Archives of Medical Science* **13**, 851–863 (2017).
49. Al-Goblan, A. S., Al-Alfi, M. A. & Khan, M. Z. Mechanism linking diabetes mellitus and obesity. *Diabetes Metab Syndr Obes* **7**, 587–591 (2014).
50. Blasco, B. V., García-Jiménez, J., Bodoano, I. & Gutiérrez-Rojas, L. Obesity and Depression: Its Prevalence and Influence as a Prognostic Factor: A Systematic Review. *Psychiatry Investig* **17**, 715–724 (2020).
51. Deuschle, M. Effects of antidepressants on glucose metabolism and diabetes mellitus type 2 in adults. *Curr Opin Psychiatry* **26**, 60–65 (2013).

Table1. Summary statistics of the seven shared loci between type 2 diabetes and PHQ-9 after MP-GWAS

							Type 2 diabetes			PHQ-9		
CHR	SNP	BP	NEAREST GENE	EA	NEA	EAF	BETA	SE	*P*	BETA	SE	*P*
2	rs2972144	227101411	*IRS1*	G	A	0.65	0.0031	0.00038	$4.89x10^{-16}$	0.065	0.012	$1.80x10^{-08}$
3	rs16860235	185512361	*IGF2BP2*	A	G	0.28	0.0046	0.00040	$1.41x10^{-30}$	0.103	0.012	$4.19x10^{-17}$
6	rs7766070	20686573	*CDKAL1**	A	C	0.26	0.0046	0.00041	$7.79x10^{-29}$	0.103	0.014	$1.24x10^{-12}$
9	rs10811662	22134253	*CDKN2B-AS1*	G	A	0.83	0.0049	0.00048	$7.88x10^{-25}$	0.09	0.015	$4.30x10^{-09}$
10	rs7903146	114758349	*TCF7L2*	T	C	0.29	0.011	0.00040	$6.88x10^{-178}$	0.21	0.012	$1.68x10^{-64}$
12	rs76895963	4384844	*CCND2-AS1*	T	G	0.98	0.015	0.0014	$3.08x10^{-27}$	0.29	0.043	$1.91x10^{-11}$
13	rs1359790	80717156	*SPRY2*	G	A	0.72	0.0035	0.00040	$8.54x10^{-17}$	0.076	0.012	$5.91x10^{-10}$

* Different SNP for PHQ-9 (rs2206734, Position-20694884, Effect Allele-G, Non-effect Allele-C; In high LD with type 2 diabetes SNP=R2=0.544)

Legend: CHR=chromosome, BP=base pair position (genome build 37), EA=effect allele, NEA=non-effect allele, EAF=effect allele frequency

564 **Table 2. Target genes of SNPs shared type 2 diabetes and depressive symptoms in relevant target tissues.**
565

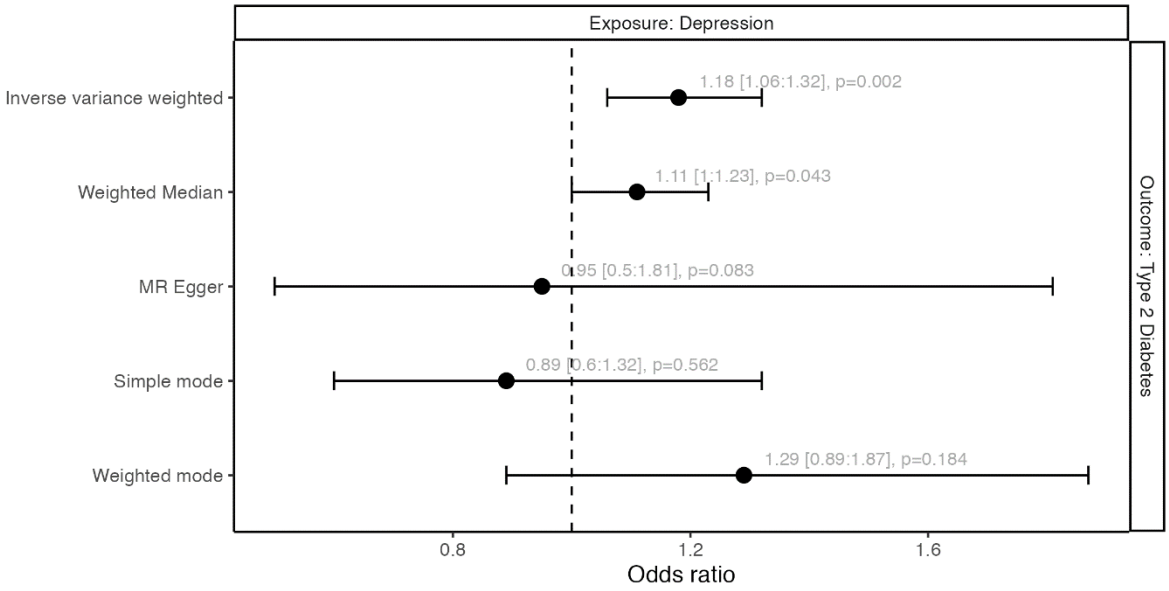
						Organ donors			
CHR	**SNP**	**BP**	**Nearest gene**	**EA**	**NEA**	**Tissue**	**Effect direction**	**Gene target**	***P***
2	rs2972144	227101411	*IRS1*	G	A	Adipose - Subcutaneous	Negative	*IRS1*	$2.60x10^{-16}$
						Adipose - Subcutaneous	Negative	*RP11-395N3.2*	$5.70x10^{-09}$
						Adipose - Subcutaneous	Negative	*RP11-395N3.1*	$1.70x10^{-07}$
						Adipose - Visceral	Negative	*IRS1*	$2.80x10^{-11}$
						Adipose - Visceral	Negative	*RP11-395N3.2*	$1.6x10^{-06}$
3	rs16860235	185512361	*IGF2BP2*	A	G	Pancreatic islets	Negative	*IGF2BP2*	$3.94x10^{-5}$
						Pancreatic islets	Negative	*C3orf70*	0.028
						Pancreatic islets	Positive	*EIF4A2*	0.045
6	rs7766070	20686573	*CDKAL1*	A	C	Pancreatic islets	Positive	*LINC005811*	0.049
9	rs10811662	22134253	*CDKN2A-AS1*	G	A	Pancreatic islets	Positive	*MTAP*	0.038
						Pancreatic islets	Positive	*CDKN2A*	0.00099
						Pancreatic islets	Positive	*CDKN2B-AS1*	$1.54x10^{-7}$
10	rs7903146	114758349	*TCF7L2*	T	C	Pancreatic islets	Positive	*TCF7L2*	0.0029
12	rs76895963	4384844	*CCND2-AS1*	T	G	Pancreatic islets	Negative	*CCND2*	1.68×10^{-6}
						Pancreatic islets	Negative	*CCND2-AS1*	0.017
						Pancreatic islets	Positive	*NDUFA9*	0.039
						Brain - Cerebellum	Negative	*CCND2*	$7.90x10^{-22}$
						Brain - Cerebellum	Negative	*CCND2-AS1*	$2.10x10^{-16}$
						Brain - Cerebellum	Negative	*CCND2-AS2*	$1.00x10^{-08}$
						Brain - Nucleus accumbens (basal ganglia)	Negative	*CCND2-AS1*	$2.70x10^{-07}$
						Brain - Putamen (basal ganglia)	Negative	*CCND2-AS1*	0.000008
						Brain - Cortex	Negative	*CCND2-AS1*	0.000011
						Muscle - Skeletal	Negative	*CCND2*	$9.30x10^{-13}$
						Adipose - Subcutaneous	Negative	*CCND2*	$2.70x10^{-10}$
13	rs1359790	80717156	*SPRY2*	G	A	NA	NA	NA	NA

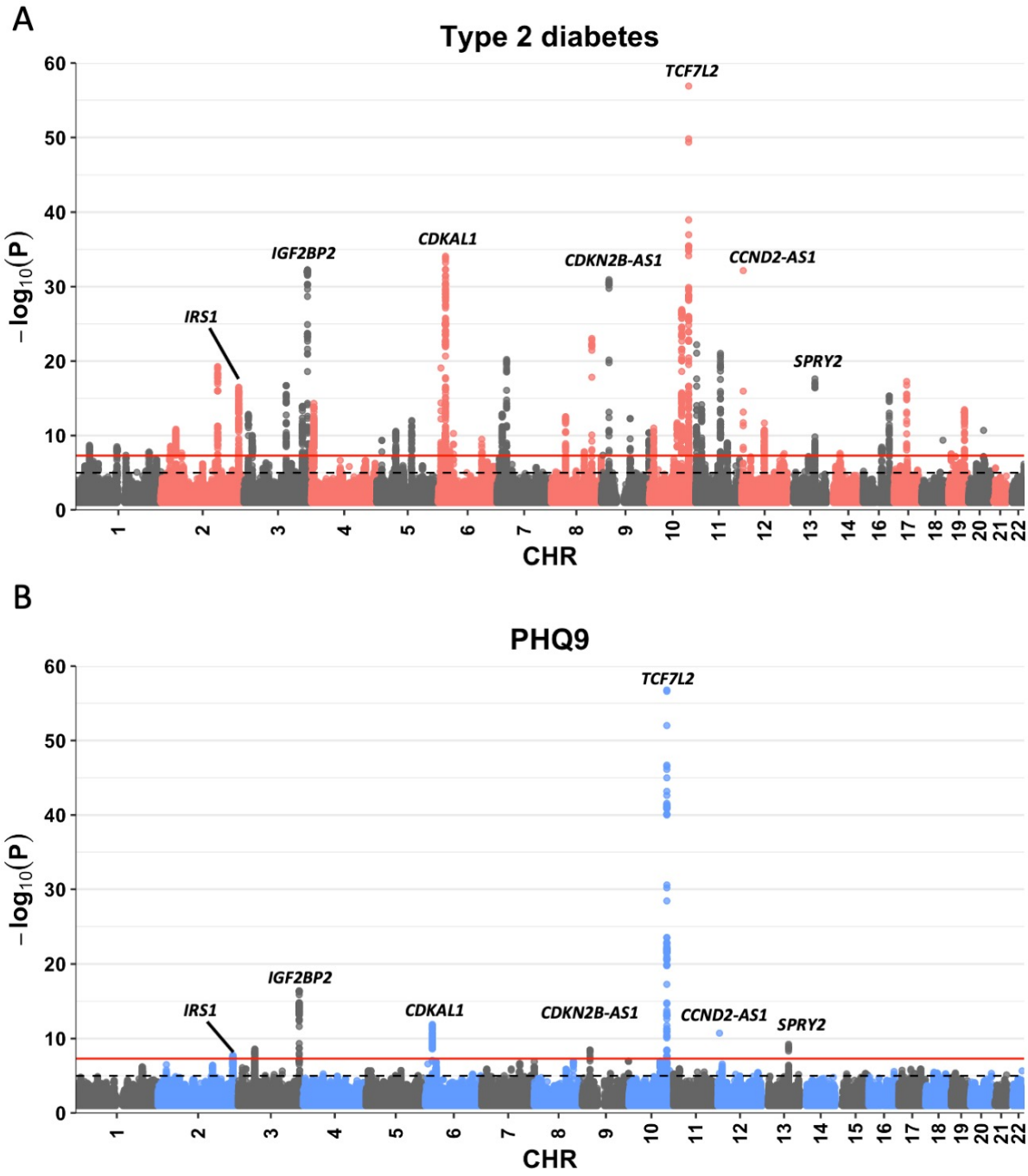
566
567 Legend: CHR=chromosome, BP=base pair position (genome build 37), EA=effect allele, NEA=non-effect
568 allele, EAF=effect allele frequency

569 Figure legends

570 **Figure 1**. Forest plot showing the Mendelian randomization analysis results between
571 depression (exposure) and type 2 diabetes (type 2 diabetes). The odds ratio (OR), their
572 95% confidence intervals and *P*-values are shown.

573 **Figure 2**. Manhattan plots of type 2 diabetes (A) and PHQ-9 (B) after MP-GWAS in
574 UK Biobank. The red horizontal line shows genome-wide significance threshold
575 ($P<5\times10^{-8}$). Grey dashed horizontal lines show suggestive genome-wide significance
576 threshold ($P<1\times10^{-5}$). The shared loci are annotated.


Exposure: Depression
Outcome: Type 2 Diabetes
Inverse variance weighted
Weighted Median
MR Egger
Simple mode
Weighted mode
1.18 [1.06:1.32], p=0.002
1.11 [1:1.23], p=0.043
0.95 [0.5:1.81], p=0.083
0.89 [0.6:1.32], p=0.562
1.29 [0.89:1.87], p=0.184
0.8
1.2
1.6
Odds ratio


A
Type 2 diabetes
TCF7L2
IGF2BP2
CDKAL1
CDKN2B-AS1
CCND2-AS1
IRS1
SPRY2
−log10(P)
CHR
B
PHQ9
TCF7L2
IGF2BP2
IRS1
CDKAL1
CDKN2B-AS1
CCND2-AS1
SPRY2
−log10(P)
CHR

6.1 Insights

This last paper presents a study investigating the genetic relationship between two common disorders: T2D and depression.

Using publicly available GWAS summary statistics of T2D and depression, I assessed the causal relationship between the two diseases using two-sample bi-directional Mendelian randomization. The results of this investigation show that depression is causal for type 2 diabetes while there was no evidence of the reverse direction being significant.

Additionally, using the UKBB, I implemented a multi-phenotype GWAS approach to jointly analyse T2D and depressive phenotypes. The depressive phenotypes included in this study were depressive symptoms based on self-report questionnaires, and clinically diagnosed major depressive disorder (MDD). Multi-phenotype GWAS demonstrated shared genetic *loci* between T2D and self-reported definitions of depression which was not seen in the standard GWAS approach of analysing phenotypes independently. Majority of the identified shared *loci* between T2D and depression have a role in insulin secretion pathways. However, T2D and the strictly MDD did not reveal shared *loci* after multi-phenotype GWAS.

I further sought to establish the target genes associated with both T2D and depression using expression quantitative trait *loci* (eQTL) analysis. Here, I used data from the GTEx and TIGER databases.

From this study I illustrate how genetic determinants that are shared between related traits can be revealed through various statistical genetics methods such as Mendelian randomization and multi-phenotype GWAS approaches. As with the obesity and cancer study, I illustrate how Mendelian randomization utilising publicly available

GWAS summary statistics can aid in the dissection of causal relationships between related traits. In addition, and perhaps most importantly, I illustrate the utility of multi-phenotype GWAS approaches in identifying shared genetic l*oci* between related traits.

7 GENERAL DISCUSSION

In my PhD project, I leveraged on both published GWAS and large-scale biobank data to assess the relationship between two distinct measures of adiposity and breast, prostate, colorectal, pancreatic and lung cancers.

By implementing statistical genetics methods that capitalize on GWAS output, such as genetic correlation, polygenic scores, and Mendelian randomization, I demonstrated that both central and overall measures of obesity relate differently to the risk of certain types of cancers.

Several considerations are currently in place for future research to build on the present work. These considerations include perspectives on the study design as well as improvement on various aspects of the methodology.

The metabolic syndrome and its components, including dyslipidaemia, insulin resistance and hypertension, should also be added to future study designs to further boost our findings surrounding central adiposity and cancer risk. In the current study, the metabolic syndrome is inferred using the WHRadjBMI phenotype. Other potential phenotypes to add to our analyses include fasting glucose levels, HDL cholesterol, systolic and diastolic blood pressure. To build on our conclusions of a significant contribution of the metabolic syndrome in cancer development, it follows logically that assessing the impact of the other components of the metabolic syndrome is needed to corroborate our findings.

The definition of the sex-specific cancer cases could be updated to improve the specificity of phenotypes. This would involve considering tumour heterogeneity by hormone receptor state as well as tumour grade in addition to overall cancer incidence definitions. In the case of breast cancer, oestrogen-receptor(ER) and progesterone-receptor positive breast cancer as well as tumour aggressiveness should be added in the case definition criteria. Similarly, for prostate cancer, ERα- and ERβ status and

tumour aggressiveness could be accounted for. Since hormone receptor status and tumour behaviour characteristics data in UKBB may be limited in both sample size and detail, there exists an opportunity to collaborate with consortia such as BCAC in actualising this study.

In the PGS and MR analyses, the next step would be to implement hierarchical clustering of the obesity SNPs (BMI and WHRadjBMI) to partition these variants into mechanistic groups. These mechanistic groups, based on their effects on the phenotype, would represent hypothesised mechanisms underlying the obesity-cancer mechanisms. The PRS based on these groups would be calculated and tested for each cancer. Likewise, I could apply MR to assess the causality between each of these groups and cancers. Furthermore, implementation of multi-phenotype GWAS approaches, such as those that use individual level data[74], could enable better definition of *loci* with pleiotropic effects, latter to be carefully evaluated for the use or exclusion from MR analyses for specific phenotype relationships[86].

A key notable strength of this PhD is the large number of BMI and WHRadjBMI variants used in both PGS and MR studies. The resultant PGS base data that were therefore of higher quality and our conclusions based on their application are thus credible. Moreover, the UKBB offers a large database with close to 500,000 individuals with both genetic and phenotypic data amenable for analyses. The wide range of phenotypic data in the UKBB also allowed for the investigation of potential confounding factors such as menopause and smoking status.

The interpretation of our findings should consider several limitations. Our analyses were based on European data due to its availability compared to data of other populations. Consequently, generalizability of our results across different populations is not recommended. Further, the lack of information of hormone receptor status in the

UKBB limited the extent to which the associations between obesity and cancer can be performed. The unavailability of public GWAS summary statistics of certain cancers such as post-menopausal breast and lung cancers also limited our ability to investigate causality using MR. While for some cancers in the UKBB, our sample sizes were low and thus GWAS, and subsequent analyses were statistically underpowered.

In conclusion, using large scale genetic data (published GWAS and biobank data), I show that central obesity, proxied using WHRadjBMI, may be a more important causal risk factor for pancreatic cancer than overall obesity. Additionally, I show an inverse association between overall obesity and prostate cancer, while central adiposity has a direct association with prostate cancer. These results additionally suggest that central obesity may be a causal risk factor for breast cancer.

8 REFERENCES

1. Obesity and overweight. https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight.

2. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* **71**, 209–249 (2021).

3. Naghavi, M. *et al.* Global, regional, and national age-sex specifc mortality for 264 causes of death, 1980-2016: A systematic analysis for the Global Burden of Disease Study 2016. *The Lancet* **390**, 1151–1210 (2017).

4. Roche, A. F., Siervogel, R. M., Chumlea, W. C. & Webb, P. Grading body fatness from limited anthropometric data. *Am J Clin Nutr* **34**, 2831–2838 (1981).

5. Alberti, K. G. M. M. *et al.* Harmonizing the Metabolic Syndrome. *Circulation* **120**, 1640–1645 (2009).

6. Geneva. Waist Circumference and Waist-Hip Ratio: Report of a WHO Expert Consultation.

7. Alberti, K. G. M. M. *et al.* Metabolic syndrome—a new world-wide definition. A Consensus Statement from the International Diabetes Federation. *Diabetic Medicine* **23**, 469–480 (2006).

8. Lalia, A. Z. *et al.* Predictors of Whole-Body Insulin Sensitivity Across Ages and Adiposity in Adult Humans. *J Clin Endocrinol Metab* **101**, 626–634 (2016).

9. Scully, T., Ettela, A., LeRoith, D. & Gallagher, E. J. Obesity, Type 2 Diabetes, and Cancer Risk. *Front Oncol* **10**, 3196 (2021).

10. Arnold, M. *et al.* Global burden of cancer attributable to high body-mass index in 2012: A population-based study. *Lancet Oncol* **16**, 36–46 (2015).

11. Calle, E. E., Rodriguez, C., Walker-Thurmond, K. & Thun, M. J. Overweight, Obesity, and Mortality from Cancer in a Prospectively Studied Cohort of U.S. Adults. *New England Journal of Medicine* **348**, 1625–1638 (2003).

12. Renehan, A. G., Tyson, M., Egger, M., Heller, R. F. & Zwahlen, M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *The Lancet* **371**, 569–578 (2008).

13. Bhaskaran, K. *et al.* Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5·24 million UK adults. *The Lancet* **384**, 755–765 (2014).

14. Calle, E. E., Thun, M. J., Petrelli, J. M., Rodriguez, C. & Heath, C. W. Body-Mass Index and Mortality in a Prospective Cohort of U.S. Adults. *New England Journal of Medicine* **341**, 1097–1105 (1999).

15. Canoy, D. *et al.* Cigarette smoking and fat distribution in 21,828 British men and women: a population-based study. *Obes Res* **13**, 1466–1475 (2005).

16. Tang, B. *et al.* Waist-hip Ratio (WHR), a Better Predictor for Prostate Cancer than Body Mass Index (BMI): Results from a Chinese Hospital-based Biopsy Cohort. *Scientific Reports 2017 7:1* **7**, 1–8 (2017).

17. Nagrani, R. *et al.* Central obesity increases risk of breast cancer irrespective of menopausal and hormonal receptor status in women of South Asian Ethnicity. *Eur J Cancer* **66**, 153 (2016).

18. Xia, B. *et al.* Metabolic syndrome and risk of pancreatic cancer: A population-based prospective cohort study. *Int J Cancer* **147**, 3384–3393 (2020).

19. Gallagher, E. J. & LeRoith, D. Obesity and diabetes: The increased risk of cancer and cancer-related mortality. *Physiol Rev* **95**, 727–748 (2015).

20. Dong, Y. *et al.* Abdominal obesity and colorectal cancer risk: systematic review and meta-analysis of prospective studies. *Biosci Rep* **37**, (2017).

21. Belfiore, A., Frasca, F., Pandini, G., Sciacca, L. & Vigneri, R. Insulin receptor isoforms and insulin receptor/insulin-like growth factor receptor hybrids in physiology and disease. *Endocrine Reviews* vol. 30 586–623 Preprint at https://doi.org/10.1210/er.2008-0047 (2009).

22. Goodwin, P. J. *et al.* High insulin levels in newly diagnosed breast cancer patients reflect underlying insulin resistance and are associated with components of the insulin resistance syndrome. *Breast Cancer Res Treat* **114**, 517–525 (2009).

23. Yee, D. Insulin-like growth factor receptor inhibitors: Baby or the bathwater? *Journal of the National Cancer Institute* vol. 104 975–981 Preprint at https://doi.org/10.1093/jnci/djs258 (2012).

24. LeRoith, D., Baserga, R., Helman, L. & Roberts, C. T. Insulin-like growth factors and cancer. in *Annals of Internal Medicine* vol. 122 54–59 (American College of Physicians, 1995).

25. Zhao, H., Dupont, J., Yakar, S., Karas, M. & LeRoith, D. PTEN inhibits cell proliferation and induces apoptosis by downregulating cell surface IGF-IR expression in prostate cancer cells. *Oncogene* **23**, 786–794 (2004).

26. Yakar, S., LeRoith, D. & Brodt, P. The role of the growth hormone/insulin-like growth factor axis in tumor growth and progression: Lessons from animal models. *Cytokine and Growth Factor Reviews* vol. 16 407–420 Preprint at https://doi.org/10.1016/j.cytogfr.2005.01.010 (2005).

27. Sen, S., Langiewicz, M., Jumaa, H. & Webster, N. J. G. Deletion of serine/arginine-rich splicing factor 3 in hepatocytes predisposes to hepatocellular carcinoma in mice. *Hepatology* **61**, 171–183 (2015).

28. Verheus, M. *et al.* Serum C-peptide levels and breast cancer risk: Results from the European prospective investigation into cancer and nutrition (EPIC). *Int J Cancer* **119**, 659–667 (2006).

29. Jenab, M. *et al.* Serum C-peptide, IGFBP-1 and IGFBP-2 and risk of colon and rectal cancers in the European prospective investigation into cancer and nutrition. *Int J Cancer* **121**, 368–376 (2007).

30. Stevens, V. L., Jacobs, E. J., Sun, J. & Gapstur, S. M. No association of plasma levels of adiponectin and c-peptide with risk of aggressive prostate cancer in the cancer prevention study II nutrition cohort. *Cancer Epidemiology Biomarkers and Prevention* **23**, 890–892 (2014).

31. Ma, J. *et al.* Prediagnostic body-mass index, plasma C-peptide concentration, and prostate cancer-specific mortality in men with prostate cancer: a long-term survival analysis. *Lancet Oncol* **9**, 1039–1047 (2008).

32. Lai, G. Y. *et al.* Association of C-peptide and leptin with prostate cancer incidence in the Health Professionals Follow-up Study. *Cancer Causes and Control* **25**, 625–632 (2014).

33. Warburg, O., Wind, F. & Negelein, E. The metabolism of tumors in the body. *Journal of General Physiology* **8**, 519–530 (1927).

34. Adekola, K., Rosen, S. T. & Shanmugam, M. Glucose transporters in cancer metabolism. *Current Opinion in Oncology* vol. 24 650–654 Preprint at https://doi.org/10.1097/CCO.0b013e328356da72 (2012).

35. de Beer, J. C. & Liebenberg, L. Does cancer risk increase with HbA 1c, independent of diabetes? *Br J Cancer* **110**, 2361–2368 (2014).

36. Vlassara, H. & Striker, G. E. Advanced Glycation Endproducts in Diabetes and Diabetic Complications. *Endocrinology and Metabolism Clinics of North America* vol. 42 697–719 Preprint at https://doi.org/10.1016/j.ecl.2013.07.005 (2013).

37. Rojas, A. Diabetes and cancer: Looking at the multiligand/RAGE axis. *World J Diabetes* **2**, 108 (2011).

38. Melvin, J. C., Holmberg, L., Rohrmann, S., Loda, M. & van Hemelrijck, M. Serum lipid profiles and cancer risk in the context of obesity: Four meta-analyses. *Journal of Cancer Epidemiology* vol. 2013 Preprint at https://doi.org/10.1155/2013/823849 (2013).

39. Chang, N. W. *et al.* Influences of apolipoprotein E polymorphism on the risk for breast cancer and HER2/neu status in Taiwan. *Breast Cancer Res Treat* **90**, 257–261 (2005).

40. Hea, Y. O. *et al.* Cholesterol level of lipid raft microdomains regulates apoptotic cell death in prostate cancer cells through EGFR-mediated Akt and ERK signal transduction. *Prostate* **67**, 1061–1069 (2007).

41. Scheinman, E. J., Rostoker, R. & LeRoith, D. Cholesterol affects gene expression of the Jun family in colon carcinoma cells using different signaling pathways. *Mol Cell Endocrinol* **374**, 101–107 (2013).

42. Alikhani, N. *et al.* Mammary tumor growth and pulmonary metastasis are enhanced in a hyperlipidemic mouse model. *Oncogene* **32**, 961–967 (2013).

43. Montgomery, R. B. *et al.* Maintenance of intratumoral androgens in metastatic prostate cancer: A mechanism for castration-resistant tumor growth. *Cancer Res* **68**, 4447–4454 (2008).

44. Pelton, K., Freeman, M. R. & Solomon, K. R. Cholesterol and prostate cancer. *Current Opinion in Pharmacology* vol. 12 751–759 Preprint at https://doi.org/10.1016/j.coph.2012.07.006 (2012).

45. Ouchi, N., Parker, J. L., Lugus, J. J. & Walsh, K. Adipokines in inflammation and metabolic disease. *Nature Reviews Immunology* vol. 11 85–97 Preprint at https://doi.org/10.1038/nri2921 (2011).

46. Garofalo, C. & Surmacz, E. Leptin and cancer. *Journal of Cellular Physiology* vol. 207 12–22 Preprint at https://doi.org/10.1002/jcp.20472 (2006).

47. Ziemke, F. & Mantzoros, C. S. Adiponectin in insulin resistance: Lessons from translational research. in *American Journal of Clinical Nutrition* vol. 91 (Am J Clin Nutr, 2010).

48. Steppan, C. M. *et al.* The hormone resistin links obesity to diabetes. *Nature* **409**, 307–312 (2001).

49. Harford, K. A., Reynolds, C. M., McGillicuddy, F. C. & Roche, H. M. Fats, inflammation and insulin resistance: Insights to the role of macrophage and T-cell accumulation in adipose tissue. *Proceedings of the Nutrition Society* **70**, 408–417 (2011).

50. Kelesidis, T., Kelesidis, I., Chou, S. & Mantzoros, C. S. Narrative Review: The Role of Leptin in Human Physiology: Emerging Clinical Applications. *Ann Intern Med* **152**, 93 (2010).

51. Surmacz, E. Leptin and adiponectin: Emerging therapeutic targets in breast cancer. *J Mammary Gland Biol Neoplasia* **18**, 321–332 (2013).

52. Sweeney, G. Leptin signalling. *Cellular Signalling* vol. 14 655–663 Preprint at https://doi.org/10.1016/S0898-6568(02)00006-2 (2002).

53. Jardé, T., Perrier, S., Vasson, M. P. & Caldefie-Chézet, F. Molecular mechanisms of leptin and adiponectin in breast cancer. *Eur J Cancer* **47**, 33–43 (2011).

54. Guo, S., Liu, M. & Gonzalez-Perez, R. R. Role of Notch and its oncogenic signaling crosstalk in breast cancer. *Biochimica et Biophysica Acta - Reviews on Cancer* vol. 1815 197–213 Preprint at https://doi.org/10.1016/j.bbcan.2010.12.002 (2011).

55. Schäffler, A., Schölmerich, J. & Buechler, C. Mechanisms of disease: Adipokines and breast cancer - Endocrine and paracrine mechanisms that connect adiposity and breast cancer. *Nature Clinical Practice Endocrinology and Metabolism* vol. 3 345–354 Preprint at https://doi.org/10.1038/ncpendmet0456 (2007).

56. Kim, H. J. *et al.* Expression of resistin in the prostate and its stimulatory effect on prostate cancer cell proliferation. *BJU Int* **108**, (2011).

57. Mantovani, A., Allavena, P., Sica, A. & Balkwill, F. Cancer-related inflammation. *Nature* vol. 454 436–444 Preprint at https://doi.org/10.1038/nature07205 (2008).

58. Ozaki, K. *et al.* Functional SNPs in the lymphotoxin-α gene that are associated with susceptibility to myocardial infarction. *Nature Genetics 2002 32:4* **32**, 650–654 (2002).

59. Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body mass index in ∼700000 individuals of European ancestry. *Hum Mol Genet* **27**, 3641–3649 (2018).

60. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* (2015) doi:10.1038/nature14177.

61. Pulit, S. L. *et al.* Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum Mol Genet* **28**, 166–174 (2019).

62. Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature 2015 518:7538* **518**, 187–196 (2015).

63. Cancer. https://www.who.int/news-room/fact-sheets/detail/cancer.

64. Zhang, H. *et al.* Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. *Nature Genetics 2020 52:6* **52**, 572–581 (2020).

65. Schumacher, F. R. *et al.* Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat Genet* **50**, 928–936 (2018).

66. Law, P. J. *et al.* Association analyses identify 31 new risk loci for colorectal cancer susceptibility. *Nat Commun* **10**, 1–15 (2019).

67. Klein, A. P. *et al.* Genome-wide meta-analysis identifies five new susceptibility loci for pancreatic cancer. *Nature Communications 2018 9:1* **9**, 1–11 (2018).

68. McKay, J. D. *et al.* Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. *Nature Genetics 2017 49:7* **49**, 1126–1132 (2017).

69. Visscher, P. M., Hill, W. G. & Wray, N. R. Heritability in the genomics era — concepts and misconceptions. *Nature Reviews Genetics 2008 9:4* **9**, 255–266 (2008).

70. Tenesa, A. & Haley, C. S. The heritability of human disease: estimation, uses and abuses. *Nature Reviews Genetics 2013 14:2* **14**, 139–149 (2013).

71. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature 2009 461:7265* **461**, 747–753 (2009).

72. Maher, B. Personal genomes: The case of the missing heritability. *Nature* **456**, 18–21 (2008).

73. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics 2010 42:7* **42**, 565–569 (2010).

74. Mägi, R. *et al.* SCOPA and META-SCOPA: software for the analysis and aggregation of genome-wide association studies of multiple correlated phenotypes. *BMC Bioinformatics* **18**, 25 (2017).

75. Turley, P. *et al.* Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat Genet* **50**, 229–237 (2018).

76. Kaakinen, M. *et al.* MARV: a tool for genome-wide multi-phenotype analysis of rare variants. *BMC Bioinformatics* **18**, 110 (2017).

77. Lewis, C. M. & Vassos, E. Polygenic risk scores: From research tools to clinical instruments. *Genome Med* **12**, 1–11 (2020).

78. Choi, S. W., Mak, T. S. H. & O'Reilly, P. F. Tutorial: a guide to performing polygenic risk score analyses. *Nature Protocols 2020 15:9* **15**, 2759–2772 (2020).

79. Willer, C. J. *et al.* Discovery and refinement of loci associated with lipid levels. *Nat Genet* **45**, 1274–1285 (2013).

80. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics* (2007) doi:10.1086/519795.

81. Vilhjálmsson, B. J. *et al.* Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am J Hum Genet* **97**, 576–592 (2015).

82. Choi, S. W. & O'Reilly, P. F. PRSice-2: Polygenic Risk Score software for biobank-scale data. *Gigascience* **8**, 1–6 (2019).

83. Verbanck, M., Chen, C. Y., Neale, B. & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nature Genetics 2018 50:5* **50**, 693–698 (2018).

84. Lawlor, D. A., Harbord, R. M., Sterne, J. A. C., Timpson, N. & Smith, G. D. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Stat Med* **27**, 1133–1163 (2008).

85. Pierce, B. L. & Burgess, S. Efficient Design for Mendelian Randomization Studies: Subsample and 2-Sample Instrumental Variable Estimators. *Am J Epidemiol* **178**, 1177–1184 (2013).

86. Hemani, G. *et al.* The MR-base platform supports systematic causal inference across the human phenome. *Elife* **7**, (2018).

87. Gao, C. *et al.* Mendelian randomization study of adiposity-related traits and risk of breast, ovarian, prostate, lung and colorectal cancer. *Int J Epidemiol* **45**, 896–908 (2016).

88. Heid, I. M. *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* **42**, 949–960 (2010).

89. Bulik-Sullivan, B. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics 2015 47:3* **47**, 291–295 (2015).

90. Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S. M. & Smoller, J. W. Pleiotropy in complex traits: challenges and strategies. *Nature Reviews Genetics 2013 14:7* **14**, 483–495 (2013).

91. Loh, P. R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nature Genetics 2015 47:3* **47**, 284–290 (2015).

9 SUPPLEMENTARY TABLES AND FIGURES

List of supplementary tables

Supplementary Table 1. Detailed results of the Mendelian randomization analyses between obesity and cancer phenotypes

Supplementary Table 2. Independent genome-wide significant signals (P<5x10^{-8}) for overall breast cancer in UK Biobank GWAS

Supplementary Table 3. Independent genome-wide significant signals (P<5x10^{-8}) for post-menopausal breast cancer in UK Biobank GWAS

Supplementary Table 4. Independent genome-wide significant signals (P<5x10^{-8}) for prostate cancer in UK Biobank GWAS

Supplementary Table 5. Independent genome-wide significant signals (P<5x10^{-8}) for colorectal cancer in UK Biobank GWAS

Supplementary Table 6. Independent genome-wide significant signals (P<5x10^{-8}) for pancreatic cancer in UK Biobank GWAS

Supplementary Table 7. Independent genome-wide significant signals (P<5x10^{-8}) for lung cancer in UK Biobank GWAS

List of supplementary figures

Supplementary Figure 1. Flowchart showing UK Biobank adiposity and pancreatic cancer definition

Supplementary Figure 2. Manhattan plot of lung cancer GWAS in UK Biobank. The red horizontal line shows genome-wide significance threshold ($P<5x10^{-8}$). The dashed grey line shows suggestive significance threshold ($P<1x10^{-5}$)

Supplementary Figure 3. Manhattan plots of overall breast cancer (top) and post-menopausal breast cancer (bottom) GWAS in UK Biobank. The red horizontal line shows genome-wide significance threshold ($P<5x10^{-8}$). The dashed grey line shows suggestive significance threshold ($P<1x10^{-5}$)

Supplementary Figure 4. Manhattan plots of prostate cancer (top) and colorectal cancer (bottom) GWAS in UK Biobank. The red horizontal line shows genome-wide significance threshold ($P<5x10^{-8}$). The dashed grey line shows suggestive significance threshold ($P<1x10^{-5}$)

Supplementary Figure 5. Manhattan plot of lung cancer GWAS in UK Biobank. The red horizontal line shows genome-wide significance threshold ($P<5x10^{-8}$). The dashed grey line shows suggestive significance threshold ($P<1x10^{-5}$)

Supplementary Figure 6. Manhattan plots of BMI (top) and WHRadjBMI cancer (bottom) GWAS in UK Biobank. The red horizontal line shows genome-wide significance threshold ($P<5x10^{-8}$)

Supplementary Figure 7. Summary of Mendelian Randomization tests performed for BMI/WHRadjBMI and breast, prostate and colorectal cancers, including the number of genetic instruments (SNPs) available from published GWAS for each test

Supplementary Figure 8. Scatter and forest plots for the BMI to cancer direction Mendelian randomization tests. A) BMI to breast cancer. B) BMI to colorectal cancer

Supplementary Figure 9. Scatter and forest plots for the cancer to BMI direction Mendelian randomization tests. A) Breast cancer to BMI. B) Colorectal cancer to BMI. C) Prostate cancer to BMI

Supplementary Figure 10. Scatter and forest plots for the WHRadjBMI to cancer direction Mendelian randomization tests. A) WHRadjBMI to breast cancer. B) WHRadjBMI to colorectal cancer

Supplementary Figure 11. Scatter and forest plots for the cancer to WHRadjBMI direction Mendelian randomization tests. A) Breast cancer to WHRadjBMI. B) Colorectal cancer to WHRadjBMI. C) Prostate cancer to WHRadjBMI

Supplementary Table 1. Detailed results of the Mendelian randomization analyses between obesity and cancer phenotypes

			Inverse variance weighted		MR Egger		Weighted median		Simple mode		Weighted mode		Heterogeneity	MR-Egger Intercept	
Exposure	**Outcome**	**NSNPs**	**OR (95% CI)**	**P**	**OR (95% CI)**	**P**	**OR (95% CI)**	**P**	**OR (95% CI)**	**P**	**OR (95% CI)**	**P**	**Q stat (P)**	**Intercept(SE)**	**P**
BMI	BrC	576	1.000 (0.995-1.005)	0.897	0.985 (0.971-1.000)	0.051	0.996 (0.988-1.003)	0.266	0.944 (0.968-1.020)	0.634	0.994 (0.977-1.011)	0.462	755.9 (5.34E-07)	0.0003 (0.0001)	0.034
BMI	PrC	574	0.993 (0.988-0.998)	0.0042	0.995 (0.982-1.009)	0.473	0.993 (0.985-0.999)	0.039	0.984 (0.960-1.009)	0.22	0.995 (0.981-1.009)	0.491	863.64 (3.93E-14)	-45.47 (0.0001)	0.713
BMI	CrC	575	1.000 (0.998-1.002)	0.92	1.001 (0.996-1.006)	0.682	1.000 (0.997-1.003)	1	0.999 (0.989-1.008)	0.768	1.000 (0.995-1.005)	0.923	656.59 (0.0094)	-20.71 (0.00004)	0.689
BrC	BMI	109	0.997 (0.985-1.008)	0.557	1.004 (0.984-1.025)	0.686	1.008 (0.998-1.017)	0.1	1.030 (0.061-17.321)	0.983	1.030 (0.076-13.925)	0.982	510.98 (1.27E-53)	-0.006 (0.0007)	0.372
PrC	BMI	74	1.004 (0.995-1.014)	0.333	0.993 (0.974-1.102)	0.471	0.998 (0.988-1.007)	0.639	0.995 (0977-1.013)	0.57	0.997 (0.987-1.008)	0.626	194.16 (6.04E-13)	0.001 1 (0.0008)	0.184
CrC	BMI	48	0.752 (0.368-1.538)	0.435	1.648 (0.352-7.729)	0.529	1.031 (0.473-2.246)	0.939	1.153 (0.021-62.240)	0.944	1.153 (0.035-38.392)	0.937	102.89 (4.71E-06)	-0.001 (0.0009)	0.268
WHRadjBMI	BrC	284	0.990 (0.983-0.997)	0.0068	1.000 (0.982-1.017)	0.974	0.991 (0.981-1.002)	0.105	1.003 (0.974-1.033)	0.83	0.993 (0.978-1.008)	0.338	529.41 (3.63E-17)	-0.0002 (0.0002)	0.226
WHRadjBMI	PrC	284	1.0046 (0.998-1.011)	0.179	1.016 (1.00018-1.032)	0.048	1.007 (0.999-1.016)	0.094	1.018 (0.990-1.045)	0.209	1.022 (1.00067-1.038)	0.0053	493.63 (1.32E-13)	-0.0002 (0.0002)	0.119
WHRadjBMI	CrC	284	1.002 (0.994-1.004)	0.125	1.000 (0.995-1.006)	0.885	1.000 (0.996-1.004)	0.917	0.995 (0.984-1.006)	0.391	1.000 (0.994-1.006)	0.988	410.68 (1.02E-06)	0.00003 (0.00006)	0.578

BrC	WHRadjBMI	117	0.993 (0.975-1.012)	0.495	0.949 (0.910-0.991)	0.018	0.976 (0.953-1.001)	0.056	1.013 (0.002-518.382)	0.997	1.013 (0.004-274.794)	0.997	180.81 (1.11E-04)	0.003 (0.001)	0.022
PrC	WHRadjBMI	81	1.008 (0.996-1.021)	0.211	1.013 (0.988-1.039)	0.324	1.005 (0.985-1.026)	0.616	0.999 (0.961-1.038)	0.947	1.005 (0.984-1.025)	0.655	79.36 (0.499)	-0.0005 (0.001)	0.668
CrC	WHRadjBMI	59	1.113 (0.362-3.423)	0.852	11.000 (1.176-102.88)	0.04	1.728 (0.314-9.509)	0.53	2.546 (0.016-40.531)	0.511	2.126 (0.342-13.222)	0.422	68.69 (0.159)	-0.003 (0.001)	0.025

Supplementary Table 2. Independent genome-wide significant signals (P<5x10^{-8}) for overall breast cancer in UK Biobank GWAS

CHR	BP	SNP	GENE	EA	NEA	EAF	BETA	SE	P
1	121280613	rs11249433	*EMBP1*	A	G	0.583	-0.0062	0.00075	1.50x10^{-16}
1	149927034	rs12048493	*OTUD7B*	A	C	0.609	-0.0045	0.00078	6.90x10^{-09}
2	121153979	rs13406182	*INHBB,LINC01101*	T	C	0.806	0.0057	0.00094	1.30x10^{-09}
2	121245613	rs12616849	*LINC01101,GLI2*	G	C	0.098	-0.0070	0.00126	2.50x10^{-08}
2	213537460	rs9967727	*ERBB4,LINC01878*	C	G	0.634	-0.0047	0.00077	8.10x10^{-10}
2	217920769	rs4442975	*LINC01921,DIRC3-AS1*	G	T	0.489	0.0087	0.00074	1.00x10^{-31}
2	217954982	rs7587558	*LINC01921,DIRC3-AS1*	T	A	0.965	0.0171	0.00205	6.40x10^{-17}
3	27374101	rs1352944	*NEK10*	C	A	0.525	0.0073	0.00074	9.80x10^{-23}
4	175850605	rs28465148	*ADAM29*	T	G	0.880	0.0071	0.00115	5.60x10^{-10}
5	1294086	rs2736098	*TERT*	C	T	0.721	0.0048	0.00083	6.80x10^{-09}
5	44706498	rs10941679	*LINC02224,BRCAT54*	A	G	0.747	-0.0088	0.00086	2.80x10^{-24}
5	45333860	rs55821517	*HCN1*	T	C	0.736	0.0052	0.00086	9.70x10^{-10}
5	56016918	rs12653202	*C5orf67,MAP3K1*	A	C	0.841	-0.0122	0.00102	3.30x10^{-33}
5	158230013	rs11135046	*EBF1*	G	T	0.457	0.0052	0.00075	3.00x10^{-12}
6	151947326	rs11155805	*CCDC170,ESR1*	A	G	0.669	-0.0072	0.00079	6.20x10^{-20}
6	152441587	rs2813550	*ESR1*	C	A	0.242	-0.0050	0.00087	7.00x10^{-09}
8	36859186	rs12681990	*KCNU1*	T	C	0.838	0.0072	0.00101	9.90x10^{-13}
8	128355618	rs13281615	*CASC21,CASC8*	A	G	0.591	-0.0065	0.00075	4.70x10^{-18}
9	110306944	rs10978911	*KLF4*	G	C	0.875	-0.0067	0.00113	2.90x10^{-09}
9	110837073	rs10816625	*KLF4*	A	G	0.936	-0.0087	0.00152	1.20x10^{-08}
9	110893030	rs628931	*KLF4*	A	G	0.377	-0.0063	0.00077	1.70x10^{-16}
10	21799726	rs12256551	*SKIDA1*	A	C	0.644	-0.0044	0.00078	2.10x10^{-08}
10	64258343	rs2393886	*ZNF365*	C	T	0.533	0.0046	0.00075	5.40x10^{-10}
10	80887957	rs10762851	*ZMIZ1*	A	G	0.841	-0.0067	0.00102	3.70x10^{-11}
10	123095209	rs9421410	*WDR11,FGFR2*	G	A	0.684	0.0052	0.00080	1.10x10^{-10}
10	123314462	rs17614209	*FGFR2*	C	G	0.976	-0.0152	0.00254	2.60x10^{-09}
10	123346116	rs2981575	*FGFR2*	G	A	0.396	0.0179	0.00076	3.80x10^{-123}
11	1902097	rs4980383	*LSP1*	C	T	0.547	-0.0056	0.00075	1.00x10^{-13}
11	69331418	rs78540526	*CCND1*	C	T	0.929	-0.0211	0.00145	3.60x10^{-48}
11	129454107	rs10736577	*BARX2*	A	G	0.395	-0.0047	0.00076	7.70x10^{-10}
12	28151609	rs812020	*PTHLH,CCDC91*	A	C	0.737	0.0059	0.00085	5.00x10^{-12}
12	28488886	rs11049539	*CCDC91*	A	T	0.698	0.0045	0.00081	3.40x10^{-08}
12	96026737	rs61938093	*PGAM1P5*	C	T	0.705	0.0075	0.00082	4.60x10^{-20}
12	115834946	rs2133317	*TBX3,MED13L*	C	G	0.615	0.0052	0.00076	1.10x10^{-11}
14	37128564	rs34914085	*PAX9*	C	A	0.789	0.0055	0.00091	1.70x10^{-09}
14	68979835	rs11624333	*RAD51B*	T	C	0.717	0.0062	0.00083	1.40x10^{-13}
16	52599188	rs4784227	*CASC16*	C	T	0.760	-0.0166	0.00087	3.00x10^{-81}
16	53810686	rs7193144	*FTO*	T	C	0.607	0.0056	0.00076	1.80x10^{-13}
16	53861592	rs7184573	*FTO*	G	A	0.624	0.0042	0.00077	5.70x10^{-08}
16	54676323	rs8044756	*LINC02140,LOC101927480*	G	A	0.487	-0.0044	0.00075	6.10x10^{-09}
16	80651109	rs17750740	*CDYL2*	T	C	0.795	-0.0052	0.00093	1.70x10^{-08}
17	29206421	rs6505216	*ATAD5*	G	T	0.766	0.0059	0.00091	8.60x10^{-11}

18	24481272	rs17621185	*AQP4-AS1*	A	G	0.789	0.0051	0.00091	1.60×10^{-08}
21	16563640	rs2823129	*NRIP1,USP25*	C	T	0.675	0.0052	0.00079	4.40×10^{-11}
22	40935593	rs183387906	*MKL1*	G	A	0.912	-0.0096	0.00134	8.60×10^{-13}
22	41027870	rs73169097	*MKL1*	C	T	0.901	-0.0101	0.00125	5.80×10^{-16}

Supplementary Table 3. Independent genome-wide significant signals (P<5x10^{-8}) for post-menopausal breast cancer in UK Biobank GWAS

CHR	BP	SNP	GENE	EA	NEA	EAF	BETA	SE	P
1	121280613	rs11249433	*EMBP1*	A	G	0.584	-0.0042	0.00064	3.90x10^{-11}
1	149927034	rs12048493	*OTUD7B*	A	C	0.609	-0.0040	0.00066	1.80x10^{-09}
2	121154536	rs72960863	*INHBB,LINC01101*	T	C	0.806	0.0047	0.00080	5.60x10^{-09}
2	213545357	rs13404902	*ERBB4,LINC01878*	C	T	0.639	-0.0037	0.00066	3.00x10^{-08}
2	217905779	rs13412666	*LINC01921,DIRC3-AS1*	G	A	0.502	0.0059	0.00063	1.40x10^{-20}
2	217957699	rs72951831	*LINC01921,DIRC3-AS1*	G	T	0.964	0.0111	0.00169	6.30x10^{-11}
3	27374101	rs1352944	*NEK10*	C	A	0.525	0.0049	0.00063	2.00x10^{-14}
5	1294086	rs2736098	*TERT*	C	T	0.721	0.0041	0.00071	1.00x10^{-08}
5	44706498	rs10941679	*LINC02224,BRCAT54*	A	G	0.747	-0.0062	0.00073	2.00x10^{-17}
5	56016918	rs12653202	*C5orf67,MAP3K1*	A	C	0.841	-0.0088	0.00087	3.70x10^{-24}
5	158230013	rs11135046	*EBF1*	G	T	0.457	0.0038	0.00064	3.00x10^{-09}
6	151969740	rs9371545	*CCDC170,ESR1*	G	A	0.926	-0.0084	0.00121	4.40x10^{-12}
6	152432902	rs910416	*ESR1*	C	T	0.490	-0.0050	0.00064	4.00x10^{-15}
8	36858483	rs13365225	*KCNU1*	A	G	0.836	0.0046	0.00086	7.50x10^{-08}
8	128355618	rs13281615	*CASC21,CASC8*	A	G	0.591	-0.0050	0.00064	6.40x10^{-15}
9	110886840	rs548980	*KLF4*	C	T	0.379	-0.0042	0.00065	1.10x10^{-10}
10	64299890	rs10995201	*ZNF365*	A	G	0.852	0.0052	0.00090	8.60x10^{-09}
10	123346116	rs2981575	*FGFR2*	G	A	0.396	0.0125	0.00065	9.60x10^{-84}
11	69331418	rs78540526	*CCND1*	C	T	0.929	-0.0146	0.00123	1.80x10^{-32}
11	129476405	rs7119897	*BARX2*	C	G	0.430	-0.0040	0.00064	3.60x10^{-10}
12	28139846	rs805510	*PTHLH,CCDC91*	T	C	0.128	-0.0057	0.00095	2.80x10^{-09}
12	96026737	rs61938093	*PGAM1P5*	C	T	0.705	0.0054	0.00070	1.50x10^{-14}
12	115836183	rs1391720	*TBX3,MED13L*	G	A	0.582	0.0042	0.00064	5.60x10^{-11}
14	68976059	rs36028293	*RAD51B*	G	A	0.720	0.0040	0.00071	1.80x10^{-08}
16	52599188	rs4784227	*CASC16*	C	T	0.760	-0.0116	0.00074	3.50x10^{-55}
16	53810686	rs7193144	*FTO*	T	C	0.607	0.0043	0.00065	3.80x10^{-11}
16	80651109	rs17750740	*CDYL2*	T	C	0.795	-0.0051	0.00079	9.40x10^{-11}
22	40935593	rs183387906	*MKL1*	G	A	0.912	-0.0076	0.00115	3.90x10^{-11}
22	41015883	rs5995881	*MKL1*	A	G	0.901	-0.0079	0.00106	9.60x10^{-14}

Supplementary Table 4. Independent genome-wide significant signals (P<5x10^{-8}) for prostate cancer in UK Biobank GWAS

CHR	BP	SNP	GENE	EA	NEA	EAF	BETA	SE	P
1	150940625	rs267738	*CERS2*	T	G	0.7804	-0.00481	0.00085	1.60x10^{-08}
1	204466176	rs4951076	*MDM4*	G	A	0.316	-0.00461	0.00076	1.10x10^{-09}
2	43738173	rs1038822	*THADA*	T	C	0.299	0.00561	0.00077	4.00x10^{-13}
2	62766723	rs11904315	*TMEM17,EHBP1*	C	A	0.889	-0.00675	0.00112	1.70x10^{-09}
2	63277843	rs58235267	*OTX1*	C	G	0.512	-0.00673	0.00071	2.10x10^{-21}
2	63443276	rs141301592	*WDPCP*	C	G	0.856	0.00552	0.00101	4.00x10^{-08}
2	85788175	rs7568458	*GGCX*	T	A	0.545	0.00473	0.00071	2.00x10^{-11}
2	173309402	rs80353656	*ITGA6*	T	C	0.938	0.01297	0.00146	7.00x10^{-19}
2	173363917	rs7596665	*ITGA6*	A	G	0.943	0.01201	0.00152	2.40x10^{-15}
2	202151163	rs3769818	*CASP8*	A	G	0.269	-0.00458	0.00079	8.00x10^{-09}
2	238389739	rs73098849	*COL6A3,MLPH*	G	A	0.833	-0.00515	0.00094	4.90x10^{-08}
3	87144004	rs139263101	*LINC00506*	C	T	0.933	-0.01122	0.00143	4.00x10^{-15}
3	113300183	rs2271494	*SIDT1*	A	T	0.582	0.00440	0.00071	6.90x10^{-10}
3	127898501	rs2811476	*EEFSEC*	A	C	0.737	-0.00437	0.000799	4.50x10^{-08}
3	170083629	rs61436251	*SKIL*	C	G	0.80002	0.00905	0.00088	8.50x10^{-25}
4	95530464	rs12639980	*PDLIM5*	C	A	0.577	-0.00495	0.00071	3.50x10^{-12}
4	106065308	rs10007915	*TET2*	C	G	0.624	0.00643	0.00073	9.50x10^{-19}
5	1282414	rs7725218	*TERT*	G	A	0.661	0.00619	0.00074	7.60x10^{-17}
5	1891821	rs10866528	*CTD-2194D22.4*	A	G	0.536	-0.00627	0.00071	1.30x10^{-18}
6	31080471	rs1265052	*C6orf15*	T	C	0.476	0.00411	0.000704	5.40x10^{-09}
6	31783507	rs1043618	*HSPA1A*	G	C	0.621	0.00401	0.00072	3.10x10^{-08}
6	32628361	rs9273501	*HLA-DQB1-AS1*	T	A	0.639	-0.00405	0.00073	3.20x10^{-08}
6	41548755	rs6917270	*FOXP4*	A	G	0.720	-0.00565	0.00078	4.70x10^{-13}
6	117207682	rs339327	*RFX6*	A	G	0.695	0.00666	0.00076	2.40x10^{-18}
6	160581374	rs651164	*SLC22A1,SLC22A2*	A	G	0.297	-0.00421	0.00077	4.20x10^{-08}
6	160835192	rs1112444	*SLC22A3*	C	A	0.695	-0.00501	0.00076	5.30x10^{-11}
7	27976563	rs10486567	*JAZF1*	G	A	0.767	0.00713	0.00083	7.60x10^{-18}
7	97773812	rs11768309	*LMTK2*	C	A	0.464	0.00606	0.000704	7.10x10^{-18}
8	23466880	rs4383983	*NKX3-1*	G	C	0.583	0.00481	0.00071	1.40x10^{-11}
8	23525543	rs13265330	*NKX3-1*	C	T	0.420	0.00719	0.00071	5.70x10^{-24}
8	127924563	rs10441523	*FAM84B,PCAT1*	C	T	0.317	0.00571	0.00076	5.20x10^{-14}
8	128030236	rs144828524	*PCAT1*	T	C	0.969	0.01140	0.00206	3.20x10^{-08}
8	128077146	rs77541621	*PCAT1,PCAT2*	G	A	0.971	-0.04196	0.00219	1.10x10^{-81}
8	128091418	rs72725868	*PCAT2*	A	G	0.913	0.00753	0.00131	8.20x10^{-09}
8	128110814	rs17765137	*PRNCR1,CASC19*	A	G	0.950	0.01232	0.00161	2.20x10^{-14}
8	128117736	rs143368544	*PRNCR1,CASC19*	C	T	0.977	-0.01346	0.00240	2.00x10^{-08}
8	128324147	rs382434	*CASC21,CASC8*	C	T	0.664	0.00810	0.00075	1.80x10^{-27}
8	128409232	rs11985829	*CASC8*	T	C	0.309	0.00841	0.00076	1.90x10^{-28}
8	128444775	rs150869774	*CASC8*	T	C	0.984	-0.01858	0.00284	6.20x10^{-11}
8	128532137	rs10090154	*CASC8,CASC11*	T	C	0.0996	0.02022	0.00117	1.60x10^{-66}
8	128540776	rs12549761	*CASC8,CASC11*	C	G	0.879	0.01087	0.00107	4.20x10^{-24}

10	47554493	rs7923435	*FAM35DP,ANTXRLP1*	A	C	0.663	-0.00569	0.00076	5.00×10^{-14}
10	51549496	rs10993994	*TIMM23B*	T	C	0.392	0.01100	0.00072	6.90×10^{-53}
10	122798413	rs2420906	*WDR11,FGFR2*	A	G	0.409	0.00418	0.00071	4.80×10^{-09}
10	126710654	rs4109292	*CTBP2*	G	A	0.506	0.00405	0.000704	8.60×10^{-09}
11	2234690	rs10840606	*MIR4686,ASCL2*	A	G	0.822	-0.01133	0.00093	4.30×10^{-34}
11	68870516	rs2924534	*TPCN2,LOC338694*	C	G	0.153	0.00624	0.00098	1.90×10^{-10}
11	68981359	rs12275055	*LOC338694,MYEOV*	A	G	0.829	-0.01335	0.00093	2.00×10^{-46}
11	69462856	rs3862792	*CCND1*	C	T	0.973	-0.01394	0.0022	2.60×10^{-10}
11	102396607	rs12285347	*MMP7*	T	C	0.547	0.00434	0.00071	7.60×10^{-10}
11	125144426	rs117015177	*PKNOX2*	A	G	0.982	-0.01580	0.00272	6.60×10^{-09}
12	53303331	rs17120257	*KRT8*	T	C	0.887	-0.01008	0.00112	1.90×10^{-19}
12	133067473	rs28435470	*FBRSL1*	G	A	0.333	0.00431	0.00075	7.70×10^{-09}
13	73714290	rs7996468	*KLF5,LINC00392*	C	T	0.210	0.00563	0.00087	9.00×10^{-11}
14	53418339	rs7158115	*FERMT2*	G	C	0.863	0.00557	0.00102	5.30×10^{-08}
16	1577184	rs12599859	*IFT140*	A	G	0.872	0.00591	0.00105	2.00×10^{-08}
17	618965	rs684232	*VPS53*	T	C	0.645	-0.00496	0.00073	1.40×10^{-11}
17	7803552	rs7224399	*CHD3*	C	T	0.940	-0.01095	0.00149	1.70×10^{-13}
17	36080165	rs17138469	*HNF1B*	G	C	0.824	0.00684	0.00093	1.70×10^{-13}
17	36099952	rs10908278	*HNF1B*	T	A	0.485	-0.01092	0.00071	1.30×10^{-53}
17	46810586	rs145922598	*HOXB13,TTLL6*	C	T	0.968	-0.01200	0.002001	2.00×10^{-09}
17	47459877	rs76778410	*LOC102724596,PHB*	A	T	0.918	-0.00816	0.00129	2.20×10^{-10}
17	69104938	rs7222314	*CASC17*	A	G	0.459	0.00770	0.00071	1.30×10^{-27}
19	38740925	rs12981216	*PPP1R14A*	C	T	0.536	0.00577	0.000705	2.60×10^{-16}
19	42020112	rs12610088	*PCAT19,LINC01480*	C	T	0.378	-0.00448	0.00072	6.00×10^{-10}
19	51342357	rs12975062	*LOC105372441*	G	A	0.802	-0.00603	0.00090	1.80×10^{-11}
19	51362537	rs62113214	*KLK3*	T	G	0.927	0.01740	0.00135	3.10×10^{-38}
20	62272411	rs6423444	*STMN3*	G	A	0.689	0.00481	0.00077	5.20×10^{-10}
22	43500212	rs5759167	*TTLL1,BIK*	G	T	0.502	0.00672	0.000703	1.10×10^{-21}
22	43506635	rs4988372	*BIK*	C	T	0.874	0.00640	0.00107	2.10×10^{-09}

Supplementary Table 5. Independent genome-wide significant signals (P<5x10^{-8}) for colorectal cancer in UK Biobank GWAS

CHR	BP	SNP	Gene	EA	NEA	EAF	BETA	SE	P
1	222218761	rs12135286	*HHIPL2*	C	T	0.806	-0.00236	0.00035	1.40x10^{-11}
3	28450527	rs114717436	*ZCWPW2*	A	G	0.989	-0.00766	0.00138	2.70x10^{-08}
5	1296486	rs2735940	*TERT*	A	G	0.510	-0.00158	0.00028	1.40x10^{-08}
5	40280202	rs1445011	*LINC00603,PTGER4*	T	C	0.717	-0.00174	0.00031	1.70x10^{-08}
6	158842827	rs341145	*TULP4*	C	T	0.649	-0.00166	0.00029	9.10x10^{-09}
8	117630683	rs16892766	*LINC00536,EIF3H*	A	C	0.920	-0.00346	0.00051	1.10x10^{-11}
8	128413305	rs6983267	*CCAT2*	G	T	0.519	0.00248	0.00028	2.30x10^{-19}
11	111166504	rs12296076	*COLCA1*	G	A	0.327	0.00178	0.00030	2.10x10^{-09}
12	112553032	rs10850001	*NAA25,TRAFD1*	T	A	0.567	0.00153	0.00028	6.70x10^{-08}
15	33001734	rs58658771	*SCG5,GREM1*	T	A	0.820	-0.00322	0.00036	4.40x10^{-19}
18	46448805	rs6507874	*SMAD7*	T	C	0.527	0.00249	0.00028	4.70x10^{-19}
20	6405479	rs13037538	*CASC20*	A	G	0.641	-0.00173	0.00029	2.00x10^{-09}
20	47340117	rs6066825	*PREX1*	A	G	0.635	0.00177	0.00029	7.80x10^{-10}
20	60983973	rs7262524	*CABLES2*	C	T	0.735	0.00193	0.00031	8.40x10^{-10}

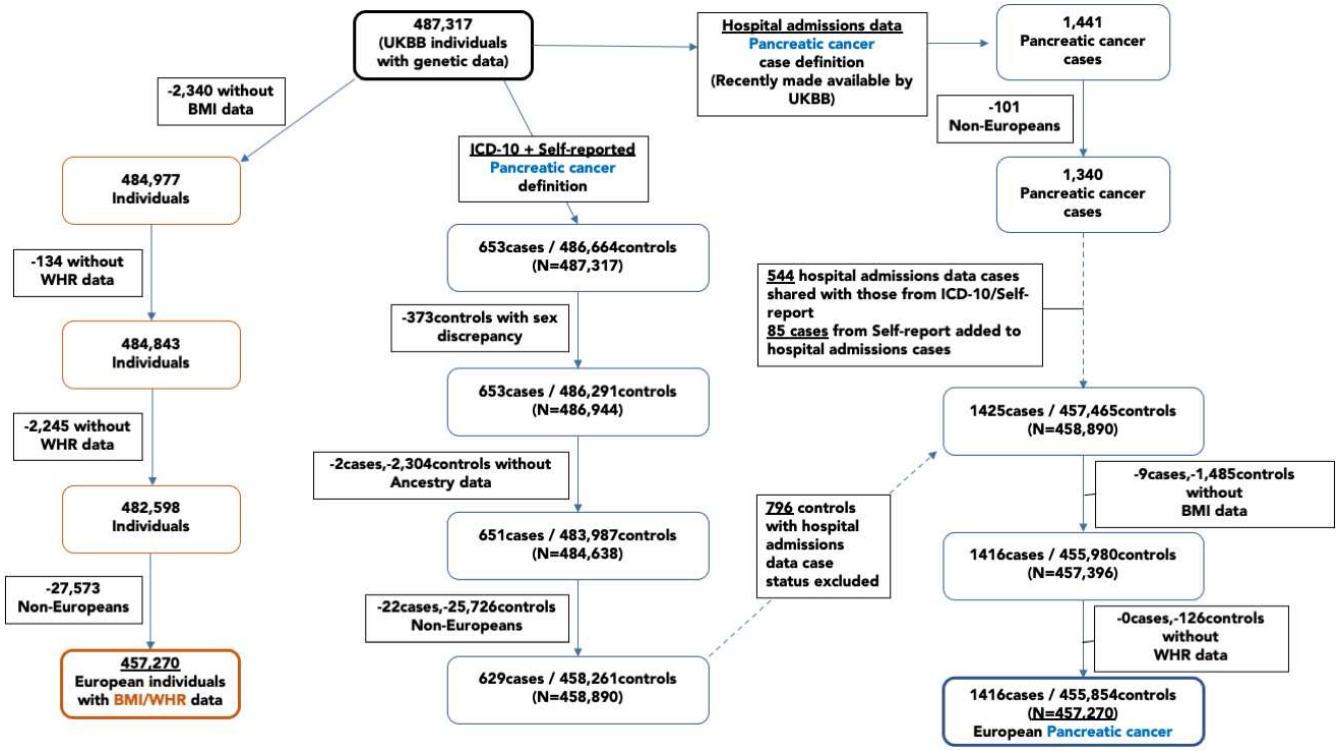
Supplementary Table 6. Independent genome-wide significant signals (P<5x10^{-8}) for pancreatic cancer in UK Biobank GWAS

CHR	BP	SNP	GENE	EA	NEA	EAF	BETA	SE	P
5	1300401	rs2736103	*TERT*	T	C	0.581	0.000702	0.00012	3.30x10^{-09}
9	136153875	rs651007	*ABO*	C	T	0.793	-0.000829	0.00014	7.60x10^{-09}
13	73916628	rs9543325	*KLF5*	C	T	0.362	0.000712	0.00012	3.90x10^{-09}
16	75234872	rs72802342	*ZFP1,CTRB2*	C	A	0.923	-0.001621	0.00022	2.50x10^{-13}

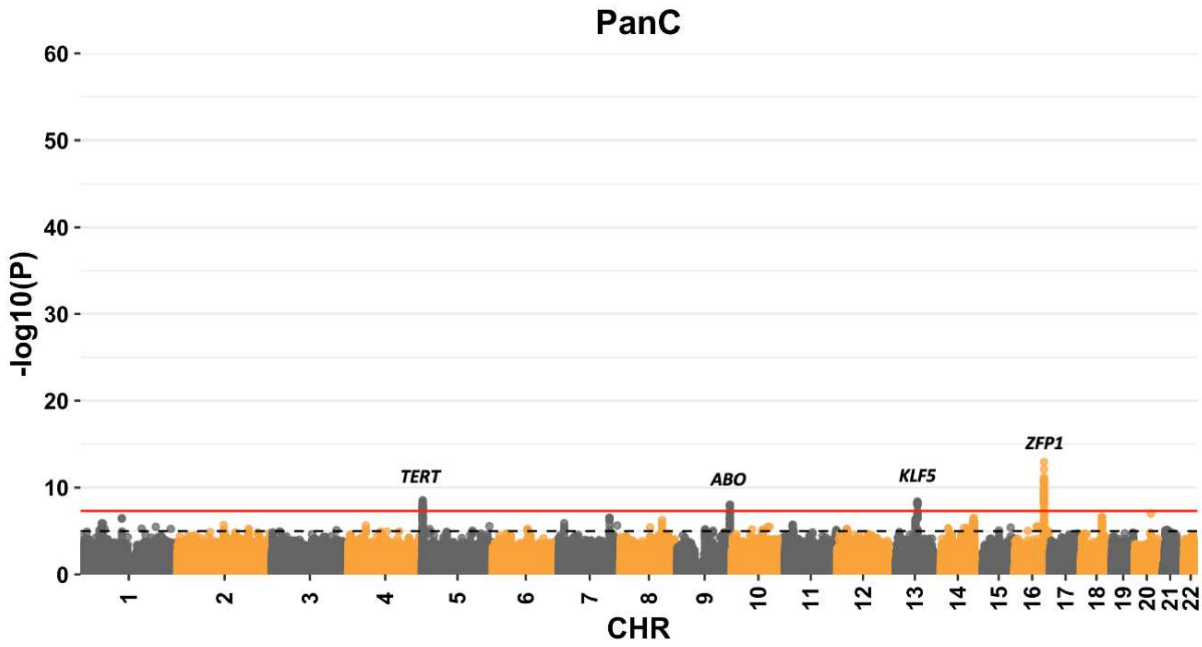
Supplementary Table 7. Independent genome-wide significant signals (P<5x10^{-8}) for lung cancer in UK Biobank GWAS

CHR	BP	SNP	GENE	EA	NEA	EAF	BETA	SE	P
5	1306165	rs4404721	*TERT*	T	C	0.369	-0.0013	0.0002	8.00x10^{-11}
15	78801394	rs11852372	*HYKK*	A	C	0.666	-0.0019	0.0002	3.10x10^{-20}
19	41342842	rs145580088	*CYP2T1P,CYP2A6*	A	G	0.978	0.0037	0.0007	3.90x10^{-08}

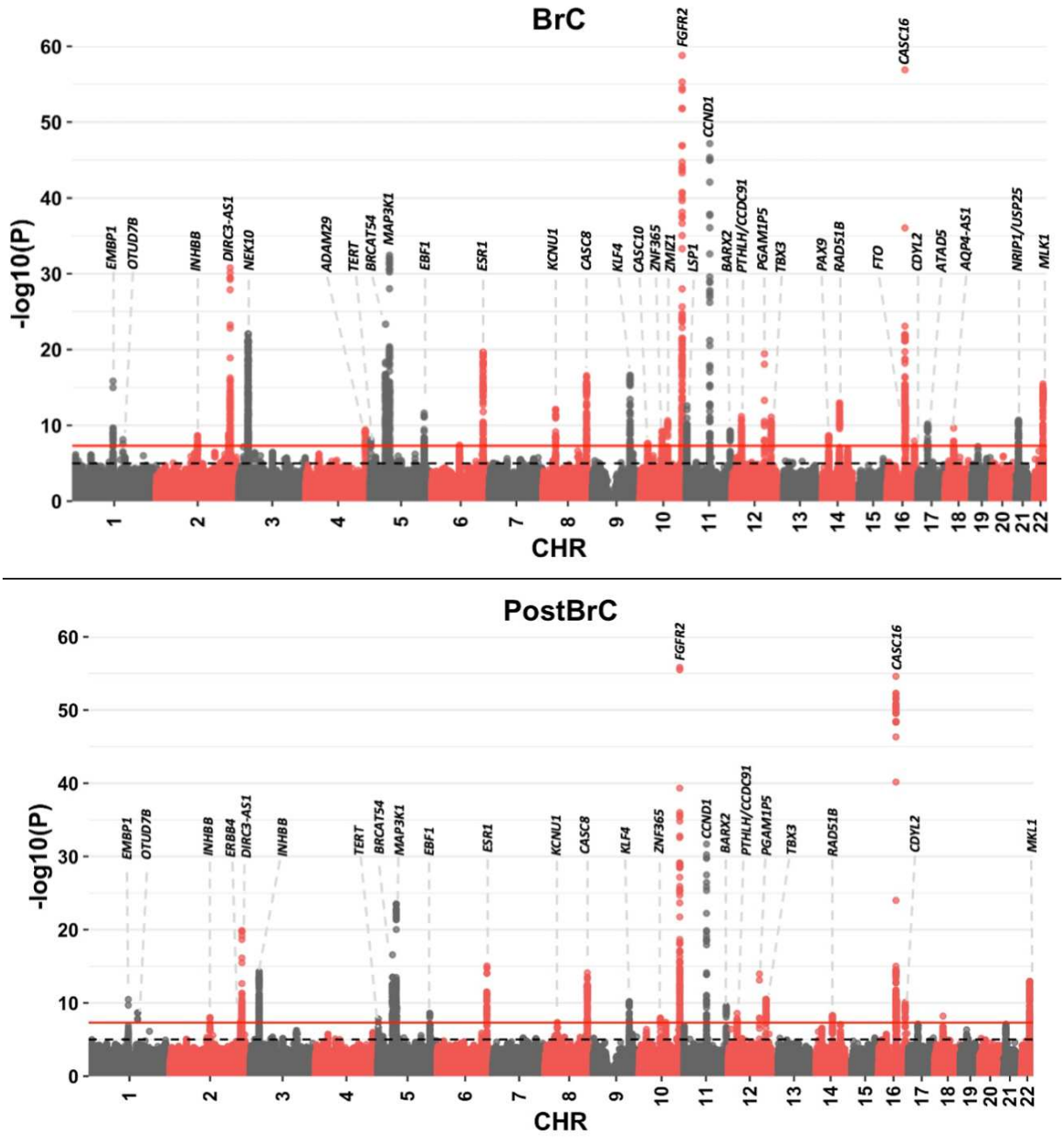
Supplementary Figure 1. Flowchart showing UK Biobank adiposity and pancreatic cancer definition



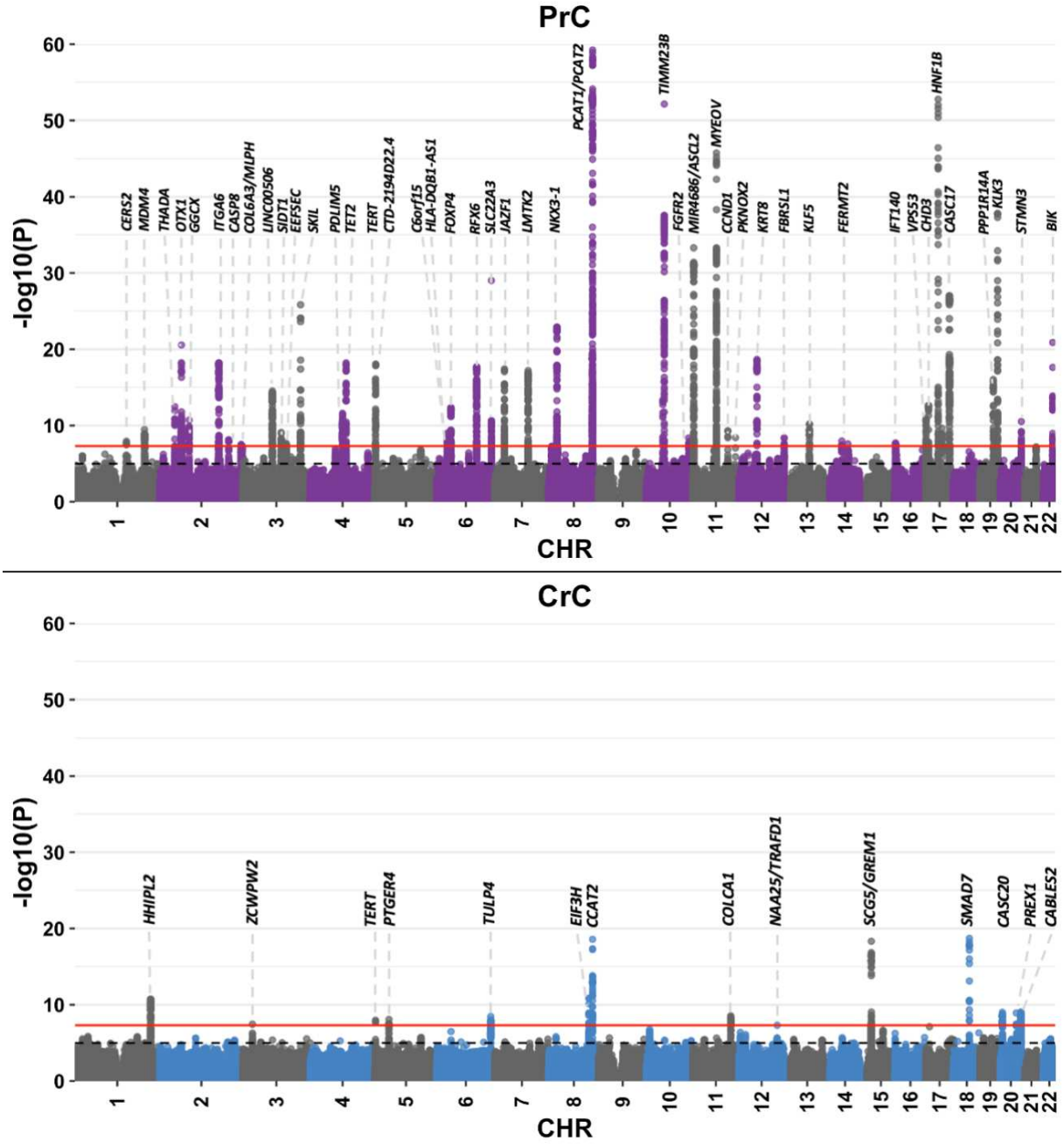
Supplementary Figure 2. Manhattan plot of lung cancer GWAS in UK Biobank. The red horizontal line shows genome-wide significance threshold (P<5x10^{-8}). The dashed grey line shows suggestive significance threshold (P<1x10^{-5})



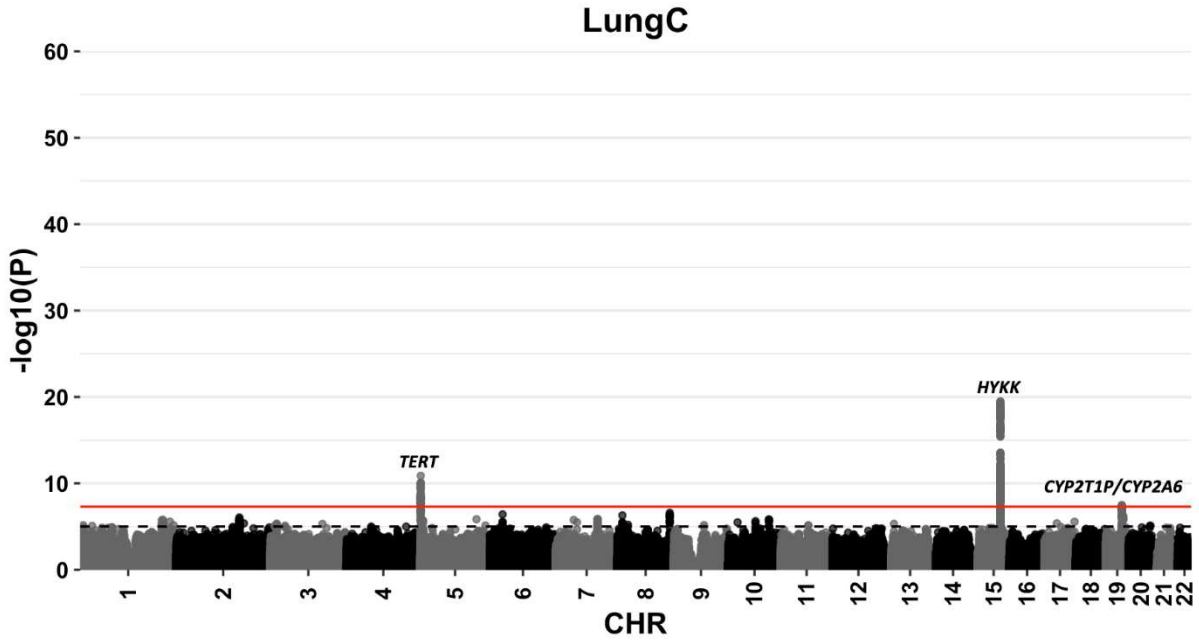
Supplementary Figure 3. Manhattan plots of overall breast cancer (top) and post-menopausal breast cancer (bottom) GWAS in UK Biobank. The red horizontal line shows genome-wide significance threshold (P<5x10^{-8}). The dashed grey line shows suggestive significance threshold (P<1x10^{-5})



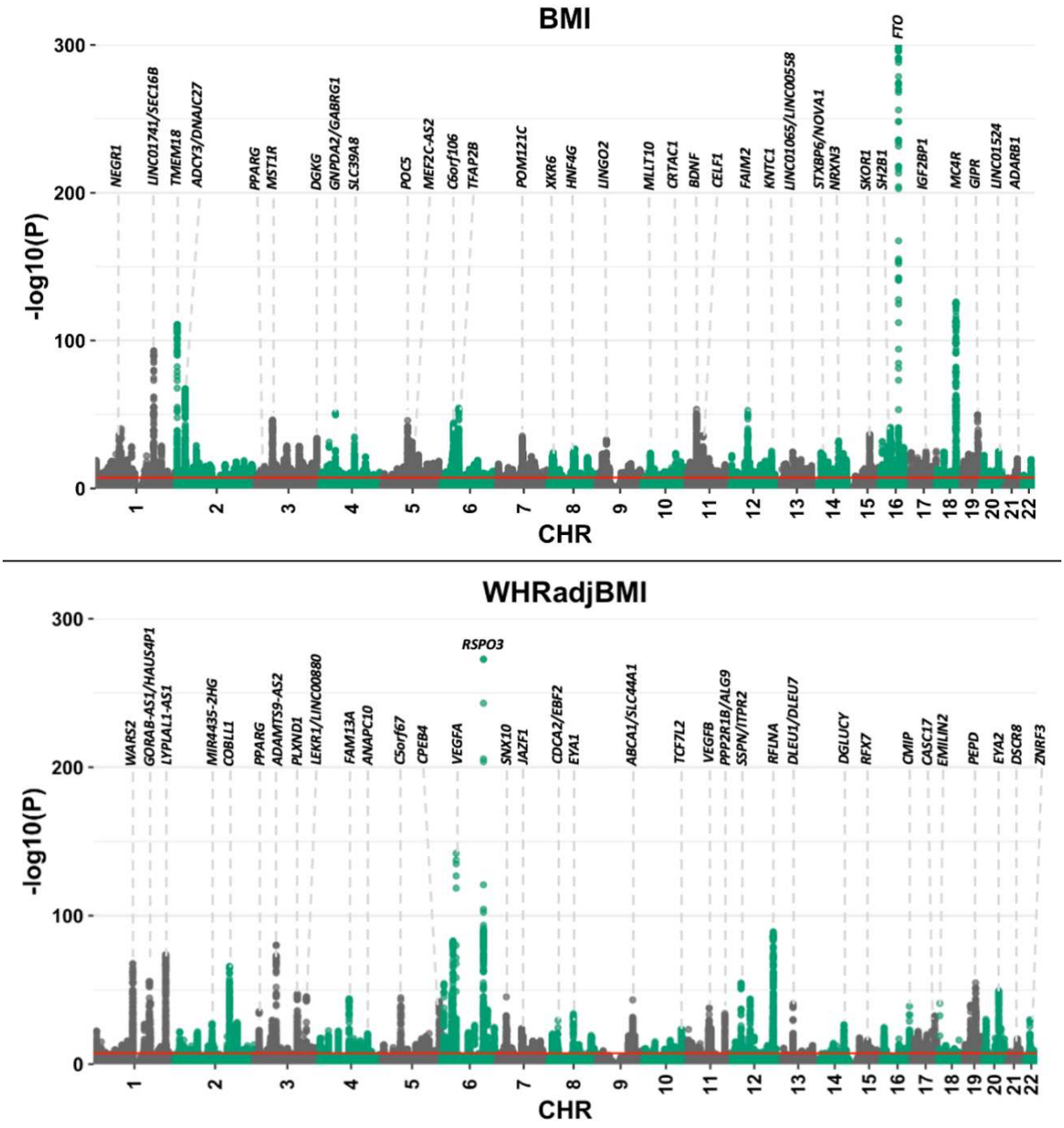
Supplementary Figure 4. Manhattan plots of prostate cancer (top) and colorectal cancer (bottom) GWAS in UK Biobank. The red horizontal line shows genome-wide significance threshold (P<5x10^{-8}). The dashed grey line shows suggestive significance threshold (P<1x10^{-5})



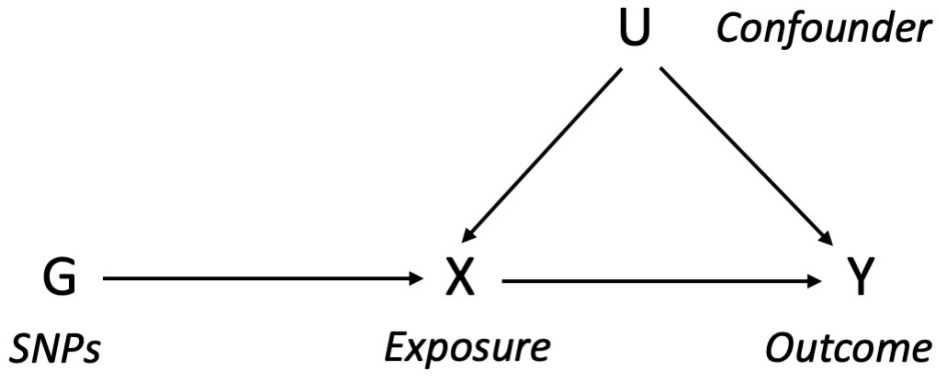
Supplementary Figure 5. Manhattan plot of lung cancer GWAS in UK Biobank. The red horizontal line shows genome-wide significance threshold (P<5x10^{-8}). The dashed grey line shows suggestive significance threshold (P<1x10^{-5})



Supplementary Figure 6. Manhattan plots of BMI (top) and WHRadjBMI cancer (bottom) GWAS in UK Biobank. The red horizontal line shows genome-wide significance threshold (P<5x10^{-8})

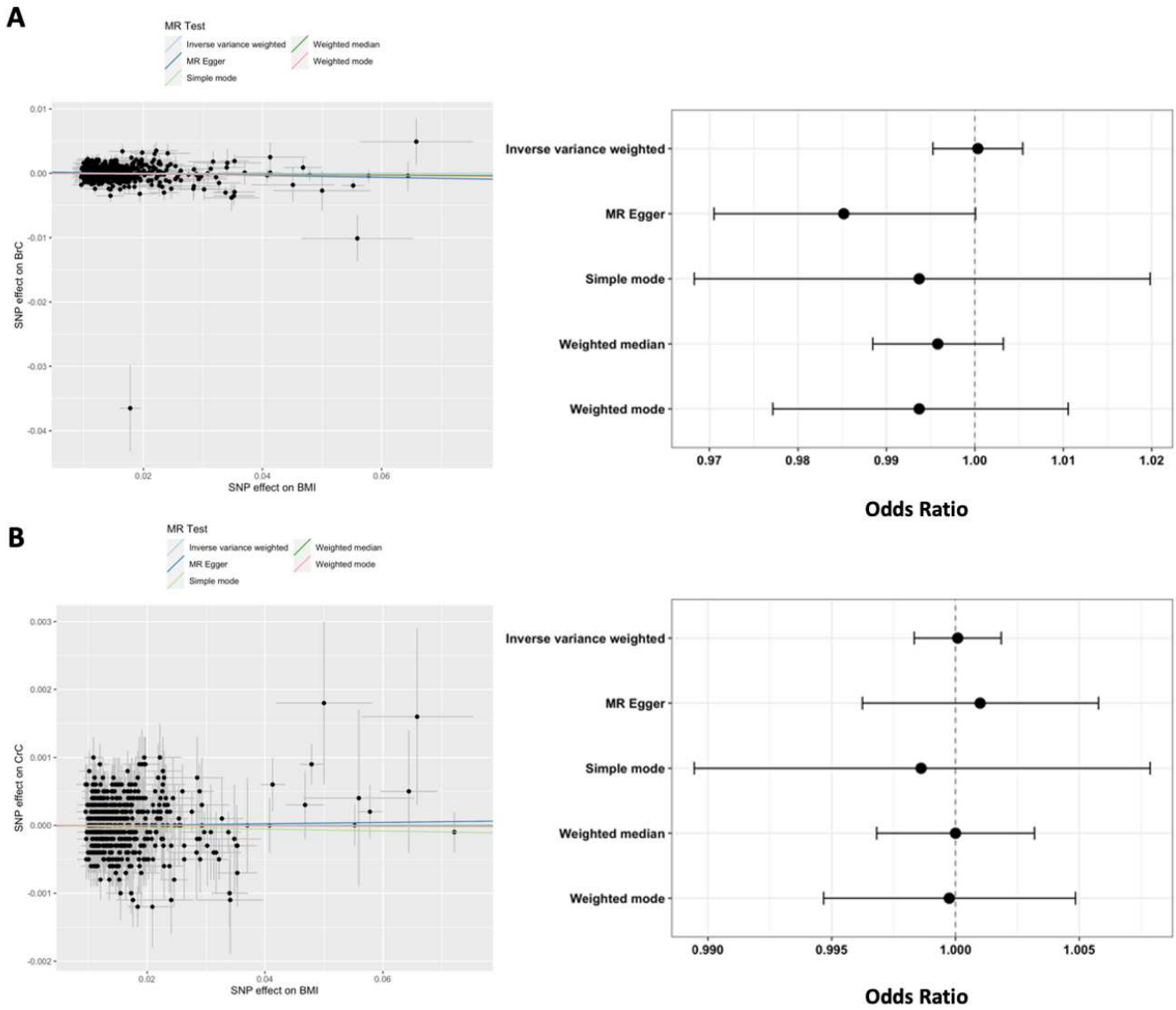


Supplementary Figure 7. Summary of Mendelian Randomization tests performed for BMI/WHRadjBMI and breast, prostate and colorectal cancers, including the number of genetic instruments (SNPs) available from published GWAS for each test

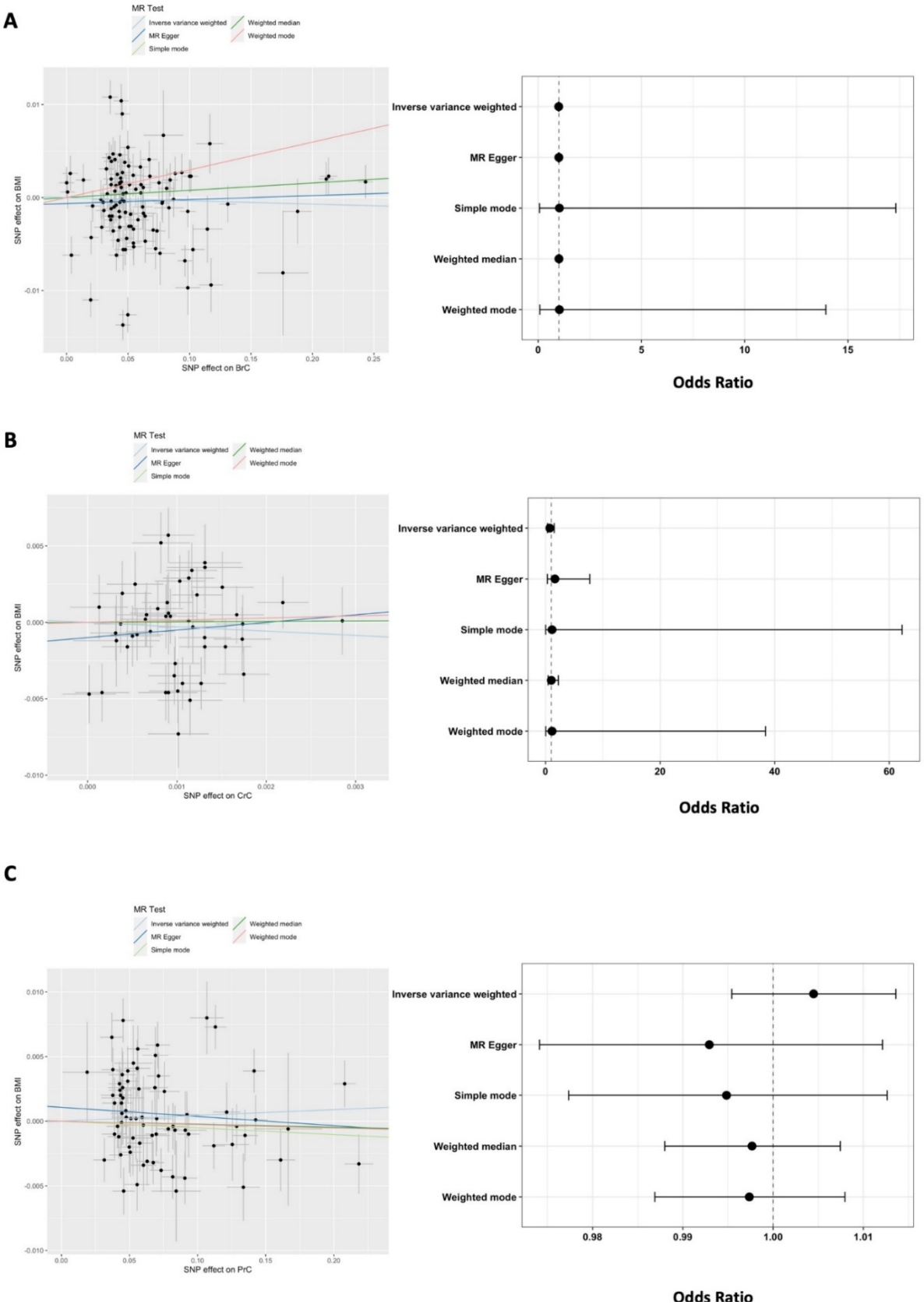


SNPs	*Exposure*	*Outcome*
668	BMI	BrC
668	BMI	CrC
668	BMI	PrC
342	WHRadjBMI	BrC
342	WHRadjBMI	CrC
342	WHRadjBMI	PrC
120	BrC	BMI
58	CrC	BMI
92	PrC	BMI
133	BrC	WHRadjBMI
66	CrC	WHRadjBMI
99	PrC	WHRadjBMI

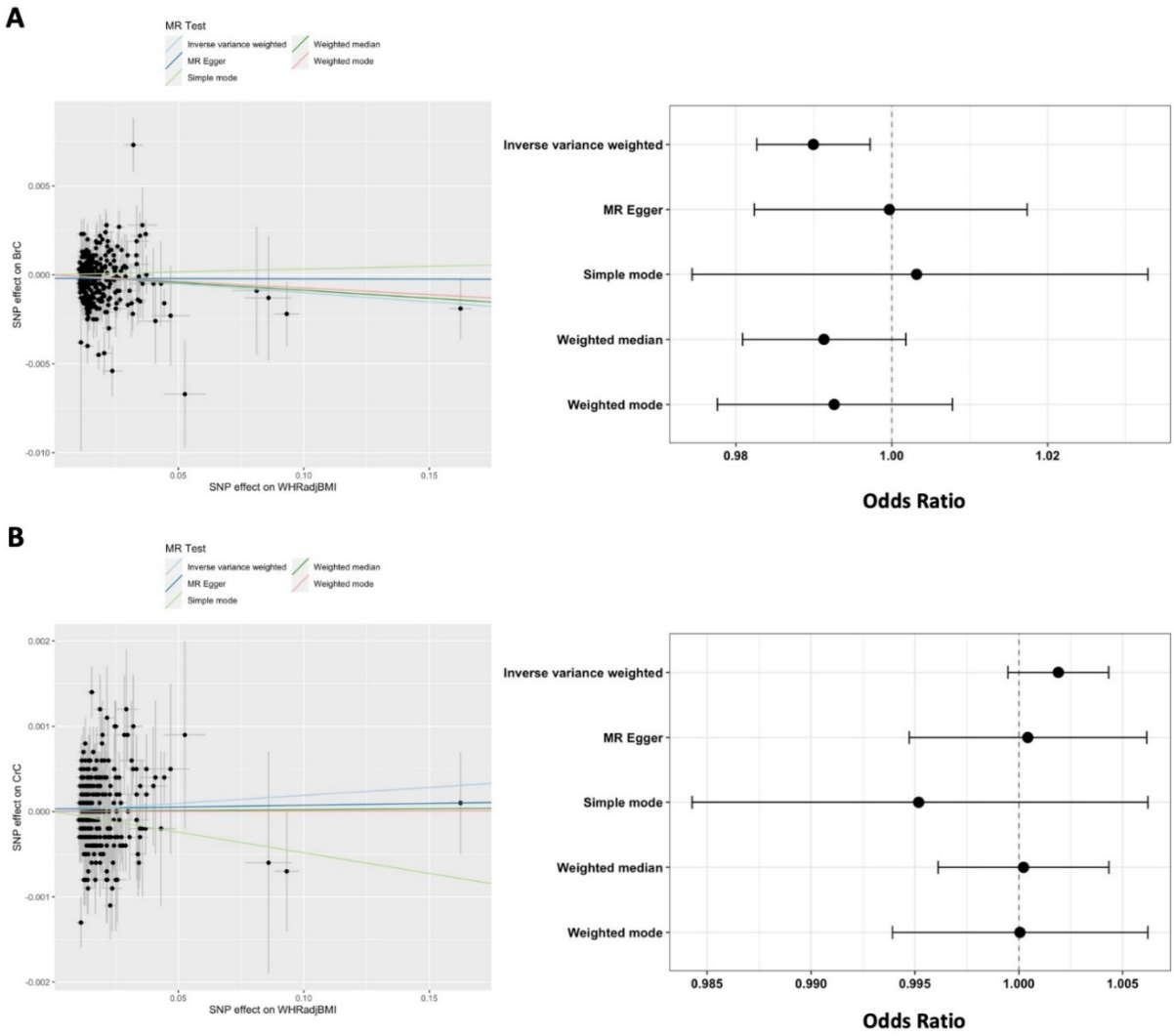
Supplementary Figure 8. Scatter and forest plots for the BMI to cancer direction Mendelian randomization tests. A) BMI to breast cancer. B) BMI to colorectal cancer



Supplementary Figure 9. Scatter and forest plots for the cancer to BMI direction Mendelian randomization tests. A) Breast cancer to BMI. B) Colorectal cancer to BMI. C) Prostate cancer to BMI



Supplementary Figure 10. Scatter and forest plots for the WHRadjBMI to cancer direction Mendelian randomization tests. A) WHRadjBMI to breast cancer. B) WHRadjBMI to colorectal cancer



Supplementary Figure 11. Scatter and forest plots for the cancer to WHRadjBMI direction Mendelian randomization tests. A) Breast cancer to WHRadjBMI. B) Colorectal cancer to WHRadjBMI. C) Prostate cancer to WHRadjBMI