

國立臺灣大學公共衛生學院流行病學與

預防醫學研究所

碩士論文

Institute of Epidemiology and Preventive Medicine

College of Public Health

National Taiwan University

Master Thesis

使用孟德爾隨機化試驗針對歐洲英國族群與漢族，探

索去脂肪體重與全身脂肪重量對乳癌的因果關係

Investigate causality of body fat mass and free fat mass

between breast cancer risk with Mendelian

randomization of European ancestry and Han Chinese

蔡佳宏

Chia-Hung Tsai

指導教授：盧子彬 博士

Advisor: Tzu-Pin Lu, Ph.D.

中華民國 111 年 6 月

June 2022



國立臺灣大學碩士學位論文
口試委員會審定書



使用孟德爾隨機化試驗針對歐洲英國族群與漢族，探索去
脂肪體重與全身脂肪重量對乳癌的因果關係

Investigate causality of body fat mass and free fat mass between
breast cancer risk with Mendelian randomization of European
ancestry and Han Chinese

本論文係 蔡佳宏 君（學號 R09849036）在國立臺灣大
學流行病學與預防醫學研究所完成之碩士學位論文，於民
國 111 年 6 月 13 日承下列考試委員審查通過及口試及格，
特此證明。

口試委員：

（簽名）

（指導教授）

馮燕臻

游宗憲

系主任、所長

（簽名）

（是否須簽章依各院系所規定）

致謝



時間就像飛鳥，一轉眼兩年就過了，台大流預所的時光也來到了尾聲。首先最想要感謝盧子彬老師，從剛進研究所開始，老師無論在課堂上或是研究室的會議，都非常有耐心地解答我許多問題，並且在論文研究遇到瓶頸時，與老師討論總會有豁然開朗的感覺，令我學到不同的角度看待自身的研究。

也非常感謝老師在論文指導上，不吝於指導我並給了我許多回饋，也願意給我許多的機會去嘗試與試錯，通過不斷地修正才有了如今的這篇論文。此外也必須謝謝定壕學長在這段期間給我的大力幫助，正是因為有了他基於學術與乳癌方面的許多建議，才避免我在研究上走更多的彎路。

再來謝謝 Amrita 與冠臻學長的幫忙，正是因為有他們的協助，我才對基因體中心的工作站以及孟德爾隨機化試驗的實際原理，有了更深刻的了解。在研究遇到問題時，冠臻學長也很熱情的願意來與我討論可能的解決方案。

也謝謝同研究室的學長姊、同學及學弟妹，在 meeting 中提供意見跟想法，讓我們在這段時間裡彼此學習、相互成長。從碩一剛進來時，我還是個對整個研究領域都一知半解的新手，再經過這兩年的訓練後，逐漸蛻變為一個合格的研究生。十分榮幸能成為台大流預所的畢業生，所上老師課程都非常有趣，也令人獲益良多。也很開心在流預所能夠交到志同道合的一群好朋友。

最後，也非常開心有家人們的支持與鼓勵，正是因為有了他們，讓我一步一步的完成了這兩年的求學之路，這一路走來，也非常慶幸遇到了許多貴人的幫助與鼓勵，謝謝你們。

蔡佳宏 謹致



中文摘要：

背景：

乳癌為女性發生率首位的癌症，其中之一的主要風險因子為肥胖，肥胖的測量指標眾多，然而在孟德爾隨機化試驗（Mendelian randomization, MR）的文獻回顧中，不同的肥胖指標對乳癌風險皆為保護因子或是不顯著，與臨床上肥胖作為乳癌的風險因子情況並不吻合。

為了探討肥胖指標對乳癌為保護因果關係的原因，我們將體重指標進行拆解，分別針對脂肪重量與去脂肪重量進行因果推論，透過拆解指標的方式來了解究竟是脂肪重量還是肌肉重量對乳癌風險有保護作用。我們希望確認肥胖與乳癌發生之間的因果關係，了解兩者之相關性不僅有助於提早找出罹患乳癌的高風險個案。在臨床意義上，更可以透過介入手段增加個體的去脂肪體重，進而作為預防措施。在本研究中，我們將使用孟德爾隨機化試驗，釐清兩者之間的因果關係。

方法：

本研究使用英國生物資料庫的女性乳癌患者資料，進行單樣本孟德爾隨機化試驗，分別篩選出 164 與 98 個變異位點作為去脂肪體重（Free fat mass, FFM）與全身脂肪體重（Body fat mass, BFM）的工具變數（Instrumental variables, IV），使用二階段迴歸、逆方差加權之估計方法衡量兩者間的因果關係，並額外利用加權中位數方法、MR-Egger 迴歸進行穩健的因果推斷。在敏感性研究中，藉由將乳癌設定為曝露變項，去脂肪體重作為結果變項，探索兩個變數間是否存在雙向的因果關係。

在外部驗證中，我們使用從英國生物資料庫中所篩選的 164 個變異位點，針對去脂肪體重進行雙樣本孟德爾隨機化試驗：首先使用乳腺癌協會聯盟（Breast Cancer Association Consortium, BCAC）的匯總性全基因組關聯分析

（Genome-wide association study, GWAS）資料進行驗證；接下來，使用臺灣生物資料庫的資料進行雙樣本孟德爾隨機化試驗，試圖在漢族人群中重現同樣的去脂肪體重與乳癌的因果關係。

結果：

其結果顯示（ $\hat{\beta}_{2SLS} = -0.012$, $OR_{2SLS} = 0.988$, $P_{2SLS} = 0.283$ ； $\hat{\beta}_{IVW} = -0.024$, $OR_{IVW} = 0.976$, $P_{IVW} = 0.012$ ），去脂肪體重與乳癌存在負向因果關係，然而在不同統計方法上的統計效力並不穩定。在敏感性研究中（ $\hat{\beta}_{2SLS} = 0.352$,

$OR_{2SLS}=1.422$, $P_{2SLS}=0.296$; $\hat{\beta}_{IVW}=0.062$, $OR_{IVW}=1.064$, $P_{IVW}=0.175$) 驗證了僅有去脂肪體重增加會降低乳癌風險的因果關係，反向因果關係不存在，罹患乳癌不會影響去脂肪體重的增加或減少。

根據乳腺癌協會聯盟的外部驗證結果 ($\hat{\beta}_{IVW}=0.002$, $OR_{IVW}=1.002$, $P_{IVW}=0.733$)，與臺灣生物資料庫的外部驗證結果 ($\hat{\beta}_{IVW}=-0.056$, $OR_{IVW}=0.946$, $P_{IVW}=0.448$)，去脂肪體重的雙樣本孟德爾隨機化試驗，並不顯著，我們認為可能由於不同資料庫間的人種差異與篩選差異所導致。

結論：

本研究使用英國生物資料庫進行的單樣本孟德爾隨機化試驗，確認了去脂肪體重為乳癌風險的保護因子，兩者存在負向的因果關係，並且不存在反向因果關係。而全身脂肪體重與乳癌風險間，則不存在任何統計顯著的因果關係。

我們針對肥胖對於乳癌，在孟德爾隨機化試驗中不尋常的保護作用，提出了可能的解釋，利用將體重拆分成去脂肪體重與全身脂肪體重，分別進行單樣本孟德爾隨機化試驗與雙樣本隨機化試驗，我們可以更清楚的了解到，究竟是脂肪重量的增加，亦或是肌肉重量的增加，導致了乳癌風險的下降。並得出了是肌肉重量（去脂肪體重）的增加降低了乳癌風險的可能。

雖然該結果於台灣族群中不具備外推性，但可能是由於乳癌在不同種族間的機轉差異導致，在未來研究中，可能需要使用漢族的去脂肪體重資料篩選出工具變數，進一步考慮乳癌族群差異造成的影響。

在本研究中，去脂肪體重作為能夠將低乳癌風險發生的保護因子，與停經後的體內激素水平息息相關，台灣女性的平均更年期為 48-52 歲，與台灣乳癌發生率高峰 45-69 歲有著很高的重合時段。站在疾病預防的角度上，應當注意女性在更年期後運動習慣是否發生改變，以及其肌肉流失的狀態，因為兩者皆有可能造成乳癌風險的增加。建議鼓勵停經期後女性進行適當的運動作為介入手段，維持體內激素水平，以降低罹患乳癌的風險。

關鍵字：乳癌，去脂肪體重，全身脂肪體重，孟德爾隨機化試驗，因果關係，英國生物資料庫，台灣生物資料庫



Abstract:

Background:

Breast cancer is the most common cancer in women, and one of the main risk factors is obesity. There are many measurements of obesity. However, in the literature review of Mendelian randomization (MR), obesity plays a different role in breast cancer. Mendelian randomization's results are either protective or insignificant, which are not consistent with the clinical result that obesity is a risk factor for breast cancer.

To investigate the causal relationship between obesity measurements and breast cancer protection, we divide body weight into body fat weight (BFM) and free fat weight (FFM), and made causal inferences for BFM and FFM respectively. Understanding the correlation between obesity indicators and breast cancer will not only help to identify high-risk cases of breast cancer early, but also can increase the individual's fat-free body weight through interventional methods, which can be used as a preventive measure. In this study, we will use Mendelian randomization to clarify the causal relationship between obesity indicators and breast cancer.

Method:

This study uses the data of female breast cancer patients from the UK Biobank to conduct one-sample Mendelian randomization, extracting 164 and 98 variants as FFM's and BFM's instrumental variables (IV). Using two-stage least squares and inverse variance weighted estimation measures the causal relationship between obesity indicators and breast cancer. We also use the weighted median method, MR-Egger regression for the robust causal inference. In a sensitivity study, by setting breast cancer as the exposure variable and FFM as the outcome variable, we investigate whether there is a bidirectional causal relationship between the two variables.

In external validation, we perform two-sample Mendelian randomization by using FFM's 164 variants extracted from the UK Biobank. We first collect Genome-wide association study (GWAS) data from the Breast Cancer Association Consortium (BCAC). Next, two-sample Mendelian randomization is performed with data from the Taiwan Biobank to reproduce the same causal relationship between FFM and breast cancer in Han Chinese.

Result:

The results show ($\hat{\beta}_{2SLS} = -0.012$, $OR_{2SLS} = 0.988$, $P_{2SLS} = 0.283$; $\hat{\beta}_{IVW} = -0.024$, $OR_{IVW} = 0.976$, $P_{IVW} = 0.012$), a negative causal relationship between FFM and breast cancer, but the statistical power is not stable in different statistical methods. The sensitivity study ($\hat{\beta}_{2SLS} = 0.352$, $OR_{2SLS} = 1.422$, $P_{2SLS} = 0.296$; $\hat{\beta}_{IVW} = 0.062$, $OR_{IVW} = 1.064$, $P_{IVW} = 0.175$) tests a causal relationship that only increased FFM is associated with a reduced risk of breast cancer, the reverse causality did not exist and having breast cancer did not affect the gain or loss of FFM.

Conclusion:

In this study, one-sample Mendelian randomization conducted by the UK Biobank confirms that FFM is a protective factor for breast cancer risk with the result that there exists a negative causality and no reverse causality. There is no statistically significant causal relationship between BFM and breast cancer risk.

We propose possible explanations for the unusual protective effect of obesity on breast cancer in previous Mendelian randomization papers by splitting off body weight into FFM and BFM and using one-sample Mendelian randomization and two-sample randomization. Therefore, we can understand whether an increase in fat mass or an increase in muscle mass leads to a decrease in breast cancer risk more clearly. And we come to the conclusion that the increase in muscle mass (FFM) may reduce the risk of breast cancer.

This result may not be significant in the Taiwanese ethnic group because of the difference in the mechanism of breast cancer among different ethnic groups. In future studies, it would be necessary to use the FFM data of the Han Chinese to extract instrumental variables and further consider breast cancer differences between ethnic.

In this study, FFM, as a protective factor that can reduce the risk of breast cancer, is closely related to hormone levels after menopause. The average menopause age of Taiwanese women is 48-52 years old, which is almost matching to the peak incidence of breast cancer in Taiwan at 45-69 years old. From the perspective of disease prevention, it should be noted that postmenopausal women may increase the risk of breast cancer if their exercise habits change after menopause or their muscle loss. It is recommended for postmenopausal women to develop exercise habits as an intervention to maintain hormone levels in the body to reduce the risk of breast cancer.

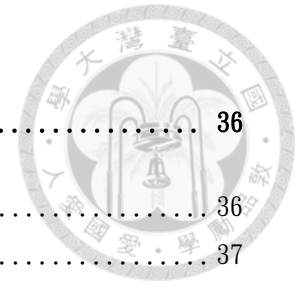
Keywords: breast cancer, body fat mass, free fat mass, Mendelian randomization, causality, UK Biobank, Taiwan Biobank





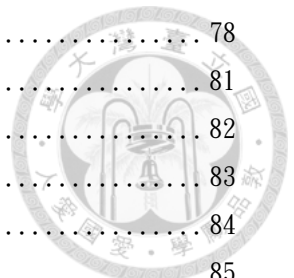
目錄

國立臺灣大學碩士學位論文口試委員會審定書	1
致謝	2
中文摘要:	3
ABSTRACT:	5
第一章 導論	11
1.1 研究背景.....	11
1.2 研究動機與目的.....	12
第二章 研究材料與方法	13
2.1 資料來源.....	13
2.2 樣本篩選-英國生物資料庫.....	15
2.3 本研究方法論述.....	16
2.4 孟德爾隨機化試驗方法論述.....	18
2.5 敏感性分析.....	19
2.6 工具變數的相關分析.....	20
2.7 外部驗證.....	21
第三章 結果	23
3.1 資料展示與變數評估.....	23
3.2 孟德爾隨機化試驗-主要結果.....	24
3.3 敏感性分析.....	24
3.4 工具變數的相關分析.....	26
3.5 外部驗證.....	27
第四章 結論與討論	27
4.1 主要發現.....	27
4.2 研究討論.....	28
4.3 結論.....	30
4.4 研究限制.....	30
參考文獻	32



圖表資料:	36
圖一：研究架構	36
圖二：樣本篩選流程	37
圖三：孟德爾隨機化試驗所使用之假設.....	38
表一：英國生物資料庫人口特徵-依照種族分群.....	39
表二：英國生物資料庫-人口學特徵資料展示.....	41
表三：英國生物資料庫-肥胖指標資料展示.....	43
圖四：英國生物資料庫-相關肥胖指標展示.....	45
表四：去脂肪體重-乳癌風險的主要孟德爾隨機化試驗結果.....	46
表五：身體脂肪體重-乳癌風險的主要孟德爾隨機化試驗結果.....	47
表六：敏感性分析-陰性對照：罹患乳癌是否會影響去脂肪體重的敏感性孟德爾隨機化試驗結果.....	48
表七：敏感性分析-孟德爾隨機化試驗 群集分析(是否停經)：停經後確診.....	49
表八：敏感性分析-孟德爾隨機化試驗 群集分析(組織亞型)：乳腺癌.....	51
表九：將去脂肪體重所使用 164 個變異位點，進行基因註釋。.....	52
圖五：單一變異位點 (RS12055445) 的全表型組關聯分析結果.....	57
表十：使用 164 個去脂肪體重工具變數變異位點，依次進行全表型組關聯分析，並彙整出顯著的變異位點與表型結果。.....	58
表十一：針對表十中的表型，統整出現次數.....	61
表十二：外部驗證-乳腺癌協會聯盟 2020：使用雙樣本孟德爾隨機化試驗，驗證在歐洲人種中，去脂肪體重與乳癌風險因果關係的一致性.....	63
表十三：外部驗證-乳腺癌協會聯盟 2020：使用雙樣本孟德爾隨機化試驗，驗證在歐洲人種中，全身脂肪體重與乳癌風險因果關係的一致性.....	64
表十四：外部驗證-臺灣生物資料庫：使用雙樣本孟德爾隨機化試驗，驗證在歐洲人與漢人間，去脂肪體重與乳癌風險的因果關係，是否存在一致性.....	65
表十五：外部驗證-臺灣生物資料庫：使用雙樣本孟德爾隨機化試驗，驗證在歐洲人與漢人間，全身脂肪體重與乳癌風險的因果關係，是否存在一致性.....	66
附錄補充-圖表資料:	67
附錄圖一：主成分分析 三維散佈圖 依照種族.....	67
附錄圖二：主成分分析 二維散佈圖 (A 圖為依照各種族的散佈圖，B 圖為針對英國白人族群的進一步篩選，僅選擇主成分分析相近的作為研究對象，如圖 B 的藍色部分所示).....	68
附錄表一：英國生物資料庫-變項與對照編碼.....	69
附錄表二：英國生物資料庫，肥胖指標重複測量之時間段與數據可用性彙整.....	71
附錄表三：台灣生物資料庫-資料展示.....	72
附錄表四：去脂肪體重之工具變數所使用之變異位點-英國生物資料庫.....	74

附錄表五：全身脂肪體重之工具變數所使用之變異位點-英國生物資料庫.....	78
附錄圖三：全基因組關聯分析曼哈頓圖-去脂肪體重.....	81
附錄圖四：全基因組關聯分析 QQ-PLOT-去脂肪體重.....	82
附錄圖五：全基因組關聯分析曼哈頓圖 全身脂肪體重.....	83
附錄圖六：全基因組關聯分析 QQ-PLOT 全身脂肪體重.....	84
附錄圖七：去脂肪體重單一突變位點的 WALD 比率估計	85
附錄圖八：全身脂肪體重單一突變位點的 WALD 比率估計	86
附錄表六：乳癌風險之工具變數所使用之變異位點-英國生物資料庫.....	87
附錄表七：體脂率-乳癌風險的單樣本孟德爾隨機化試驗結果.....	89
附錄表八：去脂肪體重指數-乳癌風險的單樣本孟德爾隨機化試驗結果.....	90
附錄表九：全身脂肪體重指數-乳癌風險的單樣本孟德爾隨機化試驗結果.....	91
中英對照與縮寫表:.....	92



第一章 導論



1.1 研究背景

全球肥胖是對公共健康的主要威脅。作為癌症的第二大常見原因，成人和兒童的超重和肥胖率仍舊在繼續增長。根據世界衛生組織（World Health Organization, WHO）的統計，從 1975 年到 2016 年，全球 5-19 歲的超重或肥胖兒童和青少年的患病率從 4% 增加到 18%，增加了 4 倍多¹。肥胖已經成為民眾亞健康的一種常見型態，同時肥胖患者往往可能增加罹患癌症、糖尿病、心血管疾病等相關風險^{2 3 4}，也因此醫療體系中，肥胖對醫療體系造成了一定比例的負擔。

在全球，乳癌是最常被診斷出的癌症，也是癌症死亡的主要原因⁵。並且根據衛生福利部-國民健康署的統計，乳癌也為我國婦女發生率第一位之癌症，女性乳癌標準化發生率及死亡率分別為 69.1 及 12.0（每十萬人口）⁶。取決於乳癌的盛行率高，且死亡率相對較其他癌症低的特點，診斷時間越早，則可以更有效的改善乳癌患者的預後，基於這樣的考量，我們希望找到一些臨床因子，能夠被驗證其與乳癌風險間的因果關係，並且研究在台灣漢族人群中是否存在一致的因果關係。

在世界癌症研究基金會（World Cancer Research Fund, WCRF）和美國癌症研究所（American Institute for Cancer Research, AICR）的報告中，肥胖與乳腺、子宮內膜、卵巢和腎臟等多種癌症皆有關聯^{7 8}，並且在針對肥胖與癌症的薈萃分析也發現了肥胖作為眾多癌症的風險因子⁹。但在結果的判讀上，肥胖容易受到潛在的混淆因子的影響¹⁰，因為肥胖與癌症的各種危險因素同時發生，例如糖尿病等，並且容易出現反向因果關係¹⁰。

因此除了上述人體臨床上的相關研究，在肥胖與乳癌的致病機轉上，也有許多文獻透過動物試驗的方式，試圖分析肥胖與乳癌風險間的關係^{11 12 13}，以及肥胖與乳癌復發間的關係。並提供了一種基因工程小鼠模型¹⁴，用於研究飲食引起的肥胖與乳腺癌復發的關係。

雖然肥胖作為多數癌症中的風險因子。若將範圍限縮於乳癌，在肥胖與乳癌間的關係上，仍有待探討的地方。透過動物試驗，可以發現肥胖如何導致乳癌發生的生物機轉。在大部分的觀察性研究中，肥胖與停經後乳癌呈現正相關，與

停經前乳癌呈現負相關亦或是沒有相關。此外，針對研究種族的不同，有時這樣的相關性也會出現不一致，在肥胖與乳癌的荟萃分析中¹⁵，觀察到了乳癌與身體質量指數的交互作用在不同種族間方向上的差異。在歐洲與非洲女性中，身體質量指數的增加，降低停經前乳腺癌的風險；在亞洲女性中，身體質量指數的增加，卻增加了停經前乳腺癌的風險。

肥胖的評估指標眾多，在過往的研究中主要透過身體質量指數來對肥胖進行定義，而體脂肪等相關資料由於較難取得，則較少被考慮。在基於英國人口的一份研究中，揭露了總計六種肥胖相關指標與癌症間的關係¹⁶，並發現乳腺癌與其餘身體質量指數、體脂率、腰臀比皆呈現線性的正相關，若將乳癌進一步的細分為停經前與停經後，則肥胖指標僅與停經後乳腺癌發病率有關。

整體來說，在觀察性研究中，肥胖為乳癌的危險因子，但是在近幾年的肥胖與乳癌的孟德爾隨機化試驗 (Mendelian randomization, MR) 中，則很少出現一致性的結論，近幾年去脂肪體重與乳癌風險的相關研究，主要有兩篇^{17 18}。分別使用雙樣本孟德爾隨機化試驗研究脂肪分布與不同癌種間的風險因果關係，並得出一般肥胖會降低患乳腺癌的風險的結論¹⁷。以及使用英國生物資料庫的資料，進行泛癌因果分析，但就去脂肪體重與乳癌等指標，則並未得出統計顯著的因果關係¹⁸。

1.2 研究動機與目的

基於精準健康的角度，為了找出罹患乳癌的高風險個案，我們必須確立肥胖與乳癌間的因果關係，同時在上述兩項研究中^{17 18}，除了兩方的因果關係存在不一致外，其與傳統的流行病學研究結果也並不一致。

肥胖的測量指標眾多，體重、身體質量指數、腰臀比、體脂率、全身脂肪體重，皆可以作為肥胖的衡量指標，肥胖指標間也存在著強烈的相關性¹⁹。2021年的一篇基於英國生物資料庫的研究，使用存活模型分析身體質量指數、體脂百分比以及腰臀比對各類癌症的影響，並發現了身體質量指數、體脂百分比以及腰臀比為停經後乳腺癌的風險因子¹⁶。然而，在另一篇同樣基於英國生物資料庫的研究中，使用孟德爾隨機化試驗來研究體重指數、體脂百分比以及腰臀比，卻得出了與觀察性研究中截然相反的結果，體重指數、體脂百分比以及腰臀比對乳腺癌具有因果保護作用²⁰，並且在其他歐洲人口的肥胖與乳癌相關的孟德爾隨機化試驗文獻中，肥胖指標對乳癌風險的影響皆為保護作用^{21 22 23}。

體重指數、體脂百分比以及腰臀比的肥胖指標，在臨床上與孟德爾隨機試驗的因果推論上，分別為風險因子與保護因子，並且在各自的研究中，皆達到統計

險中。基於這樣截然不同的結論^{16 20}，作者提出的可能原因為，脂肪組織可能在安全儲存持久性有機污染物方面發揮作用，同時也提及身體活動相關的變量（即爬樓梯的頻率和適度的身體活動）可能是乳癌風險的干擾因子來源之一²⁰。

然而在流行病學研究中，肥胖作為乳癌的風險因子，增加了人體的內源性雌激素分泌。在疾病機轉上的解釋為：女性停經後，原先分泌雌二醇的卵巢停止分泌的情況下，體內的雌激素水平將開始被脂肪影響，將會增加罹患乳癌風險²⁴。

因此本篇研究中，我們希望將體重透過生物電阻抗分析，拆分為去脂肪體重與全身脂肪體重，希望透過研究這兩項指標，來解釋肥胖對乳癌的保護作用。究竟是由於體內脂肪能夠儲存親脂性持久性有機污染物（persistent organic pollutants, POP）的原因²⁵，亦或是運動習慣造成的體重增加，作為降低乳癌風險的途徑²⁶。

同時，在肥胖相關的孟德爾隨機化試驗文獻中，針對東亞人群的研究特別稀少，目前僅有一篇探討體重指數與乳腺癌生存的研究²⁷，因此我們也希望透過本研究，使用台灣生物資料庫的資料，研究肥胖指標與乳癌風險間的關係。

第二章 研究材料與方法

2.1 資料來源

英國生物資料庫簡介：

本研究使用英國生物資料庫作為資料來源，英國生物資料庫收錄了自 2006 年到 2010 所招募的自願者，仍在不斷地更新資料，截至本論文的書寫時間為止，目前的最後追蹤日期為 2022-02-10，英國生物資料庫總計招募了約 502,505 的成年人，以及每個人擁有約 456,028 個變異位點遺傳資料，參加研究時參與者的年齡範圍在 40 至 69 歲之間，男性和女性的平均年齡分別為 56.63 歲和 57.10 歲。

乳腺癌協會聯盟以及台灣生物資料庫簡介：

在外部驗證中，本研究使用乳腺癌協會聯盟（Breast Cancer Association Consortium, BCAC）以及台灣生物資料庫的資料。由於乳腺癌協會聯盟並未提

供個人數據，僅使用全基因組關聯分析匯總性資料來進行雙樣本孟德爾隨機化試驗，總計包含了 133,384 筆病例以及 113,789 筆對照²⁸。

本研究使用台灣生物資料庫 2008 年至 2015 年所招募之自願者，總計 16000 人，收錄了 7966 位男性與 8034 位女性，篩選女性作為研究對象，並根據志願者與研究人員的面對面訪談獲得的資料，篩選出乳癌女性與非乳癌女性，維持病例與對照 1:4 的比例進行抽樣，最終使用 407 筆乳癌女性，以及 1224 筆非乳癌女性，平均年齡為五十歲，詳細的人口分佈，在（附錄表三）中作為參照。

英國生物資料庫-使用表型：

通過與英國國家癌症登記處資料所彙整的病例訊息。我們根據 ICD10 定義癌症患者與非癌症患者，將至少出現過一次 C50 的患者定義為乳癌患者；將每次的更新資料中，皆無出現 C 字首的患者，定義為非癌症患者。總計有 259,383 位非癌症患者，與 14,105 位的乳癌患者。

我們運用英國生物資料庫之編碼從中提取出以下變數進行分析，詳細編碼請參考（附錄表一），包含去脂肪體重、體脂率、身高、飲酒頻率、抽菸頻率、糖尿病、年紀、生育數量、初經年齡。針對年紀，我們對癌症患者與一般患者有各自的定義。由於本研究關注的是癌症的發生率，故癌症患者使用確診時的確診年齡，一般患者則是使用招募年齡。

英國生物資料庫採用 Tanita BC418MA-身體成分分析儀測量去脂肪體重與體脂率，俗稱生物電阻分析。利用電流通過身體，並因為脂肪與其他組織的含水量不同，導致電阻間的差異，透過這樣的差異變化計算出受試者的去脂肪體重。由於體重、體脂率、去脂肪體重等肥胖指標，會隨著時間不斷地改變，在衡量體重與癌症風險時，以什麼時間點的體重作為研究依據會極大的影響結果。但是由於英國生物資料庫的資料限制，我們僅能得到四個時間段下受試者的相關數據。

分別為初步評估訪問-1st：（ 2006-2010 ）；第一次重複評估訪問-2nd：（ 2012-2013 ）；影像訪問-3rd：（ 2014+ ）；第一次重複影像訪問-4th：（ 2019+ ）。我們將各時間段的有效樣本數列出，繪製成（附錄表二）。並發現九成以上的可用資料，皆集中於初步評估訪問-1st，其餘時間段皆有著大量的缺漏值，因此我們放棄使用重複測量的形式來評估受試者的去脂肪體重，而是使用平均的方式，將同一個受試者中四個時間段中可取得的數值進行平均。



2.2 樣本篩選-英國生物資料庫

篩選流程：

請參閱（圖二 A），為樣本篩選的具體流程，總計有 502,505 位志願者參與英國生物樣本資料庫，由於本研究的研究疾病為乳癌，因此僅保留 273,382 女性志願者，根據 ICD10 編碼將女性志願者定義為 259,383 位非癌症患者，與 14,105 位的乳癌患者。隨後刪除臨床項目的缺失值，並依據種族與基因主成分分析進行種族篩選。最後按照 1:4 的病例/對照比例隨機抽樣，作為本研究的研究樣本。總計包含了 39,388 的非癌症女性作為對照，以及 9,554 筆乳癌女性患者作為病例。

種族分群與篩選：

利用英國生物資料庫的調查資料，我們可以將受試者劃分為英國人、愛爾蘭人、其他白人、亞洲人、黑人、中國人、混血以及其他種族，參見（表一）。並從中挑選出英國籍白人作為本研究關心的族群，同時透過英國生物資料庫中所提供的篩選標準，利用主成分分析針對英國籍白人進行進一步的篩選，僅納入主成分分析非常相似的組群進行分析²⁹，如（附錄圖一）、（附錄圖二）所示。

基因分型：

英國生物資料庫中的基因數據，包含 488,377 名參與者的基因型，有兩個主要的微陣列晶片，49,950 名參與者使用 Applied Biosystems UK BiLEVE Axiom Array，438,427 名參與者使用 Applied Biosystems UK Biobank Axiom Array。兩個微陣列晶片間共享 95% 的變異位點²⁹。

品質控制-未經插補的基因變異位點資料：

在基因變異位點資料的篩選上，首先我們針對變異位點資料中，評估變異位點缺失值大於 5% 的樣本，如果每個變異位點中，超過 5% 的人沒有該變異位點，則刪除該變異位點，此步驟刪除 217 個變異位點。並基於哈溫平衡（0.000001）刪除了 31 個變異位點，以及最小等位變異位點頻率（0.01），刪除了 533 個變異位點，總計 455,271 個變異位點作為研究之基因變異位點。

品質控制-經插補的基因變異位點資料，：

我們使用英國生物資料庫中所提供的插補後基因變異位點資料，其中 Haplotype Reference Consortium (HRC) 數據作為主要的插補參考面板，因為其包含最大的廣泛歐洲單倍型²⁹。並針對進行後續的變異位點篩選。評估變異位點缺失值大於 5% 的樣本，如果每個變異位點中，超過 5% 的人沒有該變異位

點，則刪除該變異位點，此步驟刪除 637,143 個變異位點。並基於哈溫平衡 (0.000001) 刪除了 2,793,418 個變異位點，以及最小等位變異位點頻率 (0.01)，刪除了 194,594 個變異位點，總計 5,939,551 個變異位點作為研究之基因變異位點。



2.3 本研究方法論述

孟德爾隨機化試驗使用遺傳變異作為工具變量，來推斷曝露變項是否會對結果變項產生因果影響。近年來，孟德爾隨機化試驗已被用於結合大量遺傳變異的全基因組關聯分析匯總數據。其一個優點在於，根據孟德爾隨機化試驗三大假設，可以收集支持曝露變項和結果變項之間因果關係的證據。

在本研究中，我們使用孟德爾隨機化試驗，將結果變項定義為乳癌，曝露變項定義為去脂肪體重，希望研究去脂肪體重與乳癌間的因果關係。研究架構如 (圖一) 所示，使用經過篩選後的英國生物資料庫資料來進行分析。總計包含了 39,388 的非癌症女性作為對照，以及 9,554 筆乳癌女性患者作為病例。並定義去脂肪體重為曝露變項，乳癌的發生風險為所關心的結果變項。我們希望透過孟德爾隨機化，使用與去脂肪體重相關的變異位點作為工具變數，來探討去脂肪體重與乳癌間，是否存在特定的因果關係，以及因果關係的方向。

在孟德爾隨機化試驗的架構下，我們需要三大假設，來確保估計出的統計量，在通常情況下擁有較好的一致性與不偏性。這三大假設如 (圖三) 所示，分別為假設一：獨立性，即所使用的工具變數必須與干擾因子無關；假設二：相關性，所使用的工具變數必須與曝露變項相關；假設三，排他性：所使用的工具變數，僅能透過曝露變項的這條路徑來影響結果變項。

為了進行孟德爾隨機化試驗，我們先分別建立兩個全基因組關聯分析模型，分別針對了曝露變項 (去脂肪體重、全身脂肪體重) 以及結果變項 (乳癌風險)，建立全基因組關聯分析-模型一與全基因組關聯分析-模型二：

全基因組關聯分析-模型一主要是為了尋找與去脂肪體重和全身脂肪重量相關的工具變數。分別將去脂肪體重與全身脂肪重量作為被解釋變項 X 。由於兩者皆為一個連續的變項，因此使用複線性迴歸來作為全基因組關聯分析模型，針對 j 個變異位點進行全基因組關聯分析，加入了基因資料的主成份分析 (PC1~PC10)，作為基因上的調整，以調整英國生物資料庫中，由祖先以及人種導致的結構性差異。

全基因組關聯分析-模型一 (去脂肪體重)：

$$FFM | G_j = \gamma_{0,j} + \gamma_j G_j + K_{1,j} PC_1 \sim K_{10,j} PC_{10} \quad (1.1)$$

全基因組關聯分析-模型一（全身脂肪體重）：

$$BFM | G_j = \gamma_{0,j} + \gamma_j G_j + K_{1,j} PC_1 \sim K_{10,j} PC_{10} \quad (1.2)$$



我們參考了英國癌症研究網站的乳癌風險因子清單⁷，並從中選出了身高、飲酒頻率、抽菸頻率、是否擁有糖尿病、年齡、生育數量、初經年齡這些英國生物資料庫中擁有的臨床變項，作為乳癌可能的干擾因子。在（表二）中，我們列出了上述的臨床變項中，病例與對照間的兩樣本 t 檢定與皮爾森卡方檢定的 P-value，雖然飲酒頻率在該表中並沒有統計顯著的差異，但是基於臨床上的考量，我們還是將其加入後續的全基因組關聯分析以及孟德爾隨機化試驗的模型中進行分析。

在全基因組關聯分析-模型二中，將乳癌風險作為被解釋變項 Y，乳癌風險我們使用病例-對照的虛擬變數，並使用邏輯式迴歸來作為全基因組關聯分析模型，針對 j 個變異位點進行全基因組關聯分析，放入了身高、飲酒頻率、抽菸頻率、是否擁有糖尿病、年齡、生育數量、初經年齡作為臨床上的干擾因素，並加入了基因資料的主成份分析（PC1~PC10），作為基因上的調整，以調整英國生物資料庫中，由祖先以及人種導致的結構性差異。

全基因組關聯分析-模型二：

$$P = P(Y = 1 | G_j)$$

$$\log\left(\frac{p}{1-p}\right) = \gamma_0 + \Gamma_j G_j + W_{k,j} Clinical_k + M_{1,j} PC_1 \sim M_{10,j} PC_{10} \quad (2)$$

為了符合上述的孟德爾隨機化試驗三大假設，在獲得全基因組關聯分析模型一、模型二的結果後，我們依據孟德爾隨機化試驗的假設二：所使用的工具變數必須與曝露變數相關，保留全基因組關聯分析模型一中，p-value 小於 $10e^{-5}$ 的基因，作為可能的工具變數候選基因，並依照孟德爾隨機化試驗的標準流程，僅保留 F-檢定 >10 的變異位點，以避免弱工具變數效應造成的偏誤，其後再針對連鎖不平衡（Linkage disequilibrium, LD）來進行處理，使用 200 鹼基作為窗口大小，以每次 5 鹼基的方式滑動窗口，並將窗口內相關係數 >0.5 的變異位點篩選出，僅保留 p-value 最小的變異位點。處理完連鎖不平衡的問題後，我們將這些變異位點視作曝露變項對結果變項的工具變數，在本文中，即是去脂肪體重的工具變數。

我們首先使用英國生物資料庫的資料，進行單樣本的孟德爾隨機化試驗，使用兩階段方法³⁰，評估肥胖測量對乳腺癌。在第一步，我們利用所選出的去脂肪

體重工具變數，對去脂肪體重進行複線性迴歸，並在第二步，利用去脂肪體重估計值對乳癌風險進行邏輯式迴歸。

在單樣本的孟德爾隨機化試驗中，我們可以使用二階最小平方法來估計因果效應，但如直接應用雙樣本的孟德爾隨機化試驗估計方法，則會產生偏差。與雙樣本孟德爾隨機化試驗不同的是，在單樣本 MR 中，兩個估計值是從同一個人獲得的，兩個估計值間會存在共變異數，並且在存在混淆因子的情況下，孟德爾隨機化試驗的獨立性假設不成立。

不過因為在相關的孟德爾隨機化試驗方法論中提及，雖然在單樣本孟德爾隨機化試驗中，使用雙樣本孟德爾隨機化試驗的估計方法，可能會造成偏差，但工具變數強度足夠的情況下，除了 MR-Egger 以外的估計方法，會與二階最小平方法擁有一致的結果³¹。因此在單樣本孟德爾隨機化試驗中，我們同樣會使用逆方差加權（Inverse variance weighted）、MR Egger、加權中位數（weighted median）方法來輔助估計因果關係^{32 33}，並比較這四項估計方法間的一致性。

隨後，根據全基因組關聯分析模型一、模型二的結果，我們定義 $\hat{\beta}_j = \frac{\hat{\gamma}_j}{\hat{\nu}_j}$ 來衡量曝露變項與結果變項間的因果關係。我們使用孟德爾隨機化試驗的估計方法，對曝露變項的工具變數來進行估計。根據孟德爾隨機試驗方法的薈萃分析結果，目前主流的作法以逆方差加權來統合各個變異位點的 Wald 比率估計³⁴。因此在本文，將以逆方差加權作為雙樣本孟德爾隨機化試驗的主要衡量依據，並輔以 MR Egger、加權中位數方法來統合各個變異位點的值，並計算去脂肪體重對乳癌風險的因果關係。

$$\hat{\beta}_j = \frac{\hat{\gamma}_j}{\hat{\nu}_j} \quad (3)$$

本研究使用 R 中的 TwoSampleMR³⁵ 包來進行孟德爾隨機化試驗。並使用 R 4.1.0 進行分析並生成圖³⁶。

2.4 孟德爾隨機化試驗方法論述

在過去的五年中，全基因組關聯分析帶來了許多科學發現，而孟德爾隨機化試驗的使用，則是基於全基因組關聯分析的結果上，帶來另一層面的應用，試圖在回顧性的資料中，藉由基因的單核苷酸多態性（Single Nucleotide

Polymorphisms, SNPs) 來作為工具變數，研究曝露變數與我們所關心的主要變項間的因果關係。在(圖一)中，我們以去脂肪重量與乳癌作為範例，視覺化了孟德爾隨機化試驗的運作方法。

孟德爾隨機化試驗使用遺傳變異作為曝露的工具變量，以調查曝露對疾病結果的因果影響的證據。孟德爾隨機化估計代表遺傳預測的風險因素水平與結果的因果關係。由於孟德爾的隔離和獨立分類定律，與傳統的觀察性流行病學研究相比，來自孟德爾隨機化的估計不易受到干擾因子的影響產生偏差。

在研究方法中，我們介紹了孟德爾隨機化試驗的三大假設，依序分別為獨立性；相關性；排他性。由於孟德爾隨機化試驗的三大假設在實務上常常面臨一些問題，故自然的延伸出了許多不同的方法來估計 $\hat{\beta}_j = \frac{\hat{\Gamma}_j}{\hat{\gamma}_j}$ ，以下我們將著重介紹兩種方法，分別為 MR Egger 以及加權中位數方法^{32 33}。

MR Egger 用以檢視所使用的變異位點，是否具有水平的基因多效性³²，當變異位點具有水平的基因多效性 (pleiotropy) 時，則表示該變異位點同時影響了曝露變項與結果變項，因此違反了孟德爾隨機化試驗三大假設中的假設三：排他性，進而造成逆方差加權會出現偏誤的情況。

加權中位數方法則是當所使用的眾多工具變量中，如果存在一些違反上述三大假設中的工具變數，則可能會使得估計出的產生 $\hat{\beta}_j$ 偏誤。加權中位數方法的估計方法³³，處理這類的問題時，將具有比逆方差加權較為穩健的統計性質。

2.5 敏感性分析

由於去脂肪體重與乳癌間的因果關係，在本文的研究結果中並不穩定，在不同的孟德爾隨機化試驗估計方法中有不一致的結果，因此我們透過以下敏感性分析，希望能夠獲得在不同估計方法間，較為穩定的結果。為了探究在不同的子群集間，其去脂肪體重與乳癌風險的因果關係，在方向上是否會出現差異，我們使用了兩個子群集分析，來驗證孟德爾隨機化試驗的結果，分群的依據分別為是否停經以及乳癌的組織分型。

陰性對照：

在本篇論文中，去脂肪體重作為曝露變數，乳癌風險作為結果變項，探討去脂肪體重對乳癌風險的因果關係，而為了論證這樣的因果關係的方向是單向的，

不存在兩者互為因果的可能性，我們進行了陰性對照。將乳癌風險倒過來作為曝露變數，去脂肪體重作為結果變項，使用乳癌風險的全基因組關聯分析結果，找出針對乳癌風險的工具變數，所使用的工具變數參考（附錄表六）。其隱含的因果關係意義為，探究罹患乳癌後的病人，其去脂肪體重是否會因為乳癌的罹患而受到影響。

子群集分析-是否停經：

在是否停經的分群上，我們參考了臨床上的考量，乳癌是種對賀爾蒙變化敏感的癌症，也因此常見的臨床變項中，總會討論乳癌患者的動情激素受體(ER)及黃體激素受體(PR)，而在停經後的女性，由於卵巢停止分泌相關的激素，故體內的雌激素含量，開始被體內的脂肪細胞所主導，也因此，停經前後的女性患者，去脂肪體重和乳癌風險上的因果關係可能有所不同，故嘗試將其視作兩個群體，來進行孟德爾隨機化試驗的分析。並在結果中討論兩個群體間，因果關係的差異。

子群集分析-組織亞型：

在乳癌組織的亞型分群上。參照各文獻^{37 38 39}，乳癌的亞型對於預後有著不同的反應，因此在不同的亞型上，去脂肪體重和乳癌風險上的因果關係有著不同的可能。由於英國生物資料庫乳癌亞型樣本數的限制，因此我們僅採用了乳腺癌、乳小葉癌這兩類乳癌組織亞型，在將女性乳癌患者分群後，進行後續的孟德爾隨機化試驗分析流程，闡述兩個組織亞型之間，因果關係的差異。

2.6 工具變數的相關分析

基因註解：

我們使用 Varied 線上工具來進行變異位點的註釋，該項工具是由國立台灣大學-生物信息學與生物統計學核心實驗室所開發⁴⁰，並使用 hg19 做為參考序列版本。希望透過變異位點的註釋，瞭解我們使用的去脂肪體重工具變數，在基因機轉上扮演的角色與可能的功能，後續使用 Genecards 來為尋找可能的基因功能⁴¹。

工具變數相關分析-全表型組關聯分析：

通常，全表型組關聯分析的目的是為檢視特定的變異位點，在哪一些疾病與對照組中，擁有顯著的差異。透過全表型組關聯分析的結果，我們可以知道這個特定的變異位點，在哪些疾病中有顯著差異，並提供了後續的研究人員，疾病與疾病的關聯，以及可能機轉的線索。

基於孟德爾隨機化試驗的假設，我們通常不會去在乎工具變數的生物意涵，因為在數學模型的角度上，我們僅僅只需要工具變數與孟德爾隨機化試驗間，存在相關性即可，並且這些工具變數不會反過來被結果影響，即得到乳癌與否，並不會影響受試者原本自身的基因型，因為受試者的基因型在出生時就已經決定好了。

而基於探索性的目的，我們希望透過全表型組關聯分析，研究是否有著其他疾病，遵循與乳癌發生風險相同的基因路徑，被去脂肪體重來影響，進而賦予我們的工具變數，生物上的詮釋意義。孟德爾隨機化試驗中，我們選出來的這些作為工具變數的變異位點，可以理解為這些變異位點會影響受試者的去脂肪體重，並透過影響受試者去脂肪體重的方式，間接的影響乳癌發生的風險，我們希望透過去脂肪體重，找到是否有其他的疾病在機轉上與該種基因路徑相似。

操作全表型組關聯分析時，我們使用英國生物資料庫的 ICD10 作為參照，並將其轉換成 phecode 方便後續的分析，同時因為我們關注大方向上的疾病，而非特定的疾病亞型，且由於有些疾病的發生率本身就較為罕見，出於增加疾病的病例/對照比例的考量，故在處理 ICD10 時，我們選擇忽略尾碼的方式，將疾病的部分分型進行合併。

2.7 外部驗證

在外部驗證中，我們使用雙樣本孟德爾隨機化試驗，著重不同人群中的探討，在主要的孟德爾隨機化試驗流程中，我們使用英國生物資料庫的資料進行單樣本孟德爾隨機化試驗分析，然而單樣本孟德爾隨機化試驗分析的結果較容易因為使用不同資料庫的改變，造成結果的不穩定，因此我們希望分別透過乳腺癌協會聯盟與台灣生物資料庫這兩個資料庫來進行外部驗證，並使用雙樣本孟德爾隨機化試驗分析來進行因果推論，驗證在歐洲人以及漢人兩大族群中，去脂肪體重與乳癌間的因果關係，是否存在方向上的一致性。

外部驗證-乳腺癌協會聯盟：

我們使用乳腺癌協會聯盟於 2020 年所發表的資料²⁸，進行雙樣本孟德爾隨機化試驗，使用英國生物資料庫在階段一的全基因組關聯分析所篩選出的工具變數，以及乳腺癌協會聯盟在階段二基於乳癌人群的全基因組關聯分析結果以孟德爾隨機化試驗統計方法進行因果推論。由於我們僅能取得全基因組關聯分析匯總數據，因此沒有辦法對全基因組關聯分析模型進行調整，與在單樣本孟德爾隨機化試驗的全基因組關聯分析模型會有不一致。

乳腺癌協會聯盟的主要人群為歐洲族群。在這部分的外部驗證中，我們希望拓展使用英國生物資料庫所得出的因果推論，研究在皆基於歐洲族群的資料庫中，去脂肪體重與乳癌風險的因果關係，是否具有一致性。



外部驗證-台灣生物資料庫：

在完成上一步的驗證後，由於癌症在不同種族間的差異性，本身就很大，從癌症好發盛行年齡、全基因組關聯分析結果等差異中，皆可以看出，即使是同一種癌症，不同種族間應當被分別討論。故我們使用台灣生物資料庫的資料，同樣進行了雙樣本孟德爾隨機化試驗，基於英國生物資料庫在階段一全基因組關聯分析所篩選出的工具變數，以及台灣生物資料庫在階段二基於乳癌人群的全基因組關聯分析結果。希望研究在不同人群間，去脂肪體重和全身脂肪體重兩者，與乳癌風險的因果關係是否仍是一致的。因為臨床變項上與英國生物資料庫的不一致，因此在全基因組關聯分析模型上，我們僅使用主成分分析來進行調整。

全基因組關聯分析-台灣生物資料庫：

$$P = P(Y = 1 | G_j)$$

$$\ln\left(\frac{p}{1-p}\right) = \gamma_0 + \Gamma_j G_j + M_{1,j} PC_1 \sim M_{10,j} PC_{10} \quad (4)$$

第三章 結果



3.1 資料展示與變數評估

(表一)中，我們呈現了英國生物資料庫中的人口學特徵，利用英國生物資料庫所提供的種族調查資料，將受試者劃分為英國人、愛爾蘭人、其他白人、亞洲人、黑人、中國人、混血以及其他種族，並針對性別、年齡、乳癌比例、抽菸頻率、飲酒頻率、是否患有糖尿病、生育歷史、初經年齡、停經年齡與停經狀態進行呈現。並在後續的結果中，僅使用英國人的參與者作為英國生物資料庫的研究族群。

在(表二)中，我們呈現了研究樣本的相關人口學特徵，在乳癌患者以及對照受試者中，身高、抽菸頻率、飲酒頻率、是否擁有糖尿病、年齡、生育數量、初經年齡，這些我們放入全基因組關聯分析中進行干擾因素調整的變項，在兩個族群間皆有顯著的兩樣本 t 檢定統計差異。

除了去脂肪體重外，也有許多肥胖相關的指標，因此在資料呈現上，我們使用了體脂率、體重、去脂肪體重、身體脂肪重量、腰臀比來呈現不同肥胖指標的差異，在(表一)中，我們利用兩樣本 t 檢定來檢驗乳癌患者以及對照受試者中是否有統計顯著上的差別，並且使用了單變項邏輯式迴歸，來瞭解在英國生物資料庫中，這些肥胖指標是否能夠作為乳癌患者的風險因子，以及在方向上是否擁有一致性。我們發現，在英國生物資料庫中，體脂率、體重、去脂肪體重、身體脂肪重量、腰臀比皆為乳癌的風險因子，並在(圖一)中，呈現了肥胖指標間對應的相關係數與資料分佈，而在對角線上的直方圖則是變項自身的分佈。

在附錄中，我們補充了台灣生物資料庫的相關人口學特徵，請參照(附錄表三)，並且由於孟德爾隨機化試驗，主要是基於全基因組關聯分析的結果，因此在附錄中我們補充了去脂肪體重與乳癌風險的曼哈頓圖與 QQ-plot，QQ-plot 的膨脹因子，請參照(附錄圖三)至(附錄圖六)，去脂肪體重的全基因組關聯分析膨脹因子為 1.179，全身脂肪體重的全基因組關聯分析膨脹因子為 1.16。



3.2 孟德爾隨機化試驗-主要結果

(表四)、(表五)為主要的孟德爾隨機化試驗分析結果，使用英國生物資料庫，來進行單樣本孟德爾隨機化試驗。分別以去脂肪體重以及身體脂肪體重來作為曝露變數，乳癌風險作為結果變數。我們分別使用了原始變異位點資料，以及經過插補的變異位點資料。再經過孟德爾隨機化試驗三大假設的篩選後，我們在原始資料與插補資料，分別使用了 97 與 164 個變異位點來作為去脂肪體重的工具變數，並使用了 51 與 98 個變異位點來作為身體脂肪體重的工具變數。

首先針對去脂肪體重的單樣本孟德爾隨機化試驗，無論是沒有經過插補亦或是經過插補的估計結果，使用二階最小平方法的方式皆沒有統計顯著 ($\hat{\beta}_{2SLS} = -0.012$, $OR_{2SLS} = 0.988$, $P_{2SLS} = 0.283$)，在方向上二階最小平方法、逆方差加權、MR Egger 與加權中位數方法皆為負值，然而僅僅只有逆方差加權與加權中位數方法達到統計顯著 ($\hat{\beta}_{IVW} = -0.024$, $OR_{IVW} = 0.976$, $P_{IVW} = 0.012$)。去脂肪體重與乳癌發生風險兩者的因果關係為負，即代表去脂肪體重的增加，將有助於降低乳癌發生的風險。同時在 MR Egger 的截距項估計並沒有統計顯著的差異，故並不存在水平多效性的問題，來違反孟德爾隨機化試驗的假設。

再來則是身體脂肪體重的單樣本孟德爾隨機化試驗，無論是沒有經過插補亦或是經過插補的估計結果，在二階最小平方法、逆方差加權、MR Egger 與加權中位數方法的四種估計方法中，方向上均為一致，並且皆沒有統計顯著。

此外，除了去脂肪體重以及全身脂肪體重以外，我們也針對了其他肥胖指標進行了相同流程的單樣本孟德爾隨機化試驗。分別為體脂率、去脂肪體重指數以及全身脂肪體重指數。去脂肪體重指數以及全身脂肪體重指數為去脂肪體重與全身脂肪體重各自除以身高的平方，單樣本孟德爾隨機化試驗的結果如(附錄表七)、(附錄表八)與(附錄表九)所示，皆無統計顯著的因果推論。

3.3 敏感性分析

陰性對照：

為了論證這樣的因果關係的方向是單向的，不存在兩者互為因果的可能性，我們進行了陰性對照，來探討如果得了乳癌，會不會反過來影響患者的去脂肪體重。首先使用乳癌風險的全基因組關聯分析結果，找出 74 變異位點代表針對乳癌風險的工具變數，所使用的工具變數參考(附錄表六)。如(表六)所示 ($\hat{\beta}_{2SLS} = 0.352$, $OR_{2SLS} = 1.422$, $P_{2SLS} = 0.296$; $\hat{\beta}_{IVW} = 0.062$, $OR_{IVW} = 1.064$,

$P_{IVW} = 0.175$)，罹患乳癌並不會影響患者的去脂肪體重，研究中所發現的因果關係為單向的，去脂肪體重能夠影響乳癌發生的風險，反之則不成立，進一步驗證了主分析的結果。

子群集分析-是否停經：

在（表七）探索性研究-孟德爾隨機化試驗群集分析(是否停經)：我們針對不同的月經狀態子群集來分別進行孟德爾隨機化試驗的分析。因為在臨床上，我們可以合理地認為兩種患者的致病機轉是不相同的。停經後確診 ($\hat{\beta}_{2SLS} = -0.043$, $OR_{2SLS} = 0.958$, $P_{2SLS} = 0.014$ ； $\hat{\beta}_{IVW} = -0.083$, $OR_{IVW} = 0.920$, $P_{IVW} = 0.013$)；停經前確診 ($\hat{\beta}_{2SLS} = -0.028$, $OR_{2SLS} = 0.972$, $P_{2SLS} = 0.539$ ； $\hat{\beta}_{IVW} = 0.002$, $OR_{IVW} = 1.002$, $P_{IVW} = 0.914$)。

可以發現在孟德爾隨機化試驗的因果推論上，僅有停經後確診的子群有統計顯著的負相關，並且其負向因果關係的係數相較於不分亞型討論的因果推論係數更具有負向關係，去脂肪體重對乳癌風險的保護力也顯得更強。故在子群集分析中，將兩者拆開討論因果關係確實有其必要性。此外，針對停經後確診的雙樣本試驗中，MR Egger 的截距項統計顯著，可能有潛在的基因多效性，導致孟德爾隨機化試驗的排他性假設被違反，但因為 MR Egger 在單樣本孟德爾隨機化試驗中的估計結果本身存在較多問題³¹，故在此處並沒有著重於討論 MR Egger 截距項的顯著問題。

子群集分析-組織亞型：

在（表八）探索性研究-孟德爾隨機化試驗群集分析(組織亞型)：我們希望針對不同的乳癌組織亞型，分別進行孟德爾隨機化試驗的處理，但是由於主要的組織亞型，樣本集中在乳腺癌與乳小葉癌上，故我們僅使用這兩個亞型來進行孟德爾隨機化試驗的敏感性分析，乳腺癌 ($\hat{\beta}_{2SLS} = -0.021$, $OR_{2SLS} = 0.979$, $P_{2SLS} = 0.134$ ； $\hat{\beta}_{IVW} = -0.030$, $OR_{IVW} = 0.970$, $P_{IVW} = 0.005$)，乳小葉癌 ($\hat{\beta}_{2SLS} = -0.135$, $OR_{2SLS} = 0.873$, $P_{2SLS} = 0.015$ ； $\hat{\beta}_{IVW} = -0.035$, $OR_{IVW} = 0.966$, $P_{IVW} = 0.265$)。

在組織亞型的子群集分析中，因果關係的推論並不穩定，乳腺癌的子群集中，雖然逆方差加權與加權中位數方法的估計方法顯著為負向因果關係，然而二階最小平方方法卻並不顯著。在乳小葉癌的子群集中，情況恰恰相反，僅有二階最小平方方法有顯著的負向因果關係。



3.4 工具變數的相關分析

針對基因的註釋分析：

在（表九），我們將通過英國生物資料庫篩選出的 164 個去脂肪體重工具變數進行註釋，將變異位點對應的基因找出，並刪除了部分找不到對應基因的變異位點，依照染色體位置依序排列。在 164 個去脂肪體重工具變數中，共有 9 個工具變數之變異位點，集中位於基因 *DLEU1* 上，分別有 5 個變異位點落在 *L3MBTL3*、*LINC01122* 等基因上。

DLEU1：

DLEU1 (Deleted In Lymphocytic Leukemia 1) 是一個 RNA 基因，屬於 lncRNA 類。與 *DLEU1* 相關的疾病包括卵巢上皮癌和白血病、慢性淋巴細胞。

L3MBTL3：

該基因編碼惡性腦瘤 (MBT) 染色質相互作用轉錄抑制因子家族的成員。編碼的蛋白質可以調節造血功能。已在患有髓母細胞瘤的人類患者中觀察到該基因的缺失。

LINC01122：

LINC01122 (Long Intergenic Non-Protein Coding RNA 1122) 是一種 RNA 基因，屬於 lncRNA 類。與 *LINC01122* 相關的疾病包括大麻依賴症。

工具變數相關分析-全表型組關聯分析：

我們基於每一個工具變數單獨進行全表型組關聯分析，隨後統合全表型組關聯分析顯著的相關疾病有哪一些。如（圖五）所示，該圖為 rs12055445 的全表型組關聯分析的曼哈頓圖，針對該變異位點而言，有三個臨床病徵通過了邦費羅尼校正，即與 rs12055445 有統計顯著的相關。

依序針對使用英國生物資料庫資料所計算出的 164 個去脂肪體重工具變數，都單獨進行全表型組關聯分析分析，我們將通過邦費羅尼校正的臨床病徵統合成（表十）與（表十一）。腸道吸收不良、第一型糖尿病與多發性硬化症，分別與部分的變異位點有著統計顯著的相關，並且站在臨床的角度，這三項病徵與肥胖間，也有著合理的機轉可能，在後續的研究中，我們或許能針對上述三項表型進行孟德爾隨機化試驗的因果分析，針對肥胖可能造成的其他疾病，來進行探討。



3.5 外部驗證

外部驗證-乳腺癌協會聯盟：

在（表十二），我們透過從英國生物資料庫中，使用 impute data 所找出的 164 的變異位點作為去脂肪體重的工具變數，但由於乳腺癌協會聯盟的匯總性資料中，僅有 141 個變異位點與英國生物資料庫中吻合，因此僅採用 141 個變異位點來進行雙樣本的孟德爾隨機化試驗。我們利用乳腺癌協會聯盟 2020 所發佈的資料，進行雙樣本孟德爾隨機化試驗的外部驗證。並得到結果。在逆方差加權、MR Egger、加權中位數方法的因果推論皆不顯著，因此我們認為在歐洲人群中，去脂肪體重與乳癌風險之間，應當不存在因果關係。

外部驗證-台灣生物資料庫：

在（表十四），利用台灣生物資料庫的資料，進行雙樣本孟德爾隨機化試驗的外部驗證，我們同樣使用由英國生物資料庫所得到的 164 的變異位點來進行雙樣本孟德爾隨機化試驗，但僅有 93 個變異位點與台灣生物資料庫的資料吻合，並作為第一階段的去脂肪體重工具變數。在第二階段的全基因組關聯分析結果則是使用台灣生物資料庫的乳癌病例/對照進行分析，得到 ($\hat{\beta}_{IVW}=-0.056$, OR=0.946, P-value=0.448) 的因果推論結果，使用漢族族群進行雙樣本孟德爾隨機化試驗分析，並沒有得到具備統計顯著的因果推論。

。

第四章 結論與討論

4.1 主要發現

在（表三）中，我們使用體脂率、體重、去脂肪體重、身體脂肪體重以及臀腰比來進行單變項邏輯式迴歸，發現其實從觀察性研究的角度而言，無論是哪一種肥胖指標皆與乳癌的發生風險呈現正相關。並且透過繪製相關矩陣，我們可以在（圖四）中觀察到，大部分的肥胖指標，其實都有著非常高的相關性，因此不同肥胖指標間的使用，並不應當影響後續觀察到的因果關係。

正如同在研究動機中所提及的，本篇論文的研究目的，是希望透過將體重拆分為去脂肪體重以及身體脂肪體重，來探討在前人文獻中，肥胖對乳癌的影響在觀察性研究與孟德爾隨機化試驗有著相反方向的原因^{16 17 18 20}。

在本研究中，首先針對歐洲族群中的英國白人進行單樣本孟德爾隨機化試驗，以二階最小平方法的估計結果為主要因果推論，並以逆方差加權、MR Egger 與加權中位數方法進行輔助，我們在主要流程中，同時使用了原始的基因變異位點資料，以及經過插補後的基因變異位點資料，兩者在因果推論的方向與 P 值皆具有一致性，因此在後續的敏感性與外部驗證中，我們皆以經過插補後的基因變異位點資料來進行分析。

透過英國生物資料庫的單樣本孟德爾隨機化試驗，我們發現在去脂肪體重與乳癌風險間存在負向的因果關係，但這樣的因果關係在不同的估計方式中並不穩定。而全身脂肪重量與乳癌風險間，則與臨床上的常理認知：肥胖為乳癌的風險因子相反，在各種因果估計方法間，皆不存在因果關係。

有鑒於去脂肪體重與乳癌風險間的負向因果關係並不穩定，因此我們在後續分析中，又進行了陰性對照、子群集分析。在陰性對照中，我們證明了乳癌的罹患與否，並不能反過來影響去脂肪體重的增減，這是一條單向的因果關係。而在子群集分析中，去脂肪體重與乳癌風險間的負向關係，依舊不穩定，在主要的二階最小平方法估計方法中，僅有停經後確診以及乳小葉癌這兩類子群集有統計顯著的負向因果關係。

同時我們針對所使用的 164 個去脂肪工具變數，進行了基因變異位點註釋以及全表型組關聯分析，但是在基因功能上，大部分的基因與乳癌並沒有相關。

最後，我們使用雙樣本孟德爾隨機化試驗，來進行外部驗證，同時雙樣本孟德爾隨機化試驗相較於單樣本孟德爾隨機化試驗有著更加穩健的結果。然而不管是使用基於歐洲人群的乳腺癌協會聯盟資料庫，亦或是台灣生物資料庫的漢族資料，我們皆沒有觀察到去脂肪體重與乳癌風險間的因果關係，基於這樣的結果，我們認為去脂肪體重與乳癌風險間的因果關係，並不穩定，會隨著樣本資料的改變，產生不一致的結果。

4.2 研究討論

在本研究中，我們將體重指標進行拆分，全身脂肪體重在單樣本孟德爾化隨機試驗與雙樣本，皆沒有明顯的因果關係。這違反了在肥胖與乳癌相關的臨床研究中，肥胖作為停經後乳癌的印象^{15 16}。然而，在近幾年針對肥胖與乳癌的孟德爾隨機化試驗中，在肥胖指標與乳癌風險上的因果推論，我們找到了支持本研究結果的相關證據。皆為負相關抑或是不相關^{20 21 22 23}。臨床上的相關解釋，脂肪除了透過增加內源性激素的方式來增加乳癌風險外²⁴，也可能透過儲

存親脂性持久性有機污染物的方式，作為乳癌的保護因子存在^{25 25}。因此我們推測，脂肪細胞在兩種不同因果方向的拉扯下，造成了全身脂肪體重與乳癌風險沒有明顯的因果關係。

我們特別關心，在針對肥胖與乳癌的孟德爾隨機化試驗中，作者所提出的假說：身體活動相關的變量（即爬樓梯的頻率和適度的身體活動）可能是乳癌風險的干擾因子來源之一²⁰。並且在另一篇文獻中確認了，基於加速度計測量的運動狀態，能夠降低乳癌的發生風險²⁶。透過本研究的單樣本孟德爾隨機化試驗，我們發現了去脂肪體重與乳癌風險間的負向因果關係，去脂肪體重作為乳癌的保護因子存在，印證了上述兩篇研究中所提出的觀點。雖然在不同的統計方法間，去脂肪體重的負向因果關係並不穩定，即使我們在敏感性試驗中，將實驗樣本限縮於部分的子群間，這樣的情況也並未改善。主要的原因有兩種可能，一為MR-Egger本身在應用於單樣本孟德爾隨機化試驗的估計時較為不穩定³¹，二為沒有體重的長期觀測資料，由於英國生物資料庫的限制，如同我們在（附錄表二）所示，九成以上的體重相關數據皆集中在2006-2010的初步評估訪問中，並沒有辦法有效反應受試者在不同時期體重的變化情況。

與單樣本孟德爾隨機化試驗不同，在雙樣本孟格爾隨機化試驗中，不論是基於歐洲人口的乳腺癌協會聯盟還是漢族人口的台灣生物資料庫，我們都並未獲得統計顯著的因果關係推論。這部分的原因可能會與人種差異以及收案條件有關。在人種差異的部分，英國生物資料庫使用的人種為英國籍白人，而乳腺癌協會聯盟所使用的為整個歐洲族群，不同種族間的研究，以薈萃分析的方式整合為一全基因組關聯分析結果。並且在台灣生物資料庫的部分，相較於英國生物資料庫所使用的英國籍白人，台灣生物資料庫收納的為台灣漢人組群，西方人種與東方人種的差異本身非常大，因此直接使用由英國生物資料庫所選出的工具變數，來進行雙樣本的孟德爾隨機化試驗並不嚴謹，但是礙於體脂肪的測定資料在取得上的不易，我們無法取得去脂肪體重與全身脂肪體重基於東亞族群的全基因組關聯分析結果，因此在本文中僅能使用基於英國生物資料庫所篩選出的工具變數來進行跨族群的雙樣本孟德爾隨機化試驗。

此外，不同生物資料庫中，收案條件的不同也造成了流行病學上的選擇偏誤，在英國生物資料庫的收案中，並未針對癌症患者來進行排除，然而台灣生物資料庫的收案標準中，預先排除了癌症患者⁴²。兩個資料庫間的收案差異，也將很大幅度的影響雙樣本孟格爾隨機化試驗的結果。



4.3 結論

在前人文獻中，肥胖對乳癌的影響在觀察性研究與孟德爾隨機化試驗有著相反方向的原因^{16 17 18 20}，在上述提及的孟德爾隨機化試驗的研究中，肥胖指標皆為乳癌風險的保護因子。因此在本研究中，我們將體重指標進行拆分為去脂肪體重與全身脂肪體重，希望評估針對乳癌風險的保護作用，究竟是基於脂肪重量的多寡還是肌肉重量的多寡。

本研究使用英國生物資料庫進行的單樣本孟德爾隨機化試驗，確認了去脂肪體重為乳癌風險的保護因子，兩者存在負向的因果關係，並且不存在反向因果關係。而全身脂肪體重與乳癌風險間，則不存在任何統計顯著的因果關係。在雙樣本孟德爾隨機化試驗中，我們基於乳腺癌協會聯盟與台灣生物資料庫的資料進行外部驗證，但是並沒有得到統計顯著上的因果推論。我們在 4.2 研究討論中提出了基於人種差異以及收案條件上的可能原因。

透過本研究，我們針對肥胖對於乳癌，在孟德爾隨機化試驗中不尋常的保護作用，提出了可能的解釋，利用將體重拆分成去脂肪體重與全身脂肪體重，分別進行單樣本孟德爾隨機化試驗與雙樣本隨機化試驗，我們可以更清楚的了解到，究竟是脂肪重量的增加，抑或是肌肉重量的增加，導致了乳癌風險的下降。並得出了是肌肉重量（去脂肪體重）的增加降低了乳癌風險的可能。

經過合理研究設計，與孟德爾隨機化試驗的統計推論，我們發現去脂肪體重的增加，有助於預防乳癌風險的發生，基於衛教的角度，除了鼓勵民眾多加運動外，針對更年期後的女性，我們更應該注意其在更年期後，運動習慣是否發生改變，亦或是其肌肉流失的狀態，皆有可能造成乳癌風險的增加。而倘若從治療的角度出發，基於脂肪與乳癌風險間的因果關係，主要的可能途徑是透過激素水平這一方式來影響，在後續的研究中，我們也許可以嘗試研究在女性更年期後，給予賀爾蒙藥物治療的方式，是否會造成其乳癌風險增加這樣的因果關係研究。

4.4 研究限制

進行雙樣本的孟德爾隨機化試驗時，針對乳腺癌協會聯盟資料庫，基於隱私問題，我們僅能取得全基因組關聯分析的資料，因此在使用不同的資料庫來進行分析時，難以維持全基因組關聯分析模型的一致性，這部分可能會造成因果推論上產生偏誤。並且在台灣生物資料庫的外部驗證中，由於去脂肪體重與全身脂肪體重的資料在東亞族群上取得不易，因此我們只好以英國生物資料庫所選

出的工具變數來進行雙樣本的孟德爾隨機化試驗，此舉可能為我們的雙樣本孟德爾隨機化試驗的推論帶來偏差。

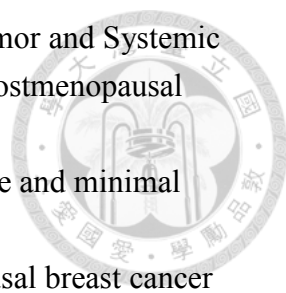
此外，在常規的乳癌風險因子中，ER、PR、HER2 這些賀爾蒙受器的臨床分型，是非常必要且有效的分類方式，基於這些分型差異，預後的差距也非常大，可以預見在癌症的發病機轉上，也會有很大的差異。但由於在英國生物資料庫中的資料中，並沒有這些相關的乳癌風險因子，故沒有辦法確認不同的賀爾蒙受器亞型，是否在去脂肪體重與乳癌風險的因果關係上，會產生不一樣的推論。

最後，由於長期追蹤資料所需花費的成本太大，在目前的肥胖與乳癌的孟德爾隨機化試驗的相關研究中，皆使用單次的資料來進行因果推論，並未結合長期追蹤的體重變化。並且肥胖是一項非常容易受到生活環境影響的個人表型⁴³，因此在這樣的研究設計上，孟德爾隨機化試驗的研究結果並沒有辦法解釋受試者在飲食或是生活環境上帶來的改變，會對乳癌帶來什麼樣的影響。

參考文獻

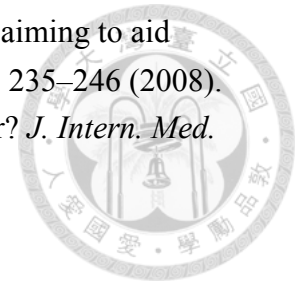


1. World Health Organization, WHO. Obesity. https://www.who.int/health-topics/obesity#tab=tab_1.
2. Bhaskaran, K. *et al.* Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5·24 million UK adults. *The Lancet* **384**, 755–765 (2014).
3. Keimling, M. *et al.* Comparison of Associations of Body Mass Index, Abdominal Adiposity, and Risk of Colorectal Cancer in a Large Prospective Cohort Study. *Cancer Epidemiol. Biomarkers Prev.* **22**, 1383–1394 (2013).
4. Wang, L. *et al.* Cancer incidence in relation to body fatness among 0.5 million men and women: Findings from the China Kadoorie Biobank. *Int. J. Cancer* **146**, 987–998 (2020).
5. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* **68**, 394–424 (2018).
6. 衛生福利部國民健康署. 衛生福利部國民健康署-乳癌防治.
7. World Cancer Research Fund, WCRF. Obesity, weight and cancer risk - Preventing cancer. *World Cancer Research Fund* <https://www.wcrf-uk.org/preventing-cancer/what-can-increase-your-risk-of-cancer/obesity-weight-and-cancer-risk/>.
8. American Institute for Cancer Research, AICR. What You Need To Know About Obesity and Cancer. *American Institute for Cancer Research* <https://www.aicr.org/resources/media-library/what-you-need-to-know-about-obesity-and-cancer/>.
9. Kyrgiou, M. *et al.* Adiposity and cancer at major anatomical sites: umbrella review of the literature. *BMJ* **356**, j477 (2017).
10. Fewell, Z., Davey Smith, G. & Sterne, J. A. C. The Impact of Residual and Unmeasured Confounding in Epidemiologic Studies: A Simulation Study. *Am. J. Epidemiol.* **166**, 646–655 (2007).
11. Matthews, S. B. *et al.* Excess Weight Gain Accelerates 1-Methyl-1-Nitrosourea-Induced Mammary Carcinogenesis in a Rat Model of Premenopausal Breast Cancer. *Cancer Prev. Res. (Phila. Pa.)* **7**, 310–318 (2014).
12. De Angel, R. E. *et al.* The enhancing effects of obesity on mammary tumor growth and Akt/mTOR pathway activation persist after weight loss and are reversed by RAD001. *Mol. Carcinog.* **52**, 446–458 (2013).

- 
13. Giles, E. D. *et al.* Obesity and Overfeeding Affecting Both Tumor and Systemic Metabolism Activates the Progesterone Receptor to Contribute to Postmenopausal Breast Cancer. *Cancer Res.* **72**, 6490–6501 (2012).
14. Ecker, B. L. *et al.* Impact of obesity on breast cancer recurrence and minimal residual disease. *Breast Cancer Res.* **21**, 41 (2019).
15. Amadou, A. *et al.* Overweight, obesity and risk of premenopausal breast cancer according to ethnicity: a systematic review and dose-response meta-analysis. *Obes. Rev.* **14**, 665–678 (2013).
16. Parra-Soto, S. *et al.* Associations of six adiposity-related markers with incidence and mortality from 24 cancers—findings from the UK Biobank prospective cohort study. *BMC Med.* **19**, 7 (2021).
17. Freuer, D. *et al.* Body Fat Distribution and Risk of Breast, Endometrial, and Ovarian Cancer: A Two-Sample Mendelian Randomization Study. *Cancers* **13**, 5053 (2021).
18. Vithayathil, M. *et al.* Body size and composition and risk of site-specific cancers in the UK Biobank and large international consortia: A mendelian randomisation study. *PLOS Med.* **18**, e1003706 (2021).
19. Kyle, U. G., Schutz, Y., Dupertuis, Y. M. & Pichard, C. Body composition interpretation: Contributions of the fat-free mass index and the body fat mass index. *Nutrition* **19**, 597–604 (2003).
20. Amin, H. A. *et al.* Mendelian randomisation analyses of UK Biobank and published data suggest that increased adiposity lowers risk of breast and prostate cancer. *Sci. Rep.* **12**, 909 (2022).
21. Guo, Y. *et al.* Genetically Predicted Body Mass Index and Breast Cancer Risk: Mendelian Randomization Analyses of Data from 145,000 Women of European Descent. *PLOS Med.* **13**, e1002105 (2016).
22. 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).
23. Ahmed, M. *et al.* Adiposity and cancer: a Mendelian randomization analysis in the UK biobank. *Int. J. Obes.* **45**, 2657–2665 (2021).
24. Kabuto, M., Akiba, S., Stevens, R. G., Neriishi, K. & Land, C. E. A Prospective Study of Estradiol and Breast Cancer in Japanese Women1. *Cancer Epidemiol. Biomarkers Prev.* **9**, 575–579 (2000).
25. Jansen, A., Lyche, J. L., Polder, A., Aaseth, J. & Skaug, M. A. Increased blood levels of persistent organic pollutants (POP) in obese individuals after weight loss—A review. *J. Toxicol. Environ. Health Part B* **20**, 22–37 (2017).

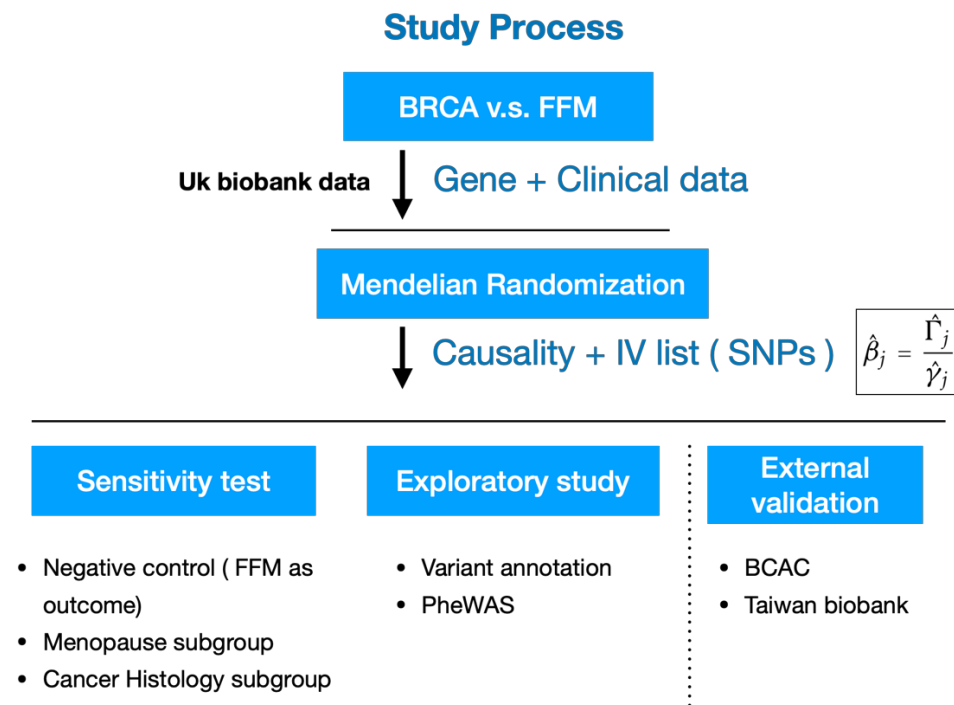
26. Papadimitriou, N. *et al.* Physical activity and risks of breast and colorectal cancer: a Mendelian randomisation analysis. *Nat. Commun.* **11**, 597 (2020).
27. Liu, Y.-S. *et al.* Body mass index and type 2 diabetes and breast cancer survival: a Mendelian randomization study. *Am. J. Cancer Res.* **11**, 3921–3934 (2021).
28. Zhang, H. *et al.* Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. *Nat. Genet.* **52**, 572–581 (2020).
29. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
30. Burgess, S., Small, D. S. & Thompson, S. G. A review of instrumental variable estimators for Mendelian randomization. *Stat. Methods Med. Res.* **26**, 2333–2355 (2017).
31. Minelli, C. *et al.* The use of two-sample methods for Mendelian randomization analyses on single large datasets. *Int. J. Epidemiol.* **50**, 1651–1659 (2021).
32. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–525 (2015).
33. 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).
34. Boef, A. G. C., Dekkers, O. M. & le Cessie, S. Mendelian randomization studies: a review of the approaches used and the quality of reporting. *Int. J. Epidemiol.* **44**, 496–511 (2015).
35. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *eLife* **7**, e34408 (2018).
36. R: The R Project for Statistical Computing. <https://www.r-project.org/>.
37. Colleoni, M. *et al.* Outcome of special types of luminal breast cancer. *Ann. Oncol.* **23**, 1428–1436 (2012).
38. McCart Reed, A. E., Kalinowski, L., Simpson, P. T. & Lakhani, S. R. Invasive lobular carcinoma of the breast: the increasing importance of this special subtype. *Breast Cancer Res.* **23**, 6 (2021).
39. Luveta, J., Parks, R. M., Heery, D. M., Cheung, K.-L. & Johnston, S. J. Invasive Lobular Breast Cancer as a Distinct Disease: Implications for Therapeutic Strategy. *Oncol. Ther.* **8**, 1–11 (2020).
40. | VariED. <http://varied.cgm.ntu.edu.tw/>.
41. Safran, M. *et al.* GeneCards Version 3: the human gene integrator. *Database* **2010**, baq020 (2010).

42. Fan, C.-T., Lin, J.-C. & Lee, C.-H. Taiwan Biobank: a project aiming to aid Taiwan's transition into a biomedical island. *Pharmacogenomics* **9**, 235–246 (2008).
43. Loos, R. J. F. & Bouchard, C. Obesity – is it a genetic disorder? *J. Intern. Med.* **254**, 401–425 (2003).

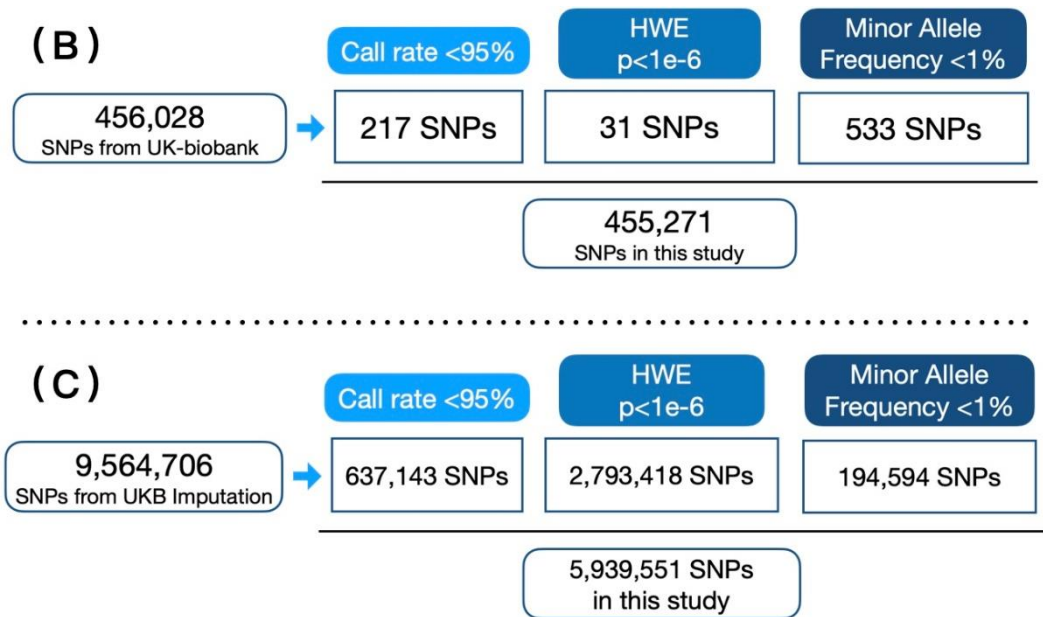
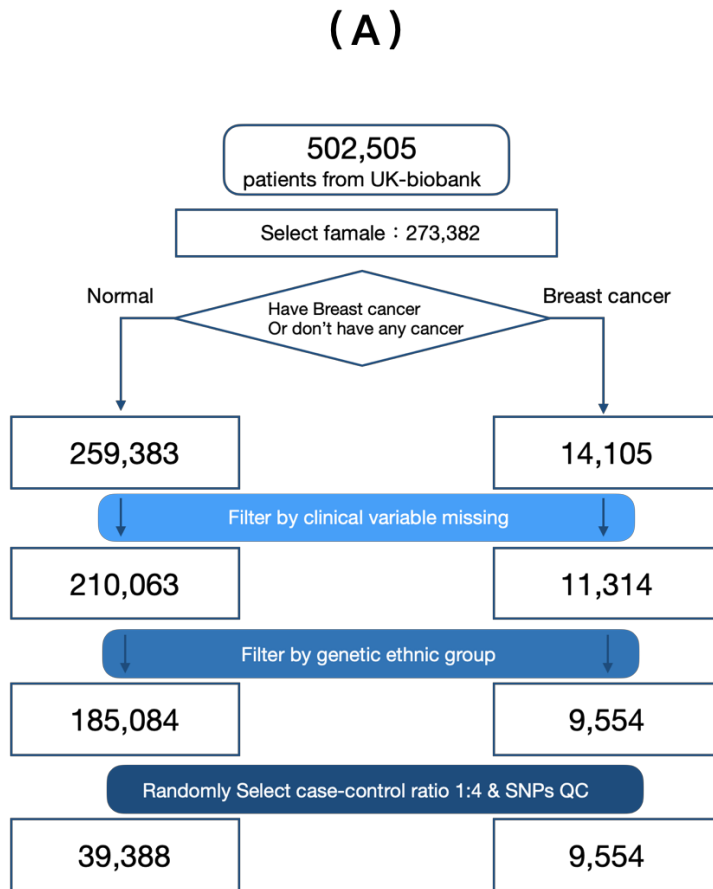




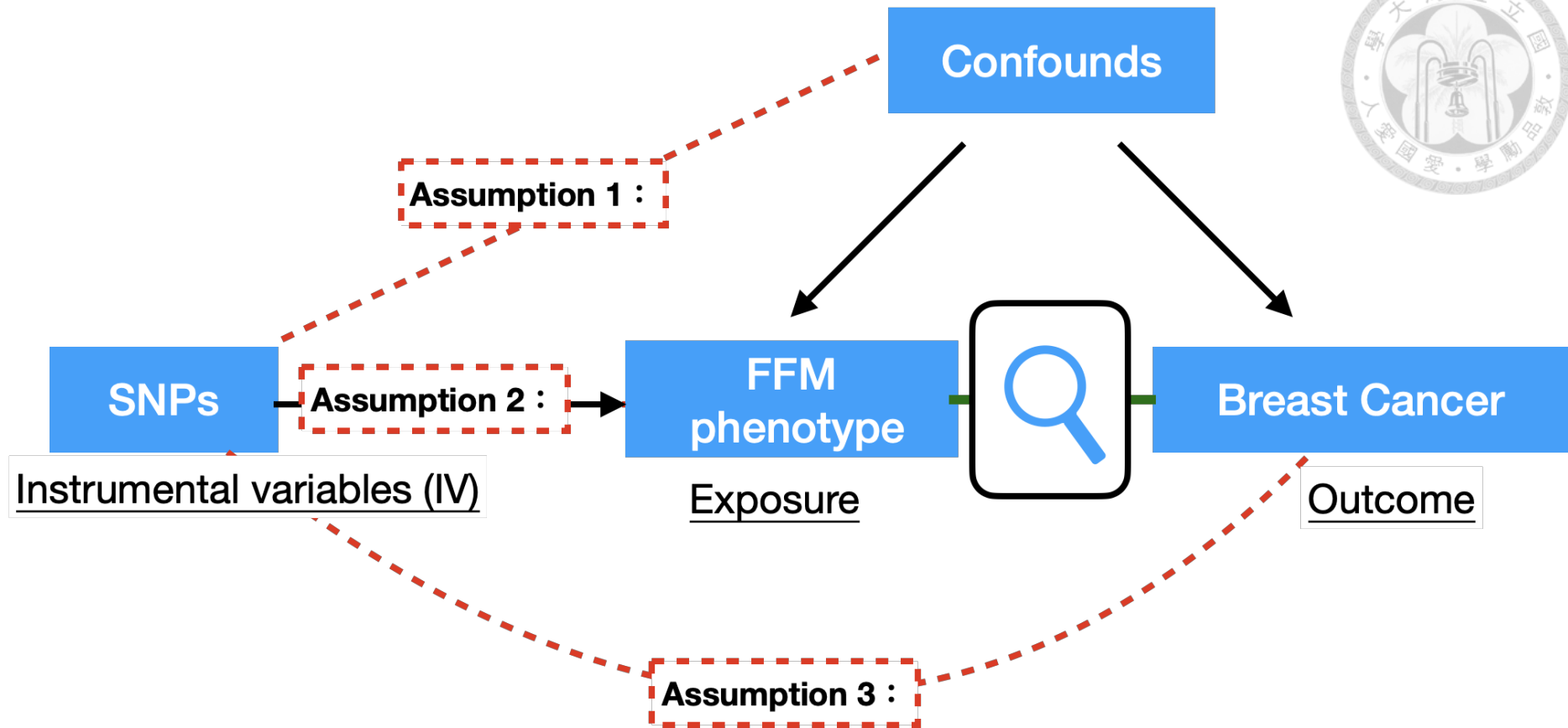
圖表資料:



圖一：研究架構



圖二：樣本篩選流程

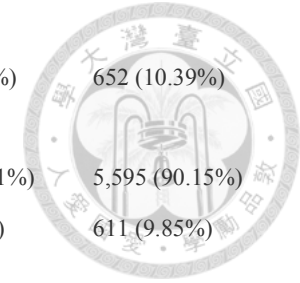


圖三：孟德爾隨機化試驗所使用之假設



表一：英國生物資料庫人口特徵-依照種族分群

Characteristic	1. British, N = 442,586 ¹	2. Irish, N = 13,207 ¹	3. Any other white background, N = 16,902 ¹	4. Asian or Asian British, N = 9,882 ¹	5. Black or Black British, N = 7,938 ¹	6. Chinese, N = 1,574 ¹	7. Mixed, N = 2,958 ¹	8. Other ethnic group, N = 7,335 ¹
sex								
men	202,332 (45.72%)	6,310 (47.78%)	6,617 (39.15%)	5,294 (53.57%)	3,365 (42.39%)	584 (37.10%)	1,105 (37.36%)	3,473 (47.35%)
women	240,254 (54.28%)	6,897 (52.22%)	10,285 (60.85%)	4,588 (46.43%)	4,573 (57.61%)	990 (62.90%)	1,853 (62.64%)	3,861 (52.65%)
age	56.86 (8.27)	56.24 (8.43)	54.55 (8.52)	53.33 (8.55)	51.94 (8.15)	52.49 (7.82)	51.73 (8.25)	53.93 (8.48)
breast cancer								
no	429,895 (97.13%)	12,847 (97.27%)	16,409 (97.08%)	9,714 (98.30%)	7,812 (98.41%)	1,542 (97.97%)	2,879 (97.33%)	7,181 (97.90%)
yes	12,691 (2.87%)	360 (2.73%)	493 (2.92%)	168 (1.70%)	126 (1.59%)	32 (2.03%)	79 (2.67%)	154 (2.10%)
Past tobacco smoking								
1 Never smoked	174,974 (42.98%)	4,330 (36.78%)	5,453 (35.73%)	6,571 (71.53%)	4,582 (63.16%)	1,022 (68.96%)	1,035 (41.24%)	2,893 (50.90%)
2 Tried once or twice	65,348 (16.05%)	1,718 (14.59%)	2,547 (16.69%)	1,009 (10.98%)	997 (13.74%)	220 (14.84%)	396 (15.78%)	795 (13.99%)
3 Occasionally	57,764 (14.19%)	1,869 (15.87%)	2,761 (18.09%)	767 (8.35%)	924 (12.74%)	126 (8.50%)	430 (17.13%)	796 (14.00%)
4 most or all days	109,038 (26.78%)	3,857 (32.76%)	4,501 (29.49%)	839 (9.13%)	752 (10.37%)	114 (7.69%)	649 (25.86%)	1,200 (21.11%)
Alcohol intake frequency								
1 Never	29,475 (6.66%)	1,113 (8.43%)	1,414 (8.37%)	4,463 (45.44%)	1,802 (22.80%)	428 (27.24%)	341 (11.55%)	1,590 (25.33%)
2 Special occasions only	47,507 (10.74%)	1,374 (10.41%)	2,495 (14.78%)	1,807 (18.40%)	2,233 (28.25%)	563 (35.84%)	538 (18.22%)	1,452 (23.13%)
3 One to three times a month	49,449 (11.18%)	1,327 (10.06%)	2,014 (11.93%)	724 (7.37%)	1,041 (13.17%)	172 (10.95%)	433 (14.66%)	677 (10.78%)
4 Once or twice a week	116,950 (26.44%)	3,360 (25.46%)	3,978 (23.56%)	1,344 (13.68%)	1,568 (19.84%)	207 (13.18%)	705 (23.87%)	1,157 (18.43%)
5 Three or four times a week	105,806 (23.92%)	3,326 (25.20%)	3,372 (19.97%)	818 (8.33%)	759 (9.60%)	96 (6.11%)	494 (16.73%)	750 (11.95%)



6 Daily or almost daily	93,085 (21.05%)	2,696 (20.43%)	3,612 (21.39%)	665 (6.77%)	501 (6.34%)	105 (6.68%)	442 (14.97%)	652 (10.39%)
diabetes								
no	420,034 (95.12%)	12,582 (95.51%)	16,031 (95.29%)	7,979 (82.50%)	6,949 (88.56%)	1,442 (94.00%)	2,765 (94.11%)	5,595 (90.15%)
yes	21,538 (4.88%)	592 (4.49%)	793 (4.71%)	1,692 (17.50%)	898 (11.44%)	92 (6.00%)	173 (5.89%)	611 (9.85%)
birth history								
no	43,989 (18.32%)	1,637 (23.76%)	2,719 (26.49%)	633 (13.92%)	745 (16.42%)	219 (22.23%)	449 (24.30%)	694 (20.87%)
yes	196,086 (81.68%)	5,252 (76.24%)	7,546 (73.51%)	3,913 (86.08%)	3,793 (83.58%)	766 (77.77%)	1,399 (75.70%)	2,632 (79.13%)
age of menarche	12.96 (1.60)	13.12 (1.63)	12.94 (1.61)	13.25 (1.82)	13.30 (1.95)	13.10 (1.68)	12.87 (1.72)	13.03 (1.79)
age of menopause	49.75 (5.13)	49.44 (5.02)	50.00 (4.82)	48.59 (5.06)	48.30 (5.52)	50.00 (4.49)	48.80 (5.44)	49.17 (5.10)
menopause status								
no	6,901 (4.81%)	210 (4.99%)	304 (5.48%)	72 (3.04%)	64 (3.45%)	39 (8.25%)	60 (7.67%)	65 (3.85%)
yes	136,659 (95.19%)	3,998 (95.01%)	5,243 (94.52%)	2,298 (96.96%)	1,792 (96.55%)	434 (91.75%)	722 (92.33%)	1,623 (96.15%)

¹ n (%); Mean (SD)



表二：英國生物資料庫-人口學特徵資料展示

Characteristic	Total	Case	Control	P-value ²
Sample size	N = 48,942 ¹	N = 9,554 ¹	N = 39,388 ¹	
Sex				
women	48,942 (100%)	9,554 (100%)	39,388 (100%)	
Age	56.64 (8.09)	57.48 (8.67)	56.43 (7.93)	<0.001
Smoke				<0.001
1 Never smoked	23,101 (47.20%)	4,320 (45.22%)	18,781 (47.68%)	
2 Tried once or twice	7,870 (16.08%)	1,469 (15.38%)	6,401 (16.25%)	
3 Occasionally	6,972 (14.25%)	1,397 (14.62%)	5,575 (14.15%)	
4 most or all days	10,999 (22.47%)	2,368 (24.79%)	8,631 (21.91%)	
Alcohol				0.036
1 Never	3,769 (7.70%)	783 (8.20%)	2,986 (7.58%)	
2 Special occasions only	6,601 (13.49%)	1,324 (13.86%)	5,277 (13.40%)	
3 One to three times a month	6,435 (13.15%)	1,208 (12.64%)	5,227 (13.27%)	
4 Once or twice a week	13,080 (26.73%)	2,524 (26.42%)	10,556 (26.80%)	
5 Three or four times a week	10,908 (22.29%)	2,071 (21.68%)	8,837 (22.44%)	
6 Daily or almost daily	8,149 (16.65%)	1,644 (17.21%)	6,505 (16.52%)	
Diabetes				<0.001
no	47275 (96.59%)	9151 (95.78%)	38124 (96.79%)	

yes	1,667 (3.41%)	403 (4.22%)	1,264 (3.21%)
Age of menarche	12.93 (1.60)	12.88 (1.60)	12.95 (1.60)
Number of live births	1.81 (1.14)	1.78 (1.14)	1.82 (1.14)
Age of menopause	49.84 (5.01)	49.74 (5.04)	49.86 (5.00)



¹ n (%); Mean (SD)

² Two Sample t-test; Pearson's Chi-squared test



表三：英國生物資料庫-肥胖指標資料展示

Characteristic	Total	Case	Control	P-value ²	Univariate logistic regression		
					OR ³	95% CI ³	p-value
Sample size	N = 48,942 ¹	N = 9,554 ¹	N = 39,388 ¹				
BFP	36.76 (6.76)	37.33 (6.56)	36.62 (6.80)	<0.001	1.02	1.01, 1.02	<0.001
Weight	71.65 (13.71)	72.46 (13.33)	71.45 (13.80)	<0.001	1.01	1.00, 1.01	<0.001
FFM	44.55 (4.89)	44.70 (4.81)	44.51 (4.91)	<0.001	1.01	1.00, 1.01	<0.001
BFM	27.10 (9.87)	27.76 (9.58)	26.94 (9.93)	<0.001	1.01	1.01, 1.01	<0.001
WHR	0.82 (0.07)	0.82 (0.07)	0.82 (0.07)	<0.001	4.81	3.49, 6.64	<0.001

¹ Mean (SD)

² Two Sample t-test

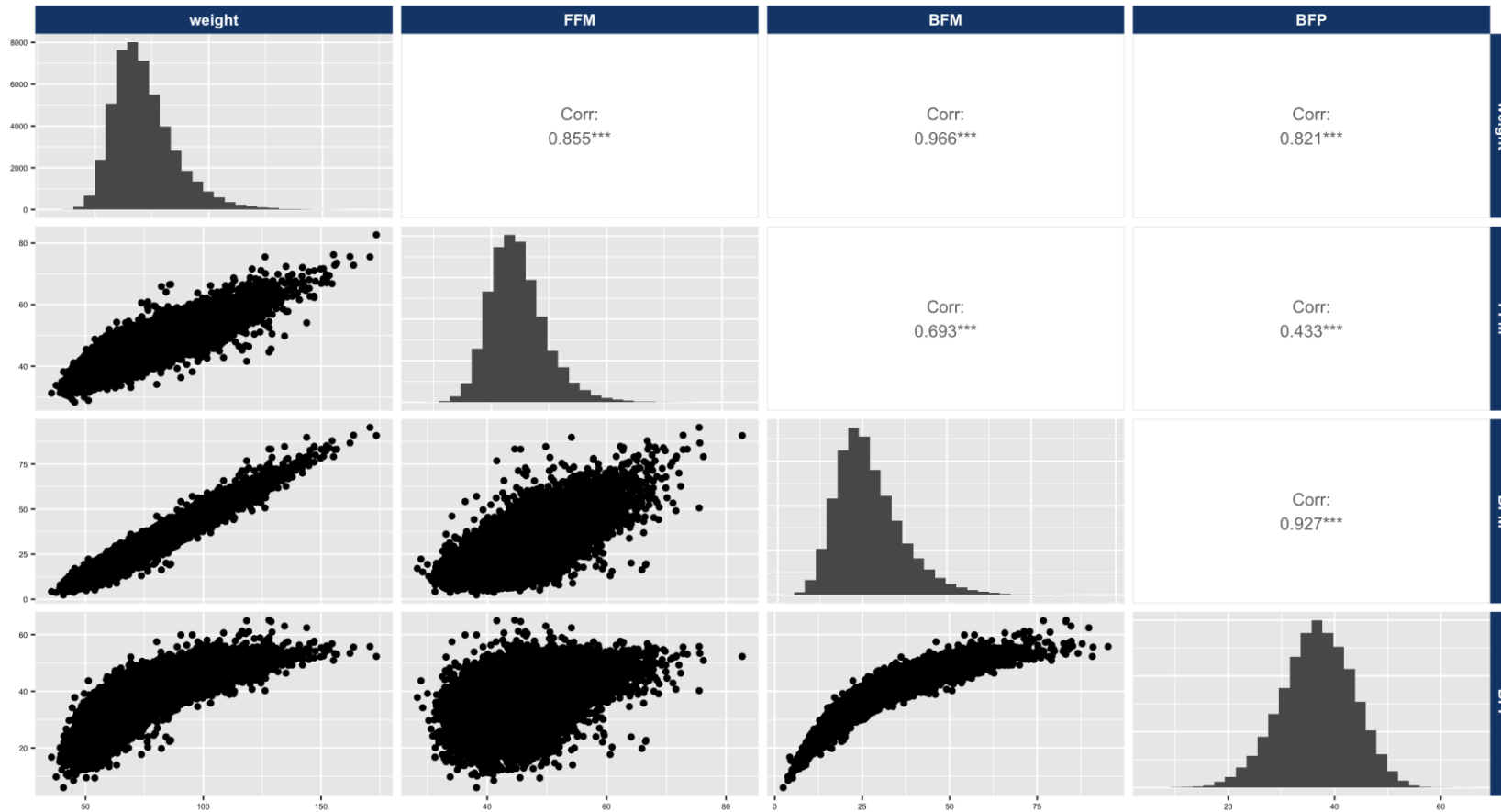
³ OR = Odds Ratio, CI = Confidence Interval

⁴ BFP (%)：體脂率；FFM (Kg)：去脂肪體重；BFM (Kg)：身體脂肪重量；WHR：腰臀比





圖四：英國生物資料庫-相關肥胖指標展示



BFP (%)：體脂率；FFM (Kg)：去脂肪體重；BFM (Kg)：身體脂肪重量；WHR：腰臀比



表四：去脂肪體重-乳癌風險的主要孟德爾隨機化試驗結果

(A) 沒有經過插補

outcome	exposure	method	SNPs	beta	standard error	p value
breast cancer	Free fat mass	2SLS	97	-0.015	0.013	0.258
breast cancer	Free fat mass	Inverse variance weighted	97	-0.031	0.014	0.027
breast cancer	Free fat mass	MR Egger	97	-0.087	0.069	0.209
breast cancer	Free fat mass	MR Egger Intercept	97	0.011	0.013	0.404
breast cancer	Free fat mass	Weighted median	97	-0.039	0.016	0.017

(B) 經過插補

outcome	exposure	method	SNPs	beta	standard error	p value
breast cancer	Free fat mass	2SLS	164	-0.012	0.011	0.283
breast cancer	Free fat mass	Inverse variance weighted	164	-0.024	0.009	0.012
breast cancer	Free fat mass	MR Egger	164	-0.050	0.042	0.236
breast cancer	Free fat mass	MR Egger Intercept	164	0.005	0.008	0.518
breast cancer	Free fat mass	Weighted median	164	-0.033	0.013	0.009



表五：身體脂肪體重-乳癌風險的主要孟德爾隨機化試驗結果

(A) 沒有經過插補

outcome	exposure	method	SNPs	beta	standard error	p value
breast cancer	Body fat mass	2SLS	51	-0.012	0.009	0.171
breast cancer	Body fat mass	Inverse variance weighted	51	-0.011	0.008	0.194
breast cancer	Body fat mass	MR Egger	51	-0.008	0.037	0.828
breast cancer	Body fat mass	MR Egger Intercept	51	-0.001	0.014	0.945
breast cancer	Body fat mass	Weighted median	51	-0.013	0.011	0.224

(B) 經過插補

outcome	exposure	method	SNPs	beta	standard error	p value
breast cancer	Body fat mass	2SLS	98	-0.010	0.007	0.134
breast cancer	Body fat mass	Inverse variance weighted	98	-0.004	0.007	0.548
breast cancer	Body fat mass	MR Egger	98	-0.010	0.027	0.704
breast cancer	Body fat mass	MR Egger Intercept	98	0.002	0.010	0.812
breast cancer	Body fat mass	Weighted median	98	-0.002	0.008	0.758



表六：敏感性分析-陰性對照：罹患乳癌是否會影響去脂肪體重的敏感性孟德爾隨機化試驗結果 ($\hat{\beta}_{2SLS} = 0.352$, $OR_{2SLS} = 1.422$,

$P_{2SLS} = 0.296$; $\hat{\beta}_{IVW} = 0.062$, $OR_{IVW} = 1.064$, $P_{IVW} = 0.175$)

outcome	exposure	method	SNPs	beta	standard error	p value
breast cancer	Free fat mass	2SLS	74	0.352	0.337	0.296
breast cancer	Free fat mass	Inverse variance weighted	74	0.062	0.046	0.175
breast cancer	Free fat mass	MR Egger	74	0.143	0.133	0.286
breast cancer	Free fat mass	MR Egger Intercept	74	-0.010	0.015	0.519
breast cancer	Free fat mass	Weighted median	74	0.126	0.064	0.051



表七：敏感性分析-孟德爾隨機化試驗 群集分析(是否停經)：停經後確診 ($\hat{\beta}_{2SLS} = -0.043$, $OR_{2SLS} = 0.958$, $P_{2SLS} = 0.014$; $\hat{\beta}_{IVW} = -$

0.083 , $OR_{IVW} = 0.920$, $P_{IVW} = 0.013$), 停經前確診 ($\hat{\beta}_{2SLS} = -0.028$, $OR_{2SLS} = 0.972$, $P_{2SLS} = 0.539$; $\hat{\beta}_{IVW} = 0.002$,

$OR_{IVW} = 1.002$, $P_{IVW} = 0.914$)

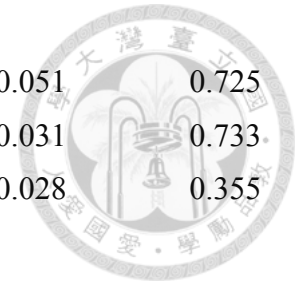
(A) : Menopause (N=6, 060)

outcome	exposure	method	SNPs	beta	standard error	p value
breast cancer	Free fat mass	2SLS	50	-0.043	0.018	0.014
breast cancer	Free fat mass	Inverse variance weighted	50	-0.083	0.033	0.013
breast cancer	Free fat mass	MR Egger	50	-0.428	0.171	0.016
breast cancer	Free fat mass	MR Egger Intercept	50	0.091	0.044	0.045
breast cancer	Free fat mass	Weighted median	50	-0.079	0.044	0.074

(B) : No Menopause (N=1, 028)

outcome	exposure	method	SNPs	beta	standard error	p value
breast cancer	Free fat mass	2SLS	23	-0.028	0.045	0.539
breast cancer	Free fat mass	Inverse variance weighted	23	0.002	0.022	0.914

breast cancer	Free fat mass	MR Egger	23	0.018	0.051	0.725
breast cancer	Free fat mass	MR Egger Intercept	23	-0.011	0.031	0.733
breast cancer	Free fat mass	Weighted median	23	0.026	0.028	0.355





表八：敏感性分析-孟德爾隨機化試驗 群集分析(組織亞型)：乳腺癌 ($\hat{\beta}_{IVW} = -0.079$, OR=0.924, P-value=0.006), 乳小葉癌

($\hat{\beta}_{IVW} = -0.011$, OR=0.989, P-value=0.826)

(A) : Ductal carcinoma (N=6,648)

outcome	exposure	method	SNPs	beta	standard error	p value
breast cancer	Free fat mass	2SLS	95	-0.021	0.014	0.134
breast cancer	Free fat mass	Inverse variance weighted	95	-0.030	0.011	0.005
breast cancer	Free fat mass	MR Egger	95	-0.017	0.045	0.705
breast cancer	Free fat mass	MR Egger Intercept	95	-0.003	0.010	0.779
breast cancer	Free fat mass	Weighted median	95	-0.041	0.016	0.010

(B) : Lobular carcinoma (N=1,094)

outcome	exposure	method	SNPs	beta	standard error	p value
breast cancer	Free fat mass	2SLS	26	-0.135	0.055	0.015
breast cancer	Free fat mass	Inverse variance weighted	26	-0.035	0.032	0.265
breast cancer	Free fat mass	MR Egger	26	0.011	0.073	0.878
breast cancer	Free fat mass	MR Egger Intercept	26	-0.035	0.049	0.482
breast cancer	Free fat mass	Weighted median	26	-0.014	0.035	0.700



表九：將去脂肪體重所使用 164 個變異位點，進行基因註釋。

Chr	Pos	Ref	Alt	SNPs ID	Gene name	Gene description
1	177905174	T	C	rs2902210	RP4-798P15.3	Null
1	177922617	A	G	rs1336784	RP4-798P15.3	Null
1	177931122	G	A	rs10798585	RP4-798P15.3	Null
1	177967005	A	C	rs35308250	RP4-798P15.3	Null
1	177992402	T	C	rs11585914	RP4-798P15.3	Null
2	58935282	T	G	rs4671328	LINC01122	long intergenic non-protein coding RNA 1122 [Source:HGNC Symbol;Acc:49267]
2	58955953	A	G	rs2708146	LINC01122	long intergenic non-protein coding RNA 1122 [Source:HGNC Symbol;Acc:49267]
2	58992430	T	C	rs7597839	LINC01122	long intergenic non-protein coding RNA 1122 [Source:HGNC Symbol;Acc:49267]
2	59122090	T	C	rs42825	LINC01122	long intergenic non-protein coding RNA 1122 [Source:HGNC Symbol;Acc:49267]
2	59125540	A	G	rs1861101	LINC01122	long intergenic non-protein coding RNA 1122 [Source:HGNC Symbol;Acc:49267]
2	172401050	C	T	rs968379	CYBRD1	cytochrome b reductase 1 [Source:HGNC Symbol;Acc:20797]
2	172405989	A	G	rs11684782	CYBRD1	cytochrome b reductase 1 [Source:HGNC Symbol;Acc:20797]
2	211540507	C	A	rs1047891	CPS1	carbamoyl-phosphate synthase 1, mitochondrial [Source:HGNC Symbol;Acc:2323]
2	233816504	C	T	rs72978115	NGEF	neuronal guanine nucleotide exchange factor [Source:HGNC Symbol;Acc:7807]
2	233830837	C	A	rs11678345	NGEF	neuronal guanine nucleotide exchange factor [Source:HGNC Symbol;Acc:7807]
2	233830837	C	T	rs11678345	NGEF	neuronal guanine nucleotide exchange factor [Source:HGNC Symbol;Acc:7807]
2	233839239	C	T	rs6748986	NGEF	neuronal guanine nucleotide exchange factor [Source:HGNC Symbol;Acc:7807]
3	141076084	T	C	rs9866391	ZBTB38	zinc finger and BTB domain containing 38 [Source:HGNC Symbol;Acc:26636]
3	141154542	T	C	rs6440008	ZBTB38	zinc finger and BTB domain containing 38 [Source:HGNC Symbol;Acc:26636]
3	141234523	T	C	rs13069717	RASA2	RAS p21 protein activator 2 [Source:HGNC Symbol;Acc:9872]
3	160838885	C	T	rs142119953	NMD3	NMD3 ribosome export adaptor [Source:HGNC Symbol;Acc:24250]
3	185302885	G	A	rs10937226	SEN2	SUMO1/sentrin/SMT3 specific peptidase 2 [Source:HGNC Symbol;Acc:23116]
3	185302885	G	C	rs10937226	SEN2	SUMO1/sentrin/SMT3 specific peptidase 2 [Source:HGNC Symbol;Acc:23116]



3	185371172	A	G	rs57158761	IGF2BP2	insulin-like growth factor 2 mRNA binding protein 2 [Source:HGNC Symbol;Acc:28867]
4	17720818	G	A	rs76007570	FAM184B	family with sequence similarity 184, member B [Source:HGNC Symbol;Acc:29235]
4	17957688	T	A	rs7673321	LCORL	ligand dependent nuclear receptor corepressor-like [Source:HGNC Symbol;Acc:30776]
4	17965927	A	C	rs79320874	LCORL	ligand dependent nuclear receptor corepressor-like [Source:HGNC Symbol;Acc:30776]
5	168265891	T	C	rs1979923	SLIT3	slit homolog 3 (Drosophila) [Source:HGNC Symbol;Acc:11087]
5	179929204	C	G	rs17697839	CNOT6	CCR4-NOT transcription complex, subunit 6 [Source:HGNC Symbol;Acc:14099]
6	31239296	G	C	rs9264666	HLA-C	major histocompatibility complex, class I, C [Source:HGNC Symbol;Acc:4933]
6	32338283	G	T	rs3129942	C6orf10	chromosome 6 open reading frame 10 [Source:HGNC Symbol;Acc:13922]
6	32368558	C	T	rs9268485	BTNL2	butyrophilin-like 2 (MHC class II associated) [Source:HGNC Symbol;Acc:1142]
6	32431705	C	T	rs9268885	HLA-DRB9	major histocompatibility complex, class II, DR beta 9 (pseudogene) [Source:HGNC Symbol;Acc:4957]
6	34444536	T	A	rs9366858	PACSIN1	protein kinase C and casein kinase substrate in neurons 1 [Source:HGNC Symbol;Acc:8570]
6	34444536	T	C	rs9366858	PACSIN1	protein kinase C and casein kinase substrate in neurons 1 [Source:HGNC Symbol;Acc:8570]
6	130345835	G	A	rs1415701	L3MBTL3	l(3)mbt-like 3 (Drosophila) [Source:HGNC Symbol;Acc:23035]
6	130360617	T	C	rs34614731	L3MBTL3	l(3)mbt-like 3 (Drosophila) [Source:HGNC Symbol;Acc:23035]
6	130378833	T	A	rs12661188	L3MBTL3	l(3)mbt-like 3 (Drosophila) [Source:HGNC Symbol;Acc:23035]
6	130378833	T	C	rs12661188	L3MBTL3	l(3)mbt-like 3 (Drosophila) [Source:HGNC Symbol;Acc:23035]
6	130449280	T	G	rs12194709	L3MBTL3	l(3)mbt-like 3 (Drosophila) [Source:HGNC Symbol;Acc:23035]
6	132654831	C	T	rs11968833	MOXD1	monooxygenase, DBH-like 1 [Source:HGNC Symbol;Acc:21063]
6	161047144	G	A	rs9365187	LPA	lipoprotein, Lp(a) [Source:HGNC Symbol;Acc:6667]
6	161096446	T	A	rs1590184	RP1-81D8.4	Null
6	161096446	T	C	rs1590184	RP1-81D8.4	Null
7	2772431	A	C	rs798528	AMZ1	archaelysin family metallopeptidase 1 [Source:HGNC Symbol;Acc:22231]
7	80812398	T	C	rs7797400	AC005008.2	Uncharacterized protein [Source:UniProtKB/TrEMBL;Acc:C9IYQ3]
7	90084742	A	G	rs6465265	CTB-13L3.1	Null
7	150690079	C	T	rs2070744	NOS3	nitric oxide synthase 3 (endothelial cell) [Source:HGNC Symbol;Acc:7876]
8	97782301	A	T	rs35616182	CPQ	carboxypeptidase Q [Source:HGNC Symbol;Acc:16910]
8	135558301	T	C	rs1530318	ZFAT	zinc finger and AT hook domain containing [Source:HGNC Symbol;Acc:19899]

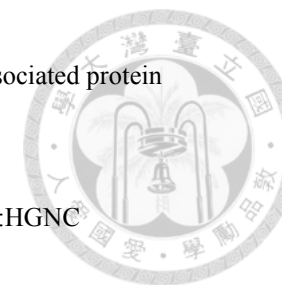


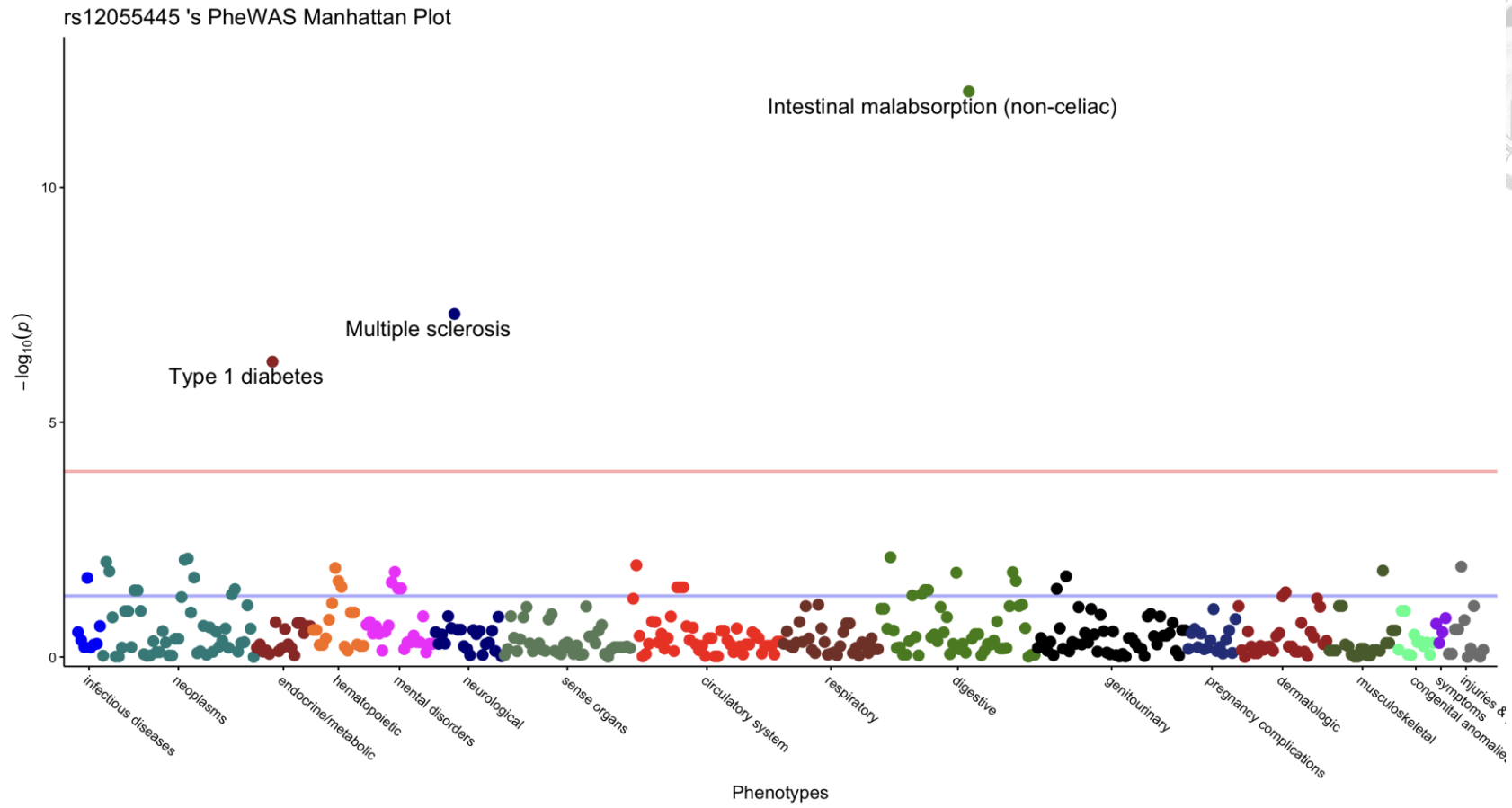
8	135581352	T	C	rs11166598	ZFAT	zinc finger and AT hook domain containing [Source:HGNC Symbol;Acc:19899]
8	135660469	A	T	rs7012213	ZFAT	zinc finger and AT hook domain containing [Source:HGNC Symbol;Acc:19899]
11	497698	G	C	rs151017225	RNHI	ribonuclease/angiogenin inhibitor 1 [Source:HGNC Symbol;Acc:10074]
11	27695464	A	G	rs11030108	BDNF-AS	BDNF antisense RNA [Source:HGNC Symbol;Acc:20608]
11	123515158	A	G	rs1274215	SCN3B	sodium channel, voltage-gated, type III, beta subunit [Source:HGNC Symbol;Acc:20665]
12	57964445	G	C	rs12821774	KIF5A	kinesin family member 5A [Source:HGNC Symbol;Acc:6323]
12	66346714	T	C	rs12371967	HMGA2	high mobility group AT-hook 2 [Source:HGNC Symbol;Acc:5009]
12	66358347	C	T	rs1042725	HMGA2	high mobility group AT-hook 2 [Source:HGNC Symbol;Acc:5009]
12	83298633	A	G	rs17726837	TMTC2	transmembrane and tetratricopeptide repeat containing 2 [Source:HGNC Symbol;Acc:25440]
12	122602305	G	A	rs28642975	MLXIP	MLX interacting protein [Source:HGNC Symbol;Acc:17055]
13	42749192	G	T	rs9590675	DGKH	diacylglycerol kinase, eta [Source:HGNC Symbol;Acc:2854]
13	42797518	C	A	rs17599277	DGKH	diacylglycerol kinase, eta [Source:HGNC Symbol;Acc:2854]
13	42797518	C	T	rs17599277	DGKH	diacylglycerol kinase, eta [Source:HGNC Symbol;Acc:2854]
13	50766240	G	A	rs1359613	DLEU1	deleted in lymphocytic leukemia 1 (non-protein coding) [Source:HGNC Symbol;Acc:13747]
13	50819769	A	G	rs2066665	DLEU1	deleted in lymphocytic leukemia 1 (non-protein coding) [Source:HGNC Symbol;Acc:13747]
13	51100097	G	A	rs3118904	RP11-175B12.2	Null
13	51105334	G	A	rs3118905	DLEU1	deleted in lymphocytic leukemia 1 (non-protein coding) [Source:HGNC Symbol;Acc:13747]
13	51129412	C	T	rs1149827	DLEU1	deleted in lymphocytic leukemia 1 (non-protein coding) [Source:HGNC Symbol;Acc:13747]
13	51131109	G	A	rs1149825	DLEU1	deleted in lymphocytic leukemia 1 (non-protein coding) [Source:HGNC Symbol;Acc:13747]
13	51176678	C	G	rs1341635	DLEU1	deleted in lymphocytic leukemia 1 (non-protein coding) [Source:HGNC Symbol;Acc:13747]
13	51176678	C	T	rs1341635	DLEU1	deleted in lymphocytic leukemia 1 (non-protein coding) [Source:HGNC Symbol;Acc:13747]
13	51254196	A	G	rs1272436	DLEU1	deleted in lymphocytic leukemia 1 (non-protein coding) [Source:HGNC Symbol;Acc:13747]
13	51254196	A	T	rs1272436	DLEU1	deleted in lymphocytic leukemia 1 (non-protein coding) [Source:HGNC Symbol;Acc:13747]
13	76365143	T	A	rs9544044	RP11-29G8.3	Null
13	76365143	T	C	rs9544044	RP11-29G8.3	Null
15	86276000	T	A	rs12907384	AKAP13	A kinase (PRKA) anchor protein 13 [Source:HGNC Symbol;Acc:371]
15	86276000	T	C	rs12907384	AKAP13	A kinase (PRKA) anchor protein 13 [Source:HGNC Symbol;Acc:371]



16	2160973	A	G	rs116092985	PKD1	polycystic kidney disease 1 (autosomal dominant) [Source:HGNC Symbol;Acc:9008]
16	2262113	G	C	rs1126	RP11-304L19.8	Null
16	2262113	G	T	rs1126	RP11-304L19.8	Null
16	19762935	A	G	rs16951276	IQCK	IQ motif containing K [Source:HGNC Symbol;Acc:28556]
16	29958216	G	A	rs8054556	TMEM219	transmembrane protein 219 [Source:HGNC Symbol;Acc:25201]
16	29958216	G	C	rs8054556	TMEM219	transmembrane protein 219 [Source:HGNC Symbol;Acc:25201]
16	29958216	G	T	rs8054556	TMEM219	transmembrane protein 219 [Source:HGNC Symbol;Acc:25201]
16	53804965	T	A	rs10852521	FTO	fat mass and obesity associated [Source:HGNC Symbol;Acc:24678]
16	53804965	T	C	rs10852521	FTO	fat mass and obesity associated [Source:HGNC Symbol;Acc:24678]
16	53818708	G	T	rs3751813	FTO	fat mass and obesity associated [Source:HGNC Symbol;Acc:24678]
17	7463102	G	A	rs143094271	TNFSF12	tumor necrosis factor (ligand) superfamily, member 12 [Source:HGNC Symbol;Acc:11927]
17	7569542	G	A	rs74351250	TP53	tumor protein p53 [Source:HGNC Symbol;Acc:11998]
17	29211667	G	A	rs7223535	ATAD5	ATPase family, AAA domain containing 5 [Source:HGNC Symbol;Acc:25752]
17	46958166	T	C	rs8075255	SUMO2P17	SUMO2 pseudogene 17 [Source:HGNC Symbol;Acc:49358]
17	62017421	A	G	rs2532111	SCN4A	sodium channel, voltage-gated, type IV, alpha subunit [Source:HGNC Symbol;Acc:10591]
17	62019103	G	A	rs56342400	SCN4A	sodium channel, voltage-gated, type IV, alpha subunit [Source:HGNC Symbol;Acc:10591]
17	62019103	G	C	rs56342400	SCN4A	sodium channel, voltage-gated, type IV, alpha subunit [Source:HGNC Symbol;Acc:10591]
17	62019103	G	T	rs56342400	SCN4A	sodium channel, voltage-gated, type IV, alpha subunit [Source:HGNC Symbol;Acc:10591]
18	57135158	C	G	rs9959290	CCBE1	collagen and calcium binding EGF domains 1 [Source:HGNC Symbol;Acc:29426]
18	57804346	A	T	rs8095404	RP11-795H16.3	Null
19	4366219	A	G	rs732716	SH3GL1	SH3-domain GRB2-like 1 [Source:HGNC Symbol;Acc:10830]
19	10770305	C	G	rs7248205	ILF3	interleukin enhancer binding factor 3, 90kDa [Source:HGNC Symbol;Acc:6038]
19	10770305	C	T	rs7248205	ILF3	interleukin enhancer binding factor 3, 90kDa [Source:HGNC Symbol;Acc:6038]
20	31939710	C	T	rs293710	BPIFB9P	BPI fold containing family B, member 9, pseudogene [Source:HGNC Symbol;Acc:16109]
20	32358906	C	G	rs6059462	ZNF341	zinc finger protein 341 [Source:HGNC Symbol;Acc:15992]
20	32358906	C	T	rs6059462	ZNF341	zinc finger protein 341 [Source:HGNC Symbol;Acc:15992]
20	33496471	C	T	rs4911448	ACSS2	acyl-CoA synthetase short-chain family member 2 [Source:HGNC Symbol;Acc:15814]

20	33642396	C	T	rs62213709	TRPC4AP	transient receptor potential cation channel, subfamily C, member 4 associated protein [Source:HGNC Symbol;Acc:16181]
20	33844938	A	C	rs2425024	MMP24-AS1	MMP24 antisense RNA 1 [Source:HGNC Symbol;Acc:44421]
20	33867697	A	G	rs619865	RP4-614O4.11	Null
20	33912371	T	C	rs2425059	UQCC1	ubiquinol-cytochrome c reductase complex assembly factor 1 [Source:HGNC Symbol;Acc:15891]
20	34030425	C	T	rs34414056	GDF5	growth differentiation factor 5 [Source:HGNC Symbol;Acc:4220]
20	34186953	T	A	rs6120997	FER1L4	fer-1-like 4 (CNull elegans), pseudogene [Source:HGNC Symbol;Acc:15801]
20	34186953	T	C	rs6120997	FER1L4	fer-1-like 4 (CNull elegans), pseudogene [Source:HGNC Symbol;Acc:15801]
20	34190870	A	G	rs3746410	FER1L4	fer-1-like 4 (CNull elegans), pseudogene [Source:HGNC Symbol;Acc:15801]
20	34190970	C	T	rs41293082	FER1L4	fer-1-like 4 (CNull elegans), pseudogene [Source:HGNC Symbol;Acc:15801]
20	34460479	C	T	rs6058339	PHF20	PHD finger protein 20 [Source:HGNC Symbol;Acc:16098]
20	35002849	A	G	rs2425224	DLGAP4	discs, large (Drosophila) homolog-associated protein 4 [Source:HGNC Symbol;Acc:24476]
20	50788017	C	A	rs6091465	ZFP64	ZFP64 zinc finger protein [Source:HGNC Symbol;Acc:15940]
20	50846917	C	T	rs6021797	RP4-723E3.1	Null
22	45611846	C	T	rs13058706	KIAA0930	KIAA0930 [Source:HGNC Symbol;Acc:1314]
22	50952570	A	C	rs131820	NCAPH2	non-SMC condensin II complex, subunit H2 [Source:HGNC Symbol;Acc:25071]
22	50952570	A	G	rs131820	NCAPH2	non-SMC condensin II complex, subunit H2 [Source:HGNC Symbol;Acc:25071]





圖五：單一變異位點 (rs12055445) 的全表型組關聯分析結果

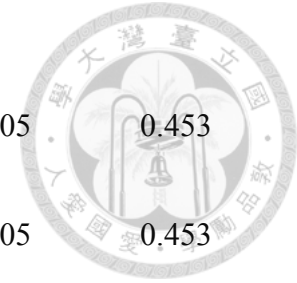


表十：使用 164 個去脂肪體重工具變數變異位點，依次進行全表型組關聯分析，並彙整出顯著的變異位點與表型結果。

SNP's name	phenotype	phenotype group	beta	P value	Allele freq
rs9946862	Cardiac shunt/ heart septal defect	congenital anomalies	1.068	4E-05	0.05
rs9275602	Intestinal malabsorption (non-celiac)	digestive	0.91	1E-22	0.202
rs9271578	Multiple sclerosis	neurological	-0.42	7E-05	0.494
rs9271578	Intestinal malabsorption (non-celiac)	digestive	-0.57	1E-09	0.492
rs9268885	Intestinal malabsorption (non-celiac)	digestive	0.93	9E-25	0.253
rs9268485	Intestinal malabsorption (non-celiac)	digestive	0.509	3E-08	0.487
rs9264666	Intestinal malabsorption (non-celiac)	digestive	0.472	1E-07	0.444
rs6433294	Other inflammatory spondylopathies	musculoskeletal	-0.43	9E-05	0.356
rs62482605	Inflammation of eyelids	sense organs	1.323	6E-06	0.029
rs62482605	Genitourinary congenital anomalies	congenital anomalies	1.345	2E-05	0.028
rs62482605	Congenital anomalies of urinary system	congenital anomalies	1.303	9E-05	0.028
rs4318925	Intestinal malabsorption (non-celiac)	digestive	0.998	2E-26	0.179

rs377763	Type 1 diabetes	endocrine/metabolic	0.494	6E-07	0.218
rs377763	Intestinal malabsorption (non-celiac)	digestive	1.128	1E-35	0.219
rs3129942	Intestinal malabsorption (non-celiac)	digestive	0.948	9E-26	0.263
rs3021302	Type 1 diabetes	endocrine/metabolic	0.494	2E-06	0.172
rs3021302	Intestinal malabsorption (non-celiac)	digestive	1.319	2E-47	0.173
rs2858332	Multiple sclerosis	neurological	-0.52	2E-06	0.474
rs2858332	Intestinal malabsorption (non-celiac)	digestive	-0.58	1E-09	0.473
rs1530318	Other nonmalignant breast conditions	genitourinary	-0.39	7E-05	0.174
rs142119953	Nervous system congenital anomalies	congenital anomalies	1.868	2E-05	0.024
rs1383266	Intestinal malabsorption (non-celiac)	digestive	-0.47	8E-05	0.243
rs12055445	Type 1 diabetes	endocrine/metabolic	0.469	5E-07	0.494
rs12055445	Multiple sclerosis	neurological	0.587	5E-08	0.493
rs12055445	Intestinal malabsorption (non-celiac)	digestive	0.672	9E-13	0.495
rs11684782	Intestinal infection	infectious diseases	0.192	1E-04	0.493
rs11684782	Other benign neoplasm of uterus	neoplasms	0.773	6E-05	0.493





`8:78165084_TTA_T`	Disorders of fluid, electrolyte, and acid-base balance	endocrine/metabolic	0.341	4E-05	0.453
`8:78165084_TTA_T`	Hypovolemia	endocrine/metabolic	0.341	4E-05	0.453
`6:32890908_AAAAC_A`	Intestinal malabsorption (non-celiac)	digestive	0.507	3E-08	0.287
`6:31181419_CAAAA_AAAAAAA_C`	Intestinal malabsorption (non-celiac)	digestive	0.348	8E-05	0.447



表十一：針對表十中的表型，統整出現次數

Phenotype	Phenotype (中文翻譯)	Significant times	Phenotype group
Intestinal malabsorption (non-celiac)	腸道吸收不良 (非腹腔)	14	digestive
Multiple sclerosis	多發性硬化症	3	neurological
Type 1 diabetes	第一型糖尿病	3	endocrine/metabolic
Cardiac shunt/ heart septal defect	心臟分流/心間隔缺損	1	congenital anomalies
Congenital anomalies of urinary system	先天性泌尿系統異常	1	congenital anomalies
Disorders of fluid, electrolyte, and acid-base balance	體液、電解質和酸鹼平衡紊亂	1	endocrine/metabolic
Genitourinary congenital anomalies	泌尿生殖系統先天性異常	1	congenital anomalies
Hypovolemia	血容量過低	1	endocrine/metabolic
Inflammation of eyelids	眼瞼發炎	1	sense organs
Intestinal infection	腸道感染	1	infectious diseases
Nervous system congenital anomalies	神經系統先天性異常	1	congenital anomalies
Other benign neoplasm of uterus	其他子宮良性腫瘤	1	neoplasms

Other inflammatory spondylopathies

其他炎症性脊柱病

1

Other nonmalignant breast conditions

其他非惡性乳腺疾病

1



musculoskeletal

genitourinary



表十二：外部驗證-乳腺癌協會聯盟 2020：使用雙樣本孟德爾隨機化試驗，驗證在歐洲人種中，去脂肪體重與乳癌風險因果關係的一

致性 ($\hat{\beta}_{IVW} = 0.002$, $OR_{IVW} = 1.002$, $P_{IVW} = 0.733$)

Case : 133,384 / Control : 113,789

outcome	exposure	method	SNPs	beta	standard error	p value
breast cancer	Free fat mass	Inverse variance weighted	141	0.002	0.007	0.733
breast cancer	Free fat mass	MR Egger	141	-0.048	0.033	0.147
breast cancer	Free fat mass	MR Egger Intercept	141	0.010	0.006	0.119
breast cancer	Free fat mass	Weighted median	141	-0.003	0.006	0.539



表十三：外部驗證-乳腺癌協會聯盟 2020：使用雙樣本孟德爾隨機化試驗，驗證在歐洲人種中，全身脂肪體重與乳癌風險因果關係的

一致性 ($\hat{\beta}_{IVW} = -0.006$, $OR_{IVW} = 0.994$, $P_{IVW} = 0.211$)

Case : 133,384 / Control : 113,789

outcome	exposure	method	SNPs	beta	standard error	p value
breast cancer	Body fat mass	Inverse variance weighted	85	-0.006	0.005	0.211
breast cancer	Body fat mass	MR Egger	85	0.001	0.020	0.940
breast cancer	Body fat mass	MR Egger Intercept	85	-0.003	0.007	0.698
breast cancer	Body fat mass	Weighted median	85	-0.005	0.004	0.227



表十四：外部驗證-臺灣生物資料庫：使用雙樣本孟德爾隨機化試驗，驗證在歐洲人與漢人間，去脂肪體重與乳癌風險的因果關係，

是否存在一致性 ($\hat{\beta}_{IVW} = -0.056$, $OR_{IVW} = 0.946$, $P_{IVW} = 0.448$)

Case: 407 / Control : 1224

outcome	exposure	method	SNPs	beta	standard error	p value
breast cancer	Fat free mass	Inverse variance weighted	93	-0.056	0.074	0.448
breast cancer	Fat free mass	MR Egger	93	0.279	0.398	0.485
breast cancer	Fat free mass	MR Egger Intercept	93	-0.062	0.073	0.393
breast cancer	Fat free mass	Weighted median	93	-0.008	0.103	0.936



表十五：外部驗證-臺灣生物資料庫：使用雙樣本孟德爾隨機化試驗，驗證在歐洲人與漢人間，全身脂肪體重與乳癌風險的因果關係，是否存在一致性 ($\hat{\beta}_{IVW} = -0.056$, $OR_{IVW} = 0.946$, $P_{IVW} = 0.448$)

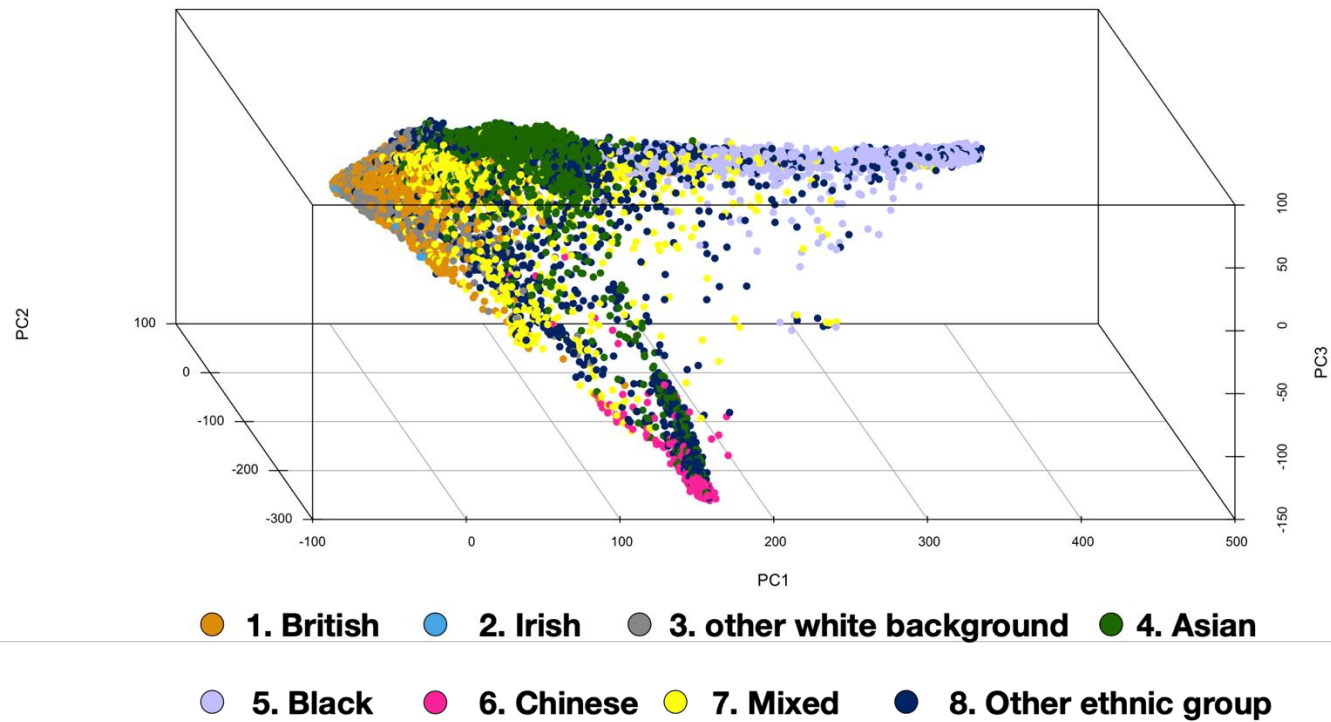
Case: 407 / Control : 1224

outcome	exposure	method	SNPs	beta	standard error	p value
breast cancer	Body fat mass	Inverse variance weighted	53	0.041	0.048	0.398
breast cancer	Body fat mass	MR Egger	53	0.284	0.210	0.182
breast cancer	Body fat mass	MR Egger Intercept	53	-0.094	0.078	0.239
breast cancer	Body fat mass	Weighted median	53	0.043	0.067	0.518



附錄補充-圖表資料:

附錄圖一：主成分分析 三維散佈圖 依照種族





附錄圖二：主成分分析 二維散佈圖（A 圖為依照各種族的散佈圖，B 圖為針對英國白人族群的進一步篩選，僅選擇主成分分析相近的作為研究對象，如圖 B 的藍色部分所示）





附錄表一：英國生物資料庫-變項與對照編碼

UKB 編碼：	變項名稱 中文	變項名稱 英文
31	性別	Sex
34	出生年份	Year of birth
48	腰圍	Waist circumference
49	臀圍	Hip circumference
1249	過去吸煙	Past tobacco smoking
1558	飲酒頻率	Alcohol intake frequency
2443	糖尿病	Diabetes diagnosed by doctor
2714	初經年齡	Age when periods started (menarche)
2734	生產數量	Number of live births
3581	停經年齡	Age at menopause (last menstrual period)
21000	種族背景	Ethnic background
21001	身體質量指數	BMI

21002	重量	Weight
21022	招聘年齡	Age at recruitment
22006	英國白人並且基於主成分分析的篩選	Genetic ethnic grouping
23099	體脂百分比	Body fat percentage
23101	全身去脂體重	Whole body fat-free mass
40001	潛在（主要）死因：ICD10	Underlying (primary) cause of death: ICD10
40008	癌症診斷年齡	Age at cancer diagnosis
40011	癌症腫瘤的組織學	Histology of cancer tumor





附錄表二：英國生物資料庫，肥胖指標重複測量之時間段與數據可用性彙整

Characteristic	身高	體重	身體質量指數	體脂率	去脂肪體重	腰圍	臀圍
Sample size	N = 502,5051	N = 502,5051	N = 502,5051	N = 502,5051	N = 502,5051	N = 502,5051	N = 502,5051
1st	499,965 (99%)	499,730 (99%)	499,399 (99%)	492,096 (98%)	492,096 (98%)	500,285 (99%)	500,344 (99%)
2nd	20,313 (4.0%)	20,303 (4.0%)	20,299 (4.0%)	20,037 (4.0%)	20,037 (4.0%)	20,317 (4.0%)	20,322 (4.0%)
3rd	47,159 (9.4%)	47,109 (9.4%)	47,055 (9.4%)	46,076 (9.2%)	46,076 (9.2%)	47,202 (9.4%)	47,200 (9.4%)
4th	3,087 (0.6%)	3,086 (0.6%)	3,080 (0.6%)	3,009 (0.6%)	3,009 (0.6%)	3,087 (0.6%)	3,087 (0.6%)

¹ n (%), FFM : 去脂肪體重

² 1st : (2006-2010); 2nd : (2012-2013); 3rd : (2014+); 4th : (2019+)



附錄表三：台灣生物資料庫-資料展示

Characteristic	Total	Case	Control	P-value ²
Sample size	N = 1,631 ¹	N = 407 ¹	N = 1,224 ¹	
sex				
women	1,631 (100%)	407 (100%)	1,224 (100%)	
age	50 (11)	53 (10)	49 (11)	<0.001
smoke				0.5
1. 沒有抽菸	1,601 (98%)	403 (99%)	1,198 (98%)	
2. 抽菸(以月作為頻率)	1 (<0.1%)	0 (0%)	1 (<0.1%)	
3. 抽菸(以週作為頻率)	5 (0.3%)	1 (0.2%)	4 (0.3%)	
4. 抽菸(以天作為頻率)	24 (1.5%)	3 (0.7%)	21 (1.7%)	
alcohol				>0.9
1. 沒有或偶爾喝酒	1,589 (97%)	398 (98%)	1,191 (97%)	
2. 已經戒酒	10 (0.6%)	2 (0.5%)	8 (0.7%)	
3. 目前仍持續喝酒	31 (1.9%)	7 (1.7%)	24 (2.0%)	
diabetes				0.8



no	1,559 (96%)	388 (95%)	1,171 (96%)
yes	72 (4.4%)	19 (4.7%)	53 (4.3%)
menarche	13.56 (1.57)	13.64 (1.51)	13.54 (1.59)
Birth history			0.3
no	271 (17%)	57 (14%)	214 (17%)
yes	1,360 (83%)	350 (86%)	1,010 (83%)

0.11

¹ n (%); Mean (SD)

² Two Sample t-test; Fisher's exact test; Pearson's Chi-squared test



附錄表四：去脂肪體重之工具變數所使用之變異位點-英國生物資料庫

rs138574676	rs4671328	rs9866391
rs4339825	rs2708146	rs6440008
rs1359940	rs7597839	rs13069717
rs543874	rs42825	rs142119953
rs2902210	rs1861101	rs10937226
rs1336784	rs6433294	rs57158761
rs10798585	rs968379	rs76007570
rs35308250	rs11684782	rs7679066
rs11585914	rs1047891	rs7673321
rs7522001	rs72978115	rs79320874
rs1327124	rs11678345	rs10866398
rs62104180	rs6748986	rs62368123
rs6759177	rs62410530	rs1378438



rs1582931	rs1383266	rs1590184
rs1979923	6:32890908_AAAAC_A	rs798528
rs17697839	rs9366858	rs7797400
rs707894	rs200627547	rs6465265
rs9264666	6:41734037_GGAAA_G	rs2070744
rs377763	rs554468	rs62482605
rs3129942	rs1415701	8:78165084_TTA_T
rs9268485	6:130349218_CTT_C	rs35616182
rs9268885	rs34614731	rs1530318
rs9271578	rs12661188	rs11166598
rs3021302	rs12194709	rs7012213
rs12055445	rs566411214	rs7908841
rs2858332	rs11968833	rs1467616
rs9275602	rs57187762	rs12217775
rs4318925	rs9365187	rs7081930

rs151017225	rs9590675	rs16951276
rs11030108	rs17599277	rs56308633
rs1274215	rs1359613	rs8054556
12:29301733_AC_A	rs2066665	rs10852521
rs11169182	rs3118904	rs3751813
rs12821774	rs3118905	rs143094271
rs7309108	rs1149827	rs74351250
rs12371967	rs1149825	rs7223535
rs1042725	rs1341635	rs8075255
rs11495610	rs1272436	rs2854218
12:83283746_GTAAC_G	rs9544044	rs2532111
rs17726837	rs28711886	rs56342400
rs1498707	rs12907384	rs8096254
rs4373984	rs116092985	rs7229491
rs28642975	rs1126	rs9959290



rs8086895

rs8095404

rs11661721

rs752720

rs56163207

rs9946862

rs732716

rs7248205

rs62105791

rs293710

rs551831885

rs6059462

rs6119422

rs6059908

rs4911448

rs62213709

20:33724557_GAA_G

rs2425024

rs619865

rs2425051

rs8120559

rs2425059

rs34414056

rs6120997

rs3746410

rs41293082

20:34391175_GAA_G

rs6058339

rs2746089

rs2425224

rs6091465

rs6021797

rs113321764

rs13058706

rs131820





附錄表五：全身脂肪體重之工具變數所使用之變異位點-英國生物資料庫

rs549596	rs6713448	4:140801333_GAAA_G
rs1318408	rs144916872	rs13105040
rs79520136	rs57703366	rs6822147
rs12125768	rs34974173	rs115947941
rs1359940	rs755892176	rs17055653
rs543874	rs4671328	rs9392975
rs10489882	rs17551326	6:8993316_CA_C
rs2902210	rs750588370	rs3864311
rs1336784	rs74418956	rs80154106
rs10798585	rs1528451	rs707894
rs1336777	rs2312193	rs2066295
rs9430020	rs78401742	rs4412193
rs7538503	rs13135092	rs6456739

rs10946844	rs1328288	rs6562009
rs4711122	rs11012732	rs1578273
rs13215054	rs10886016	13:90047856_GAAAAGAAAA_G
rs445150	rs705147	rs74464729
rs7762253	rs1359974	rs10483534
rs2814965	rs151017225	rs10146997
rs116133659	rs7127507	rs76186975
rs148828403	rs6592364	rs112162790
rs4082052	rs61910771	rs1242889
rs9341388	rs7305229	rs56165550
6:152169613_TA_T	rs297924	rs143504748
rs61250449	rs12821774	rs78748952
rs6989354	rs10862520	rs10852521
rs4072917	rs11060146	rs3751813
rs6477187	rs9527458	rs7191938



rs77483079

rs36030660

18:57835314_CT_C

18:57881478_TA_T

rs17066842

rs2620829

rs382634

rs897751

rs7250425

rs3736328

rs4808762

rs4420638

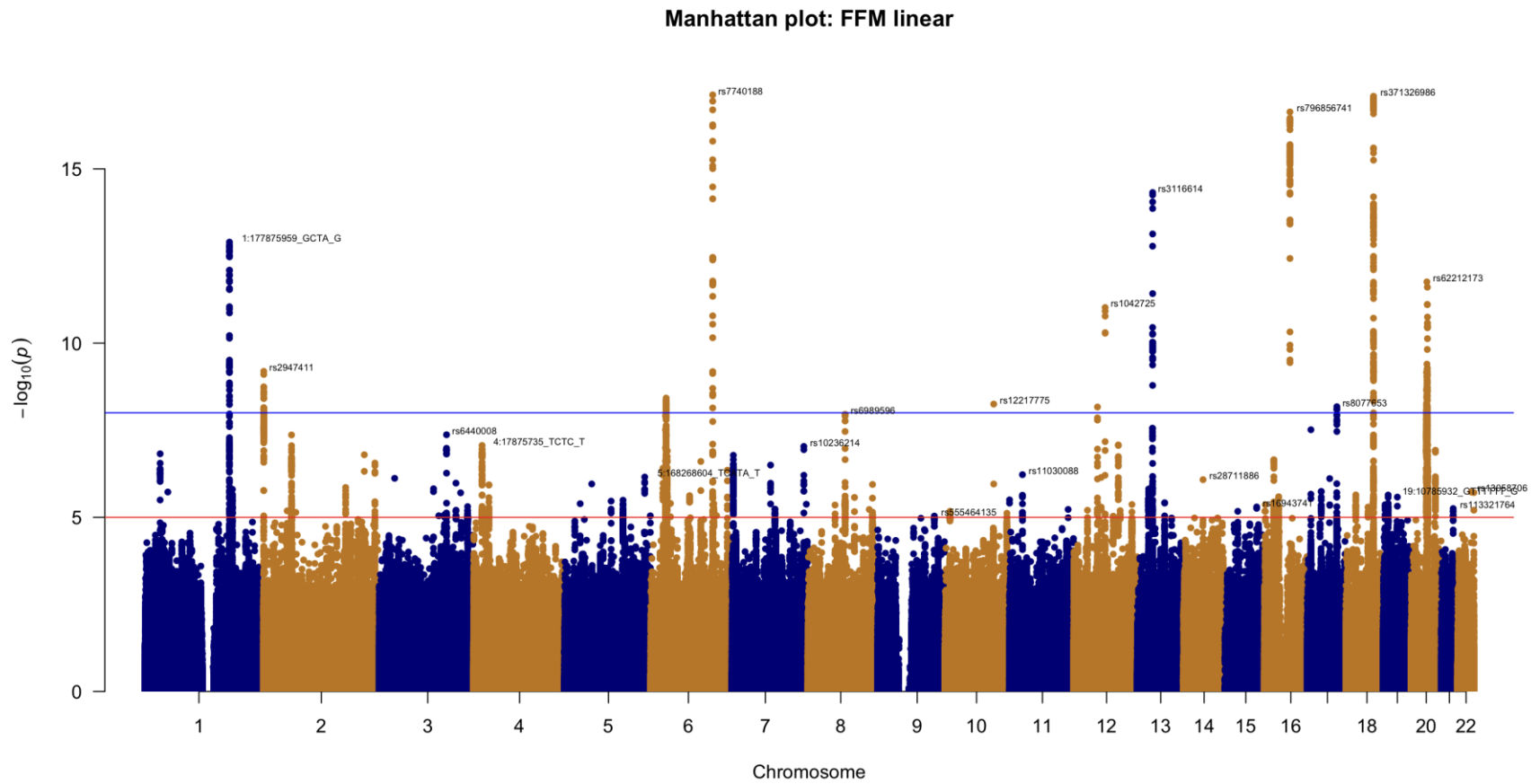
rs11879227

rs6059655



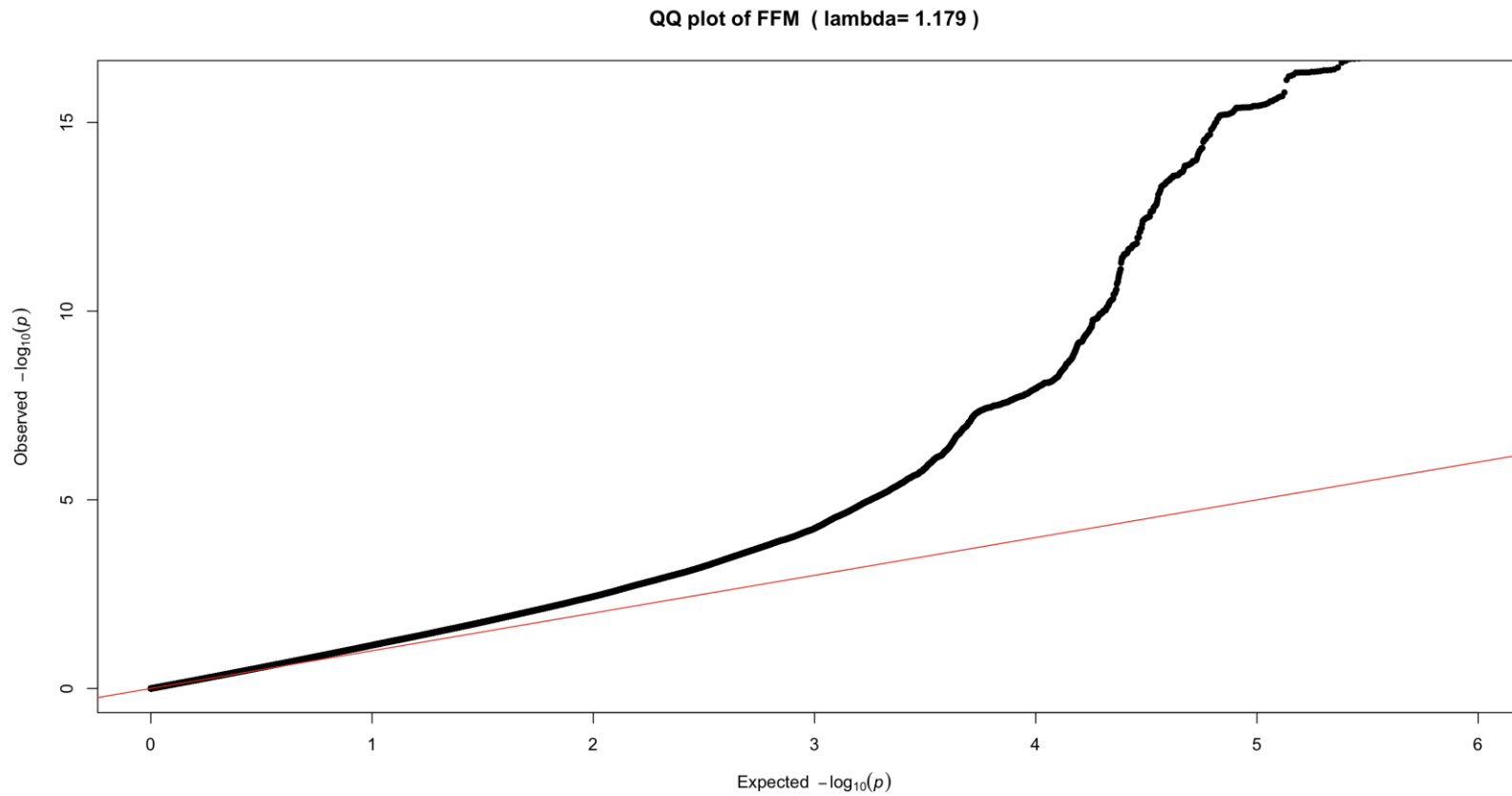


附錄圖三：全基因組關聯分析曼哈頓圖-去脂肪體重





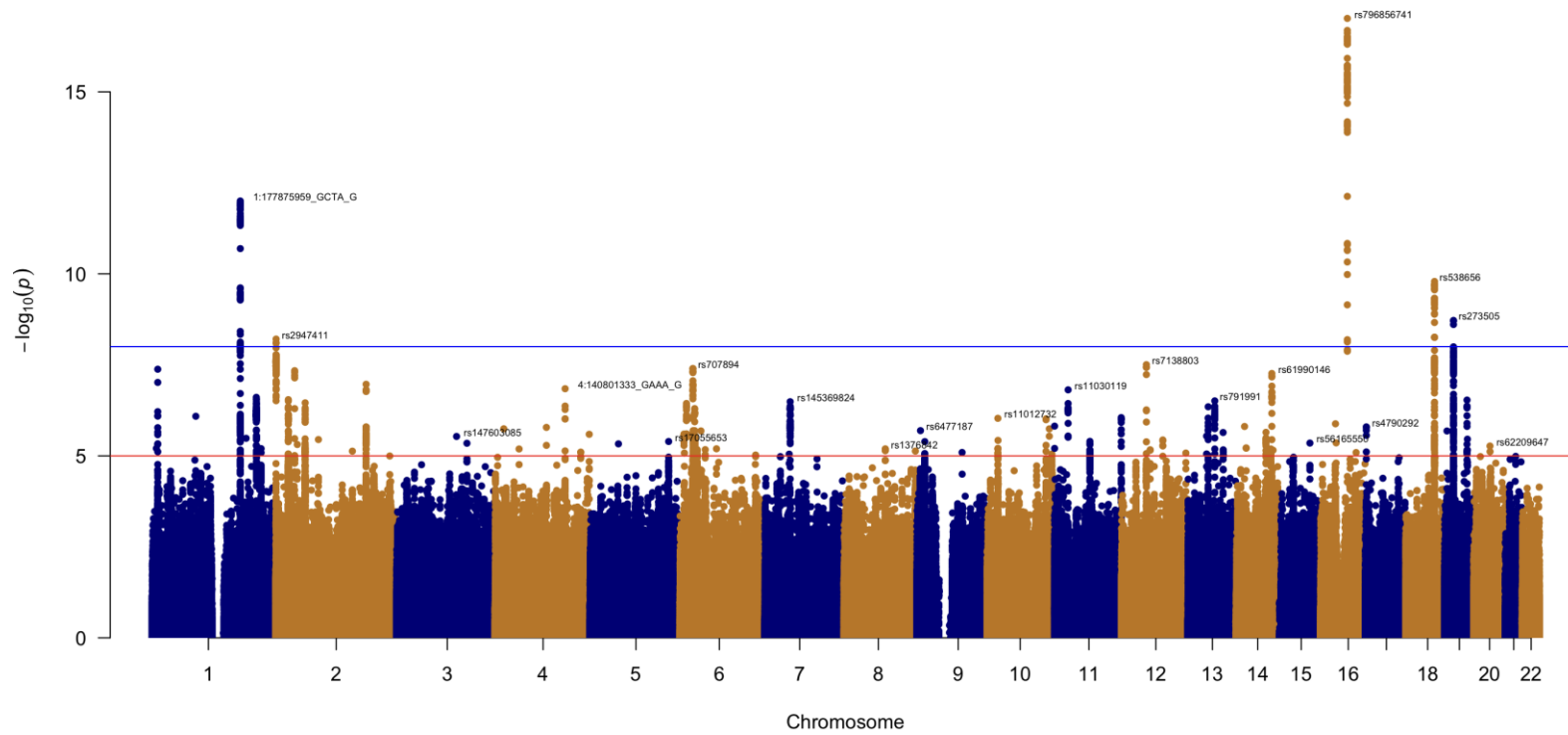
附錄圖四：全基因組關聯分析 QQ-plot-去脂肪體重





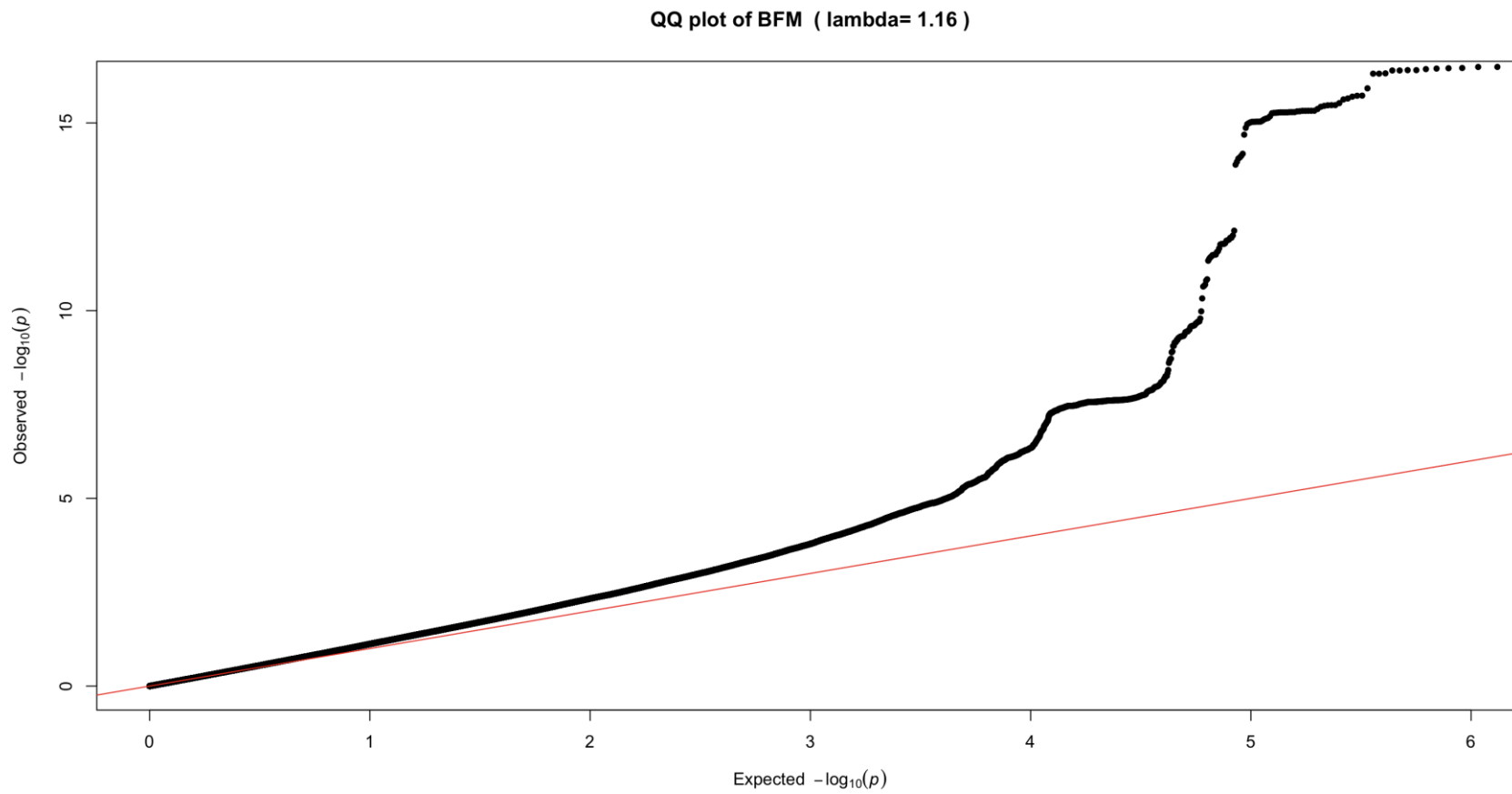
附錄圖五：全基因組關聯分析曼哈頓圖 全身脂肪體重

Manhattan plot: BFM linear



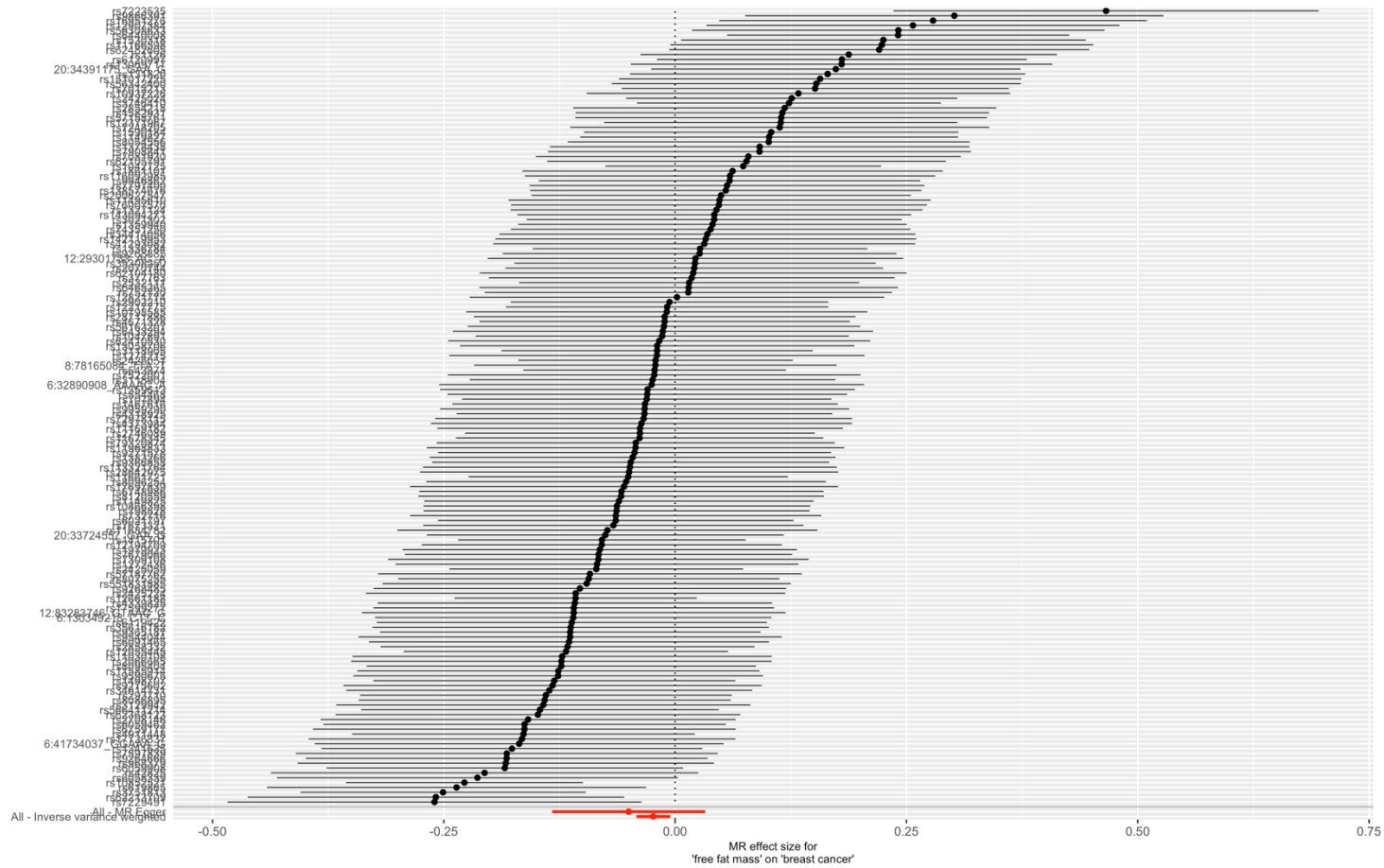


附錄圖六：全基因組關聯分析 QQ-plot 全身脂肪體重



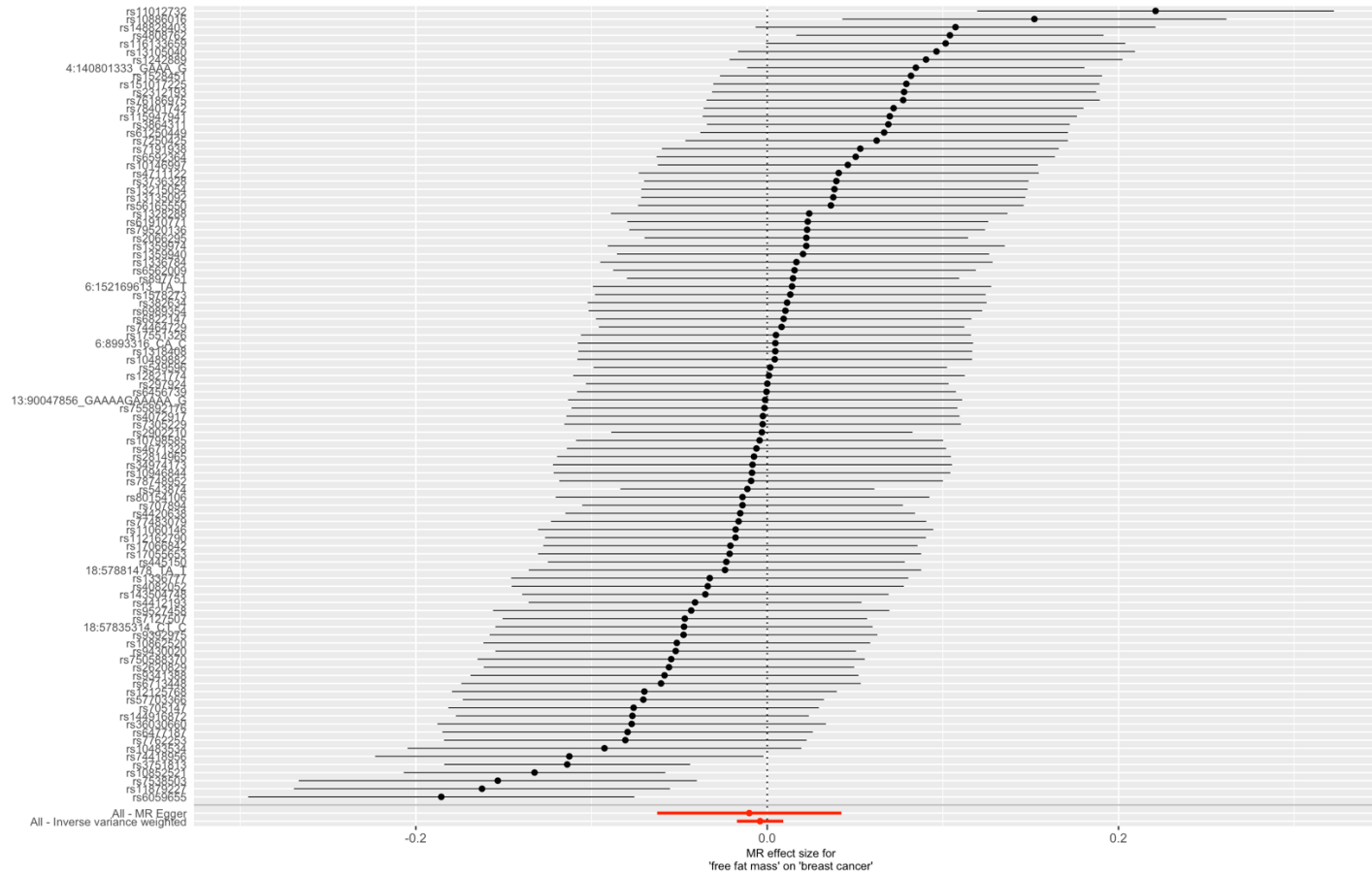


附錄圖七：去脂肪體重單一突變位點的 Wald 比率估計





附錄圖八：全身脂肪體重單一突變位點的 Wald 比率估計





附錄表六：乳癌風險之工具變數所使用之變異位點-英國生物資料庫

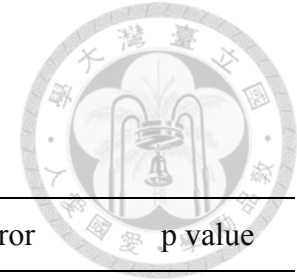
rs615022	rs11915002	rs9371545
rs10655740	rs78649327	rs80309700
rs150118532	rs2330572	rs34520389
rs142333882	rs930395	rs35640336
rs62186333	rs16885991	rs10096351
rs12465506	rs78601188	rs566510
rs12694403	rs1862625	9:110939007_CTTTTT_C
rs13409592	rs62355901	10:21807469_GA_G
rs72951831	rs6890270	rs61862866
rs2372966	rs62356591	rs17830944
rs1352941	rs58979455	rs11200007
rs9815383	rs35469600	rs4751844
rs9811454	rs75859313	rs45631611

rs45631632
rs3135724
rs7923788
rs1219515
rs7395522
rs139074399
rs680618
rs614367
rs525457
rs75915166
rs34507830
rs42294

rs805511
rs812020
12:28344563_TC_T
rs76191003
rs10879214
rs11836367
rs2042862
rs11844632
rs1468280
rs1111481
16:52564253_CTT_C
rs4784223

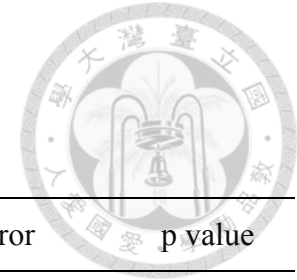
rs8053784
rs9708611
rs56971155
18:24518050_AT_A
rs62082103
rs117719194
21:16565066_TTTTAA_T
rs11089974
rs73171018
rs73161323
rs41311445





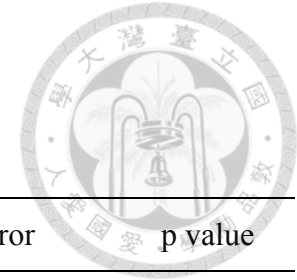
附錄表七：體脂率-乳癌風險的單樣本孟德爾隨機化試驗結果

outcome	exposure	method	SNPs	beta	standard error	p value
breast cancer	body fat percentage	2SLS	95	-0.007	0.010	0.490
breast cancer	body fat percentage	Inverse variance weighted	95	-0.007	0.009	0.488
breast cancer	body fat percentage	MR Egger	95	0.032	0.034	0.354
breast cancer	body fat percentage	MR Egger Intercept	95	-0.010	0.009	0.246
breast cancer	body fat percentage	Weighted median	95	-0.001	0.012	0.910



附錄表八：去脂肪體重指數-乳癌風險的單樣本孟德爾隨機化試驗結果

outcome	exposure	method	SNPs	beta	standard error	p value
breast cancer	Free fat mass index	2SLS	112	-0.045	0.036	0.210
breast cancer	Free fat mass index	Inverse variance weighted	112	-0.044	0.037	0.230
breast cancer	Free fat mass index	MR Egger	112	-0.113	0.116	0.333
breast cancer	Free fat mass index	MR Egger Intercept	112	0.005	0.008	0.535
breast cancer	Free fat mass index	Weighted median	112	-0.085	0.045	0.058



附錄表九：全身脂肪體重指數-乳癌風險的單樣本孟德爾隨機化試驗結果

outcome	exposure	method	SNPs	beta	standard error	p value
breast cancer	Body fat mass index	2SLS	78	-0.005	0.018	0.793
breast cancer	Body fat mass index	Inverse variance weighted	78	-0.009	0.021	0.683
breast cancer	Body fat mass index	MR Egger	78	0.006	0.072	0.930
breast cancer	Body fat mass index	MR Egger Intercept	78	-0.002	0.010	0.830
breast cancer	Body fat mass index	Weighted median	78	-0.004	0.024	0.868



中英對照與縮寫表：

孟德爾隨機化試驗 (Mendelian randomization, MR)
工具變數 (Instrumental variables, IV)
乳腺癌協會聯盟 (Breast Cancer Association Consortium, BCAC)
世界癌症研究基金會 (World Cancer Research Fund, WCRF)
美國癌症研究所 (American Institute for Cancer Research, AICR)
全基因組關聯分析 (Genome-wide association study, GWAS)
全表型組關聯分析 (Phenome-wide association studies, PheWAS)
二階最小平方法 (Two-stage least squares, 2SLS)
逆方差加權法 (Inverse variance weighted)
加權中位數法 (weighted median)
兩樣本 t 檢定 (Two Sample t-test)
皮爾森卡方檢定 (Pearson's Chi-squared test)
基因多效性 (pleiotropy)
體脂率 (Body fat percentage, BFP)
去脂肪體重 (Free fat mass, FFM)
去脂肪體重指數 (Free fat mass index, FFMI)

全身肪體重 (Body fat mass, BFM)
全身肪體重指數 (Body fat mass index, BFMI)
腰臀比 (Waist-hip ratio, WHR)
親脂性持久性有機污染物 (persistent organic pollutants, POP)