

國立臺灣大學醫學院基因體暨蛋白體醫學研究所

博士論文

Graduate Institute of Medical Genomics and Proteomics

College of Medicine

National Taiwan University

Doctoral Dissertation

利用主成分分析多性狀全基因組關聯分析

尋找慢性腎臟病的新風險位點

Discovering Novel Loci of Chronic Kidney Disease via

Principal Component Analysis based

Multiple-trait Genome-wide Association Study

莊國璣

Gwo-Tsann Chuang

指教教授：張以承博士

Advisor: Yi-Cheng Chang, M.D.-Ph.D

中華民國 113 年 6 月

June, 2024



國立臺灣大學博士學位論文

口試委員會審定書

DOCTORAL DISSERTATION ACCEPTANCE CERTIFICATE
NATIONAL TAIWAN UNIVERSITY

利用主成分分析多性狀全基因組關聯分析尋找慢性腎臟病的新風險位點

Discovering Novel Loci of Chronic Kidney Disease via Principal Component
Analysis based Multiple-trait Genome-wide Association Study

本論文係莊國璫（D06455001）在國立臺灣大學醫學院基因體暨蛋白質體醫學研究所完成之博士學位論文，於民國113年06月06日承下列考試委員審查通過及口試及格，特此證明。

The undersigned, appointed by the Graduate Institute of Medical Genomics and Proteomics, College of Medicine, National Taiwan University on 6th June 2024 have examined a Doctoral Dissertation entitled above presented by Gwo-Tsann Chuang (D06455001) candidate and hereby certify that it is worthy of acceptance.

口試委員 Oral examination committee:

張以承

李妮麗

許書育

(指導教授 Advisor)

陳建勳

高子州

系（所、學位學程）主管 Director:



誌 謝



本論文可以順利完成，承蒙恩師張以承老師悉心教導。在研究期間對研究方向、論文內容架構、資料處理給予學生諸多開導與鼓勵，師恩浩瀚，永懷於心，在此致上最高敬謝之意。由於已經進入臨床工作的緣故，我的研究時間主要是下班之後的時間進行，因此與實驗室同仁相處的時間並不多；然而遇上需要討論事情之際，我們溝通仍然很有效，也能夠順利解決問題。感謝冠靜、翠微、振豪以及奔浩等等張以承老師實驗室的同仁。

感謝臺灣人體生物資料庫的沈志陽老師、李德章老師、褚候維博士、蘇明威博士、林建維專員。感謝許書睿老師、李妮鍾教授、盧子彬教授和陳建勳教授對於口試的指教。最後，謹以此論文獻給我的家人們，謝謝您們的支持與鼓勵，使我無後顧之憂得以順利完成學業。

莊國璨 謹誌 2024.6.25

中文摘要



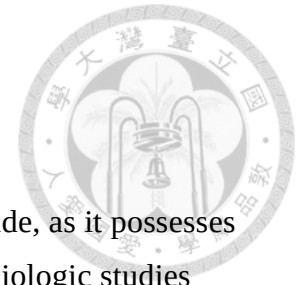
慢性腎臟病是全球各國的重要醫療議題，因為其影響醫療資源甚鉅。已經有許多流行病學研究尋找出慢性腎臟病傳統的危險因子，在許多這類的研究，慢性腎臟病通常被操作定義為腎絲球過濾率估值(estimated glomerular filtration rate; eGFR) $< 60\text{ml/min/1.73m}^2$ 。然而許多慢性腎臟病第三至第五期的人，仍然可以保有穩定的腎功能，所以事實上腎功能的變化軌跡比慢性腎臟病的現下分期還來的重要，而其中白蛋白尿(albuminuria)的程度又是影響腎功能變化的主因。近年 National Kidney Foundation 也已經建議針對臨床試驗案，可以在相對較短期的追蹤年限使用白蛋白尿和 eGFR 下降速率(eGFR slope)來替代傳統長期追蹤到腎功能大幅下降 30%或是進入腎衰竭等等定義。

藉由執行白蛋白尿、eGFR 以及 eGFR slope 個別的全基因組關聯分析(genome-wide association study; GWAS)可以尋找腎臟功能惡化進而發展慢性腎臟病的遺傳風險因子。近年在 GWAS 的領域，亦提倡藉由多性狀研究的方式來增加 GWAS 的檢力。比較廣為人知的是 Multi-Trait Analysis of GWAS (MTAG)，其方式為對數個性狀的 GWAS summary 應用廣義逆方差加權法統合分析(generalized inverse-variance-weighted meta-analysis)。從另一個面向來看，由於以上三種性狀因為都是連續變項，我們也可以用主成分分析將這些數量性狀先轉化成不同的主成分(principal component; PC)，再對這些 PC 進行 GWAS (PC-GWAS)。

本研究執行了白蛋白尿、eGFR 和 eGFR slope 的單性狀 GWAS，以 PC-GWAS。藉由比較不同方法之間的結果，以及對比文獻之後可以發現 PC-GWAS 找到 20 個新位點，其中有些鄰近的基因從文獻上由其功能描述或是小鼠實驗結果已經可以推估和腎臟功能或疾病相關；另外有 4 個位點在腎臟表現數量性狀基因座資料庫發現其具顯著的影響。未來設計功能分析研究來這些新位點是否和腎臟功能確實相關仍然是目前檢視 GWAS 結果的黃金標準。藉由本研究的經驗，PC-GWAS 的方式也值得在適合的疾病和族群應用和推廣。

關鍵詞：慢性腎臟病，白蛋白尿，腎絲球過濾率預估值，全基因組關聯分析，多性狀全基因組關聯分析

英文摘要



Chronic kidney disease (CKD) is an important issue global-wide, as it possesses tremendous medical resource burden. There are numerous epidemiologic studies finding conventional risk factors for CKD, and the operational definition for CKD patients in these studies is usually individuals with estimated glomerular filtration rate (eGFR) $<60\text{ml/min/1.73m}^2$. However, many patients with CKD stage 3-5 have stable renal function for years, and thus the change of renal function with time is more important than the current CKD stage. Proteinuria/albuminuria is the major determinant of eGFR trajectory. National Kidney Foundation has suggested change in albuminuria and eGFR slope fulfill criteria for surrogate end points in clinical trials for chronic kidney disease progression.

We can apply genome-wide association study (GWAS) on complex diseases such as CKD to find its heritable factors. Only few individuals of Taiwan biobank have eGFR below 60ml/min/1.73m^2 , and thus a case-control GWAS may not be appropriate. In addition, evaluating the risk factor for eGFR change and albuminuria is even more important for renal prognosis.

We have performed quantitative GWASs regarding eGFR per se, eGFR slope and albuminuria to find risk loci for CKD progression. There were also several multi-trait methods developed in recent years to enhance the power of GWAS. We used principal component analysis to convert these quantitative traits into pseudotraits (principal components; PC), and then perform the GWAS of the pseudotraits (PC-based GWAS). A comparison of result of different methods and literature review were done. Twenty novel risk loci associated with kidney traits were identified by PC-based GWAS. Subsequent post-GWAS analyses and functional analyses will be done accordingly in future.

Keyword: chronic kidney disease, albuminuria, estimated glomerular filtration rate, genome-wide association study, multi-trait genome-wide association study

目次



口試委員會審定書.....	i
誌謝.....	ii
中文摘要.....	iii
英文摘要.....	iv
目次.....	v
圖次.....	vii
表次.....	viii
第一章 緒論.....	1
1.1 研究背景.....	1
1.2 研究動機與目的.....	2
第二章 研究方法.....	5
2.1 研究架構.....	5
2.2 研究對象.....	5
2.3 基因體資料.....	6
2.3.1 全基因體定型資料.....	6
2.3.2 基因體定型資料品質管制.....	7
2.3.3 基因型插補以及插補後之資料品質管制.....	8
2.3.4 TWB v1.0 和 TWB v2.0 插補後資料合併.....	8

2.4 全基因組關聯分析.....	8
2.4.1 單性狀全基因組關聯分析.....	8
2.4.2 主成分分析多性狀全基因組關聯分析.....	9
第三章 結果.....	10
3.1 單性狀全基因組關聯分析結果.....	10
3.1.1 Albuminuria GWAS results.....	10
3.1.2 eGFR GWAS results.....	12
3.1.3 eGFR slope GWAS results.....	14
3.2 主成分分析多性狀全基因組關聯分析.....	16
3.2.1 eGFR, eGFR slope, albuminuria 主成分分析.....	16
3.2.2 主成分進行 GWAS 之結果.....	19
3.2.3 主成分分析多性狀全基因關聯分析額外得到之位點.....	24
第四章 討論與結論.....	48
4.1 結果討論.....	48
4.2 結論.....	50
參考文獻.....	51
附錄.....	57

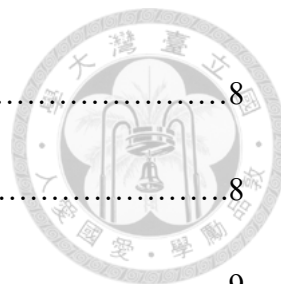


圖 次

圖 1 研究分析流程圖.....	5
圖 2 白蛋白尿 GWAS 的曼哈頓圖及 Q-Q 圖.....	11
圖 3 腎絲球過濾率預估值 GWAS 的曼哈頓圖及 Q-Q 圖.....	13
圖 4 腎絲球過濾率預估值下降率 GWAS 的曼哈頓圖及(B) Q-Q 圖.....	15
圖 5 原始性狀 eGFR, eGFR slope 和 albuminuria 在 PC1 和 PC2 的雙標圖.....	18
圖 6 主成分 PC1 GWAS 的曼哈頓圖及 Q-Q 圖.....	20
圖 7 主成分 PC2 GWAS 的曼哈頓圖及 Q-Q 圖.....	21
圖 8 主成分 PC3 GWAS 的曼哈頓圖及 Q-Q 圖.....	22
圖 9 主成分 GWAS 得到之新風險位點的區域圖.....	28

表 次



表 1 白蛋白尿 GWAS 基本資料.....	10
表 2 白蛋白尿 GWAS 顯著的 SNPs.....	12
表 3 腎絲球過濾率預估值 GWAS 基本資料.....	12
表 4 腎絲球過濾率預估值下降速率 GWAS 基本資料.....	14
表 5 腎絲球過濾率預估值下降率 GWAS 顯著的 SNPs.....	16
表 6 主成分分析多性狀全基因組關聯分析的參與者的基本資料.....	17
表 7 各主成分解釋 eGFR, eGFR slope, albuminuria 三項原始性狀之變異量.....	17
表 8 各主成分由 eGFR, eGFR slope 和 albuminuria 的負荷向量組成.....	17
表 9 主成分分析多性狀全基因關聯分析顯著的 SNPs.....	23
表 10 主成分分析多性狀全基因關聯分析得到的 novel loci 與 eQTL 結果.....	27

第一章 緒論



1.1 研究背景

複雜性疾病(complex disease)，亦稱為多因性疾病(multifactorial disease)，相對於單基因疾病而言是由多基因以及環境因子交互作用而發生[1]。尋找單基因疾病的致病基因是使用連鎖分析(linkage analysis)的方法[2]，而尋找複雜疾病的諸多相關基因目前最廣為使用的是全基因組關聯分析(genome-wide association study; GWAS)[3]。

全基因組關聯分析可以針對疾病，即病例對照(case-control)全基因組關聯分析，也可以針對數量性狀(quantitative trait)分析。一般常見疾病因為非單基因疾病而是由多個基因影響，基本上可以視為數量性狀來評估，而反之數量性狀一定是多基因影響的性狀[4]。比如高血壓是以疾病的方式以血壓切值來看待，但是尋找和高血壓有關的遺傳因子其實可以直接以血壓值作為數量性狀來進行全基因組關聯分析；效應大小的表示在病例對照 GWAS 為勝算比(odds ratio)，在數量性狀 GWAS 為 beta 係數(beta coefficient)，即每增加一個 A1 對偶基因則此數量性狀改變的量。

之前我們團隊已經執行了數個數量性狀的全基因組關聯研究並且有三篇已經發表於國際性期刊(原文詳見附錄)，包括：一、尋找調控血清纖維母細胞生長因子 21 及 23 的遺傳因子[5]，此篇為歷來文獻第一篇探索血清 FGF21 濃度相關的遺傳因子以及第二篇(東亞族群第一篇) 探索血清 FGF23 濃度相關遺傳因子的文章；二、影響褪黑激素分泌量的遺傳因子[6]，此篇為歷來第一篇探索褪黑激素夜間分泌量相關遺傳因子的文章；三、和白蛋白尿量相關的位點[7]，此篇是針對台灣族群第一篇尋找和白蛋白尿相關的遺傳因子之全基因組關聯研究，研究找到的新位點以及對應的基因提供增進瞭解東方族群白蛋白尿機轉甚至慢性病惡化的原因。以上三篇文章本人皆為主要分析者及共同第一作者，藉由這些



研究的經驗和過程，可以發現將性狀選擇適當分析的型態很重要。一般性的問題是如果原始數量性狀未呈常態分佈，需要以對數變換(log transformation)等其它變換方式處理後才來進行 GWAS 最後的線性相關分析。另外評估白蛋白尿的程度非取得尿中白蛋白濃度而已，而是同時需要尿中的肌酸酐濃度，才能做尿液濃淡的校正；以尿液評估褪黑激素的釋放量，則需要一致的以早上的尿中褪黑激素代謝物 aMT6s(6-sulfatoxymelatonin)及尿液肌酸酐的濃度來評估，原因是睡眠時夜間的褪黑激素釋放量遠大於白天時。未將欲分析的性狀做適當的調整，則分析得到的結果將不可信。

1.2 研究動機與目的

由於本人是腎臟科醫師，所以對於尋找慢性腎臟病以及腎功能惡化的風險位點極具興趣。慢性腎臟病是全球各國的重要醫療議題，因為其影響醫療資源甚鉅[8]。已經有許多流行病學研究尋找出慢性腎臟病傳統的危險因子，在許多這類的研究，慢性腎臟病通常被操作定義為腎絲球過濾率預估值(estimated glomerular filtration rate; eGFR) $< 60\text{ml/min/1.73m}^2$ ，意即慢性腎臟病第三期以上。然而許多慢性腎臟病第三至第五期的人，仍然可以保有穩定的腎功能，所以事實上腎功能的變化軌跡比慢性腎臟病的現下分期還來的重要，而其中白蛋白尿的程度又主要是影響腎功能變化的主因[9]。近年美國國家腎臟基金會(National Kidney Foundation)也已經建議針對臨床試驗案，可以在相對較短期的追蹤年限使用白蛋白尿(albuminuria)和 eGFR 下降速率(eGFR slope)來替代傳統長期追蹤到腎功能大幅下降 30%或是進入腎衰竭等等定義[10]。

慢性腎臟病為複雜性疾病，我們可以用全基因組關聯分析來尋找和其相關的遺傳因子[11]。若是原始的數量性狀近似常態分佈(如 eGFR)，則進行數量性狀 GWAS 獲得的結果將會比二元的病例對照(如慢性腎臟病常用的操作定義為 $\text{eGFR} < 60\text{ ml/min/1.73m}^2$) GWAS 佳，因其保留了比較多的原始資訊。由於在人



體生物資料庫的受試者由於主體是相對健康者，所以只有少數人的 eGFR 會落在 $< 60 \text{ ml/min/1.73m}^2$ 的數值，所以使用傳統慢性腎臟病的操作定義來進行病例對照全基因組關聯分析不太適合；在病例比重近似在族群的盛行率之下，以病例對照這樣的二元變項進行全基因組關聯分析相較於以連續變項進行分析會喪失許多檢力[12]。另外如前段落所述，評估找尋出腎功能變化以及白蛋白尿程度的因子其實對於判斷預後更有幫助。因此除了執行 eGFR 本身，合併 albuminuria GWAS 以及 eGFR slope GWAS 來找尋腎臟功能惡化進而發展慢性腎臟病的遺傳風險因子。

個別的數量性狀(quantitative trait)只能描述疾病的一個面向，並不能夠完整詮釋疾病本身，因此個別的 quantitative trait GWAS 只能尋找到部分和疾病本身相關的遺傳風險因子。不但如此，即便將這些個別的 GWAS 結果合在一起看，仍然可能因為檢力(power)不足有所漏失。舉例來說，假設一個變異點或是單核苷酸多型性(single-nucleotide polymorphism; SNP) 它對腎臟功能確實有影響，因此也對於這些腎臟的性狀都有影響，即這個 SNP 具有水平多效性(horizontal pleiotropy)，但是如果以個別單一的性狀去做關聯分析時可能不夠顯著而無法被找尋到。針對這個問題，近年有一些 multiple-trait analysis 的方法被發展出來，可以增加檢力，如 multi-trait analysis of genome-wide association (MTAG) [13]。

從另一個面向來看，我們也可以用主成分分析(principal component analysis; PCA)直接將這些數量性狀先轉化成虛構性狀(pseudotrait)，其相對於個別原始性狀可以呈現最大的變異程度，也就是對於疾病其代表性可能更大[14]，進而再對此虛構性狀做 GWAS (PC-based GWAS)。針對腎臟功能或是慢性腎臟病，都是可以用數量性狀來評估，所以適用此方法。本計畫研究目的為針對 albuminuria，eGFR 和 eGFR slope 這三個性狀，我將執行它們個別的 GWAS，再以此三個性狀主成分分析之後產出的虛構性狀為代表慢性腎臟疾病及腎功能惡化，對此虛構症狀進行 GWAS。得出各個結果之後，除了比較各方法之間的

差異，也會對比目前文獻及資料庫來檢視是否有發現到新的風險基因位點，值得進一步做功能分析來探討慢性腎臟病以及腎功能惡化的機轉。



第二章 研究方法



2.1 研究架構

本研究流程如圖 1 所示。首先會進行白蛋白尿、腎絲球過濾率，以及腎絲球過濾率斜率這三個數量性狀的全基因組關聯分析，接著以 MTAG 當作第一種 multitrait analysis 的方式。另一方面，挑選出同時具備此三性狀者，以主成分分析的方式得到虛構性狀；依據總變異量需由幾個虛構性狀能夠完整解釋，去進行這些個別虛構性狀的全基因組關聯分析。將各個方式得到的結果進行比較，再對比目前文獻及資料庫的結果來看是否有發現到和慢性腎臟病、腎功能惡化相關的新風險基因位點。

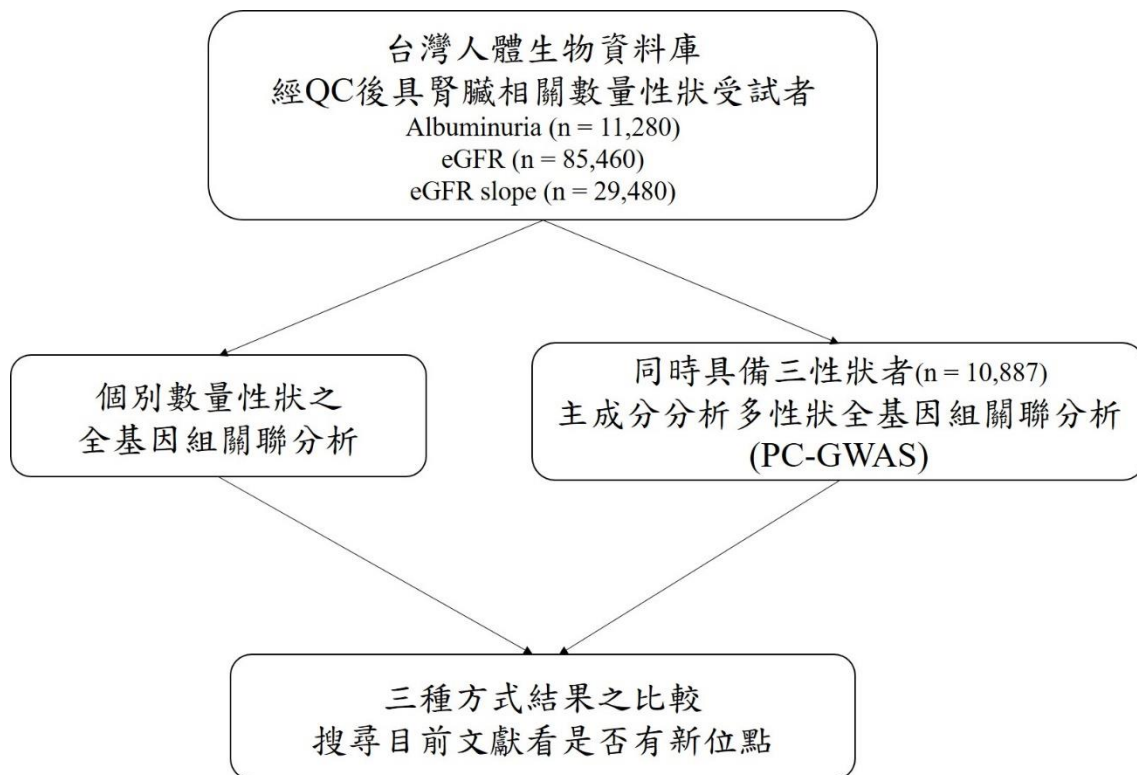


圖 1 研究分析流程圖

2.2 研究對象

本研究案對象皆來自臺灣人體生物資料庫(Taiwan Biobank)[15]。收案地點為



國內多家醫院，駐站分佈於北、中、南、東以及澎湖地區；志願參加受試者須年滿 20 歲，有行為能力且為本國籍；除參與收案前解說與簽屬知情同意書外，受試者亦須配合進行身體檢測(身高、體重、體脂肪、腰臀圍、血壓、脈搏、骨密度檢測與肺功能檢測。)，檢體採集(包含 30 毫升空腹靜脈血、15 毫升尿液與 5 管糞便)以及完成健康相關資訊問卷(個人基本資料、生活習慣、飲食習慣、生活環境暴露、生育史、個人疾病史、家族病史與人際關係互動等)。常規之血液檢驗項目包括紅血球、飯前血糖、糖化血色素、膽固醇、高密度膽固醇、三酸甘油脂、尿酸、尿素氮、肌酐酸、肝功能檢驗與肝炎病毒指標(C 型肝炎病毒抗體、B 型肝炎的抗原、表面抗體與核心抗體)；常規之尿液檢驗項目包括白蛋白以及肌酐酸。受試者若有意願可參與後續追蹤，再次提供身體檢測、檢體採集以及問卷填寫。台灣人體生物資料庫提供的所有資料皆不連結個人資訊內容，亦即全部為去連結之資料。本研究通過台大醫院研究倫理委員會核可(No.: 202312110RINA)

由於糖尿病是影響腎臟功能以及造成慢性腎臟病最重要的一個後天因子，因在此研究中我們將排除有糖尿病(問卷已回答有糖尿病，或是檢驗糖化血色素 $\geq 6.5\%$)的受試者來進行相關分析。

2.3 基因體資料 (genomic data)

2.3.1 全基因體定型資料 (whole-genome genotyping)

台灣人體生物資料庫目前有兩個晶片系統。TWB v1.0 基因型鑑定系統採用美國 Thermo Fisher Scientific 公司所研發之技術平台，與國家基因體研究中心 (National Center for Genome Medicine, NCGM) 合作設計專屬臺灣漢人的單一核苷酸多型性 (Single Nucleotide Polymorphism, SNP) 基因定型晶片，挑選共 653,291 個 SNPs，以台灣人體生物資料庫(Taiwan Biobank, TWB)的英文縮寫 TWB 命名。所挑選 SNPs 之來源，包括主體的 Axiom Genome-Wide CHB (Han Chinese in Beijing) Array (525,652 SNPs)、已經發表之癌症相關 GWAS 研究中具統

計意義之位點、其他定序方法中發現國人樣本具多型性之位點、以及其他與藥物反應或代謝相關之位點。

TWBv2.0 基因型鑑定晶片為透過資料庫 TWBv1.0 和全基因體定序的資訊和執行經驗，進一步針對臺灣族群客製化開發的系統。挑選一共約 75 萬單一核苷酸多型性位點(SNPs)，提供更適切臺灣族群及臨床使用的基因型鑑定位點。

2.3.2 基因體定型資料品質管制(quality control; QC)

為減少潛在錯誤之資料影響分析結果，全基因組關聯研究分析之前，需要對基因體定型資料做 QC，包括如下 (使用 PLINK 1.9 [16]進行)：

- 1) 進行性別確認：識別問卷自述性別和基因檢驗性別不一致的樣本
- 2) 將具有批次效應(源自晶片不同批次檢驗)之 SNPs 設為遺漏值
- 3) 基因定型率：排除 Missing rate 大於 5 %之 SNPs
- 4) 移除極端異型合子率(heterozygosity rate)者：若一個體之異型合子率相對於全體平均的異型合子率有過多偏差(超出 5 個標準差)，則此個體會被移除
- 5) 移除具有親屬關係者：排除雙胞胎、一等親或是二等親(即 identity-by-descent score 大於 0.1875 者)
- 6) 移除次等位基因頻率(minor allele frequency; MAF)小於 0.1 % 之 SNPs
- 7) 哈溫定律檢定 (Hardy-Weinberg Equilibrium, HWE)：移除 HWE 檢定 P 值小於 1×10^{-5} 之 SNPs
- 8) 排除超過 10% batch 為缺失之 SNPs
- 9) TWB v1.0 array 本來使用 Genome Reference Consortium Human Build 37 (GRCh37) 作為基因體座標 (genomic coordinate)，要將其轉換至新版的 GRCh38

2.3.3 基因型插補(genotype imputation)以及插補後之資料品質管制(post-imputation QC)

將全基因體定型資料，再以後端電腦運算以全基因體的定序的資料作為參考基因組(reference panel)，進行基因型插補以增加可分析之位點。使用的軟體為 SHAPEIT4[17]進行 phasing (對單倍型 haplotype 的推估)，再以 IMPUTE2[18]軟體進行 imputation；參考基因組是結合了 1000 Genome Project 中的 East Asian (EAS) population 資料 (504 人) 和來自 Taiwan Biobank 的全基因體定序資料 (1451 人)。插補後的 QC 包括移除 missing rate 大於 5 %之 SNPs，排除 MAF 小於 1 %之 SNPs 以及移除插補分數 information score 小於 0.3 之 SNPs。

2.3.4 TWB v1.0 和 TWB v2.0 插補後資料合併

由於研究對象的原始全基因體定型方式，一部分的人是使用 TWB v1.0，一部分的人是使用 TWB v2.0，另有少部分人兩者都有。為了資料的一致要將這些基因體資料作整合，以同時具有 TWB v1.0、TWB v2.0 晶片資料及全基因體定序資料的個案進行基因型一致性的評估，即取這三組資料交集的 SNPs 進行基因型一致性的評估；同一個案之同一 SNP 的基因型在三組資料不一致時，此 SNP 將被刪除。最後保留一共約 980 萬個 SNPs。

2.4 全基因組關聯分析 (genome-wide association study)

2.4.1 單性狀全基因組關聯分析

本篇使用的全基因組關聯分析，皆以 PLINK 1.9 軟體執行。以 Genome Reference Consortium Human Build 38 (GRCh38) 作為基因體座標，在線性遺傳模式下以線性回歸檢定 SNPs 和數量性狀(quantitative trait)的相關性，其中共變數包括年齡、性別及 PC1-PC10 以校正族群分層 (population stratification)。傳統上全基因組關聯分析將 P 值 5×10^{-8} 視為顯著閾值，但是由於其過份嚴格致使檢力下

降而許多真正的訊號會呈現偽陰性，所以本篇研究選擇 P 值小於 10^{-5} 的位點視為顯著[19]。使用 linkage disequilibrium (LD)-based clumping 來挑出各連鎖不平衡區塊中相關性最高之 SNP (top locus)。使用 R 軟體[20]之套件 qqman[21]來製作曼哈頓圖(Manhattan plot)和 Q-Q plot (quantile-quantile plot)，分別用來宏觀檢視全基因組具顯著相關的位點以及檢視結果是否存在族群分層的問題。

此研究相關的三個性狀為白蛋白尿(albuminuria)，腎絲球過濾率預估值(eGFR)，以及腎絲球過濾率預估值下降速率(eGFR slope)。白蛋白尿的原始表示方式為單一泡尿液的白蛋白尿濃度除以肌酐濃度 (urine albumin-to-creatinine ratio; UACR)，單位為 mg/g；由於 UACR 非常態分佈，接著要進行對數變換(log-transformed)，最終對 logUACR 來做關聯分析。腎絲球過濾率預估值是以抽血檢驗得到肌酐數值(S_{cr})之後採用 CKD-EPI 公式估計，其計算方式如下：

$$GFR \text{ (ml/min/1.73 m}^2\text{)} = 141 \times \min(S_{cr}/\kappa, 1)^\alpha \times \max(S_{cr}/\kappa, 1)^{-1.209} \times 0.993^{\text{Age}} \times 1.018$$

(if female) $\times 1.159$ (if black), where κ is 0.7 for females and 0.9 for males, α is -0.329 for females and -0.411 for males; min indicates the minimum of S_{cr}/κ or 1, and max indicates the maximum of S_{cr}/κ or 1 [22]。至於腎絲球過濾率預估值下降速率的表示方式為(eGFR 追蹤值 - eGFR 初始值) / 追蹤時間，單位為 ml/min/1.73 m²/year。

2.4.2 主成分分析多性狀全基因組關聯分析

在台灣人體生物資料庫受試者且非糖尿病患者之中，找尋同時具有 albuminuria，eGFR 以及 eGFR slope 這三個性狀資料的人，針對這群人進行 albuminuria，eGFR 以及 eGFR slope 的主成分分析；應用 R 軟體之內建指令 prcomp 來進行主成分分析。接著再以前幾個已經可以解釋所有變異量的主成分(principal components; PC)，亦即虛構性狀，來進行全基因組關聯分析。

第三章 結果



3.1 單性狀全基因組關聯分析結果

3.1.1 Albuminuria GWAS results

Albuminuria GWAS 參與者的基本資料如表 1。單性狀 Albuminuria GWAS 的 manhattan plot 和 Q-Q plot 如圖 2 所示，顯著結果整理如表 2 所示。

表 1 白蛋白尿 GWAS 基本資料

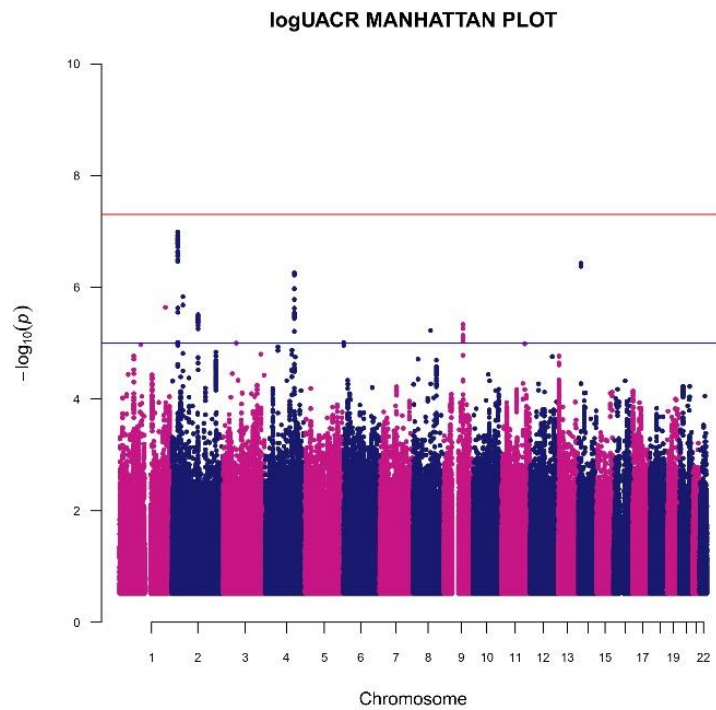
分析人數, N	11,280
年齡, year	48.8 ± 10.4
男性, N (%)	4,285 (38.0%)
UACR*, mg/g	8.355 (4.819-9.554)

*expressed as median (Q1-Q3)



圖 2 白蛋白尿 GWAS 的(A)曼哈頓圖及(B) Q-Q 圖

(A)



(B)

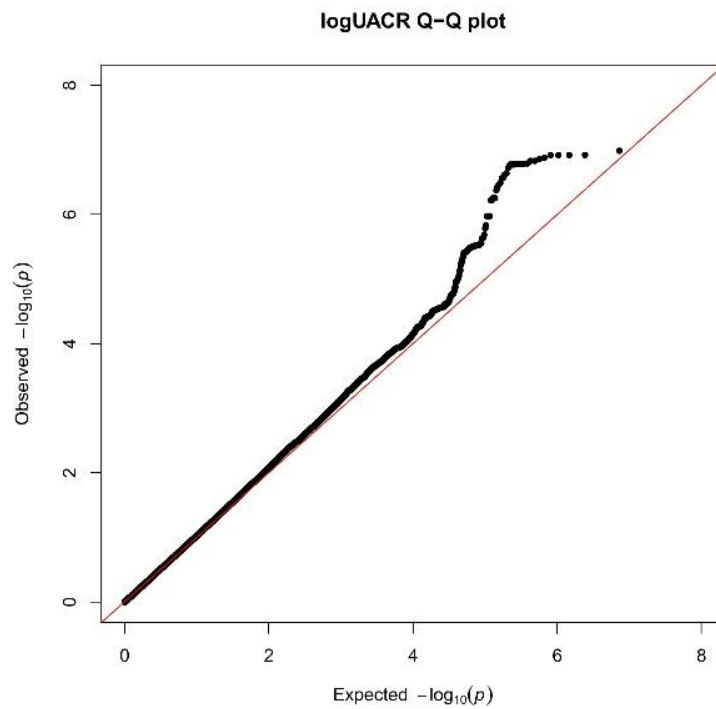


表 2 白蛋白尿 GWAS 顯著的 SNPs

Chr	BP	SNP	A1	A2	MAF	BETA	SE	P	Nearby gene
1	218026147	rs193129422	G	T	0.011	0.069	0.014	2.3E-06	SPATA17
2	27512105	rs6547692	G	A	0.488	0.015	0.003	2.7E-07	GCKR
2	27598615	rs12989678	T	C	0.466	0.016	0.003	1.0E-07	C2orf16
2	51728241	rs190931819	G	C	0.052	0.033	0.007	1.5E-06	ZNF512
2	123539652	rs12711641	A	C	0.223	0.017	0.004	3.1E-06	NRXN1
4	138579033	rs1376095	A	G	0.210	-0.020	0.004	5.6E-07	CNTNAP5
6	1245061	rs11970205	A	G	0.044	0.033	0.007	9.7E-06	SLC7A11
8	82182829	rs79425178	T	C	0.031	0.039	0.009	5.9E-06	FOXQ1
9	90576444	rs34232821	T	G	0.165	0.018	0.004	4.6E-06	SNX16
14	32571580	rs1155069	C	T	0.038	0.040	0.008	3.7E-07	DIRAS2

3.1.2 eGFR GWAS results

eGFR GWAS 參與者的基本資料如表 3。單性狀 eGFR GWAS 的 manhattan plot 和 Q-Q plot 如圖 3 所示，結果有 210 個顯著 top loci，如附錄-表 1 所示。

表 3 腎絲球過濾率預估值 GWAS 基本資料

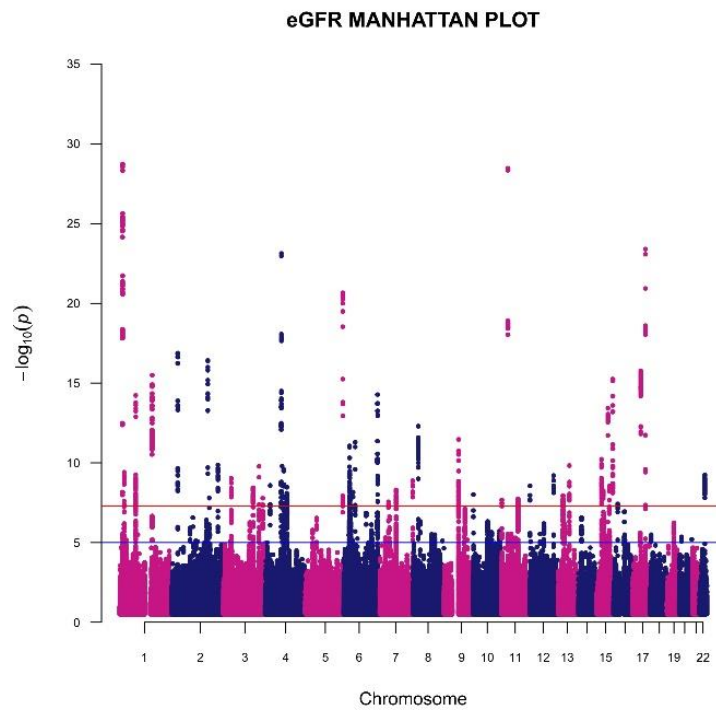
分析人數, N	85,460
年齡, year	49.1 ± 10.7
男性, N (%)	31,040 (36.3%)
eGFR, ml/min/1.73m ²	105.9 ± 13.2



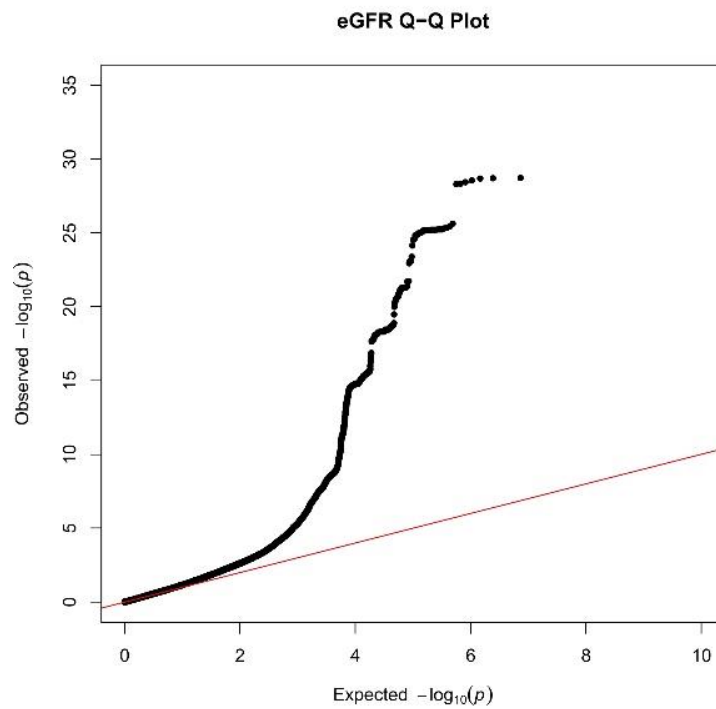
圖 3 腎絲球過濾率預估值 GWAS 的(A)曼哈頓圖及(B) Q-Q 圖

(A)

(B)



(B)





3.1.3 eGFR slope GWAS results

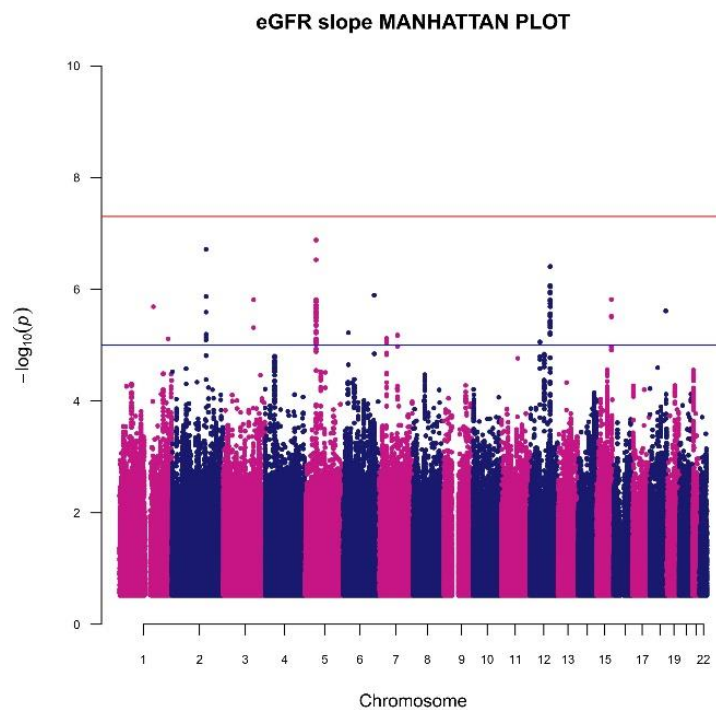
eGFR slope GWAS 參與者的基本資料如表 4。單性狀 eGFR slope GWAS 的 manhattan plot 和 Q-Q plot 如圖 4 所示，顯著結果整理如表 5 所示。

表 4 腎絲球過濾率預估值下降速率 GWAS 基本資料

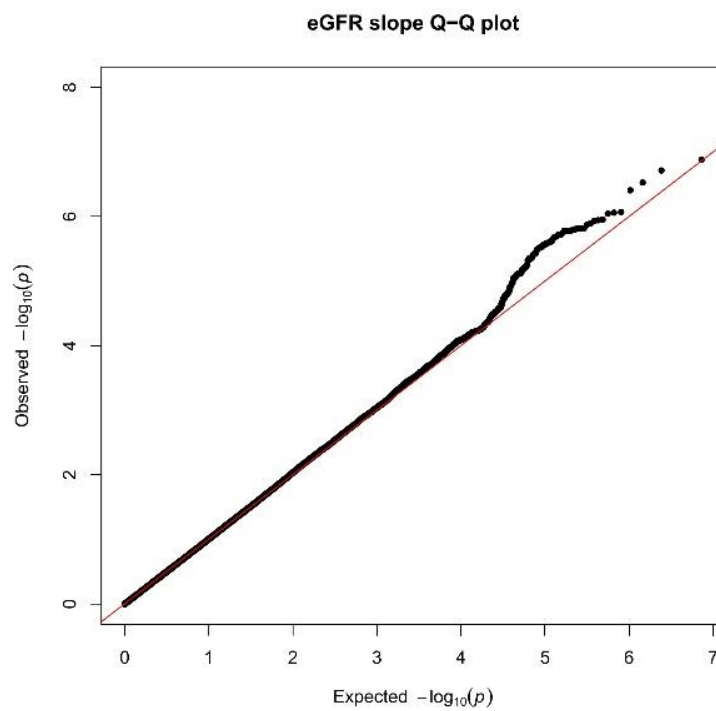
分析人數	29,480
年齡 (baseline)	49.7 ± 10.3
男性, N (%)	9,912 (33.6%)
Baseline eGFR, ml/min/1.73m ²	105.1 ± 13.0
追蹤時間, year	4.2 ± 1.2
eGFR slope, ml/min/1.73m ² /year	-0.76 ± 1.38

圖 4 腎絲球過濾率預估值下降速率 GWAS 的(A)曼哈頓圖及(B) Q-Q 圖

(A)



(B)



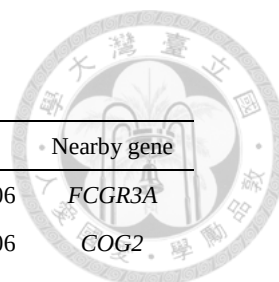


表 5 腎絲球過濾率預估值下降率 GWAS 顯著的 SNPs

Chr	BP	SNP	A1	A2	MAF	BETA	SE	P	Nearby gene
1	161538827	rs10919543	G	A	0.322	0.057	0.012	2.1E-06	<i>FCGR3A</i>
1	230689754	rs61828618	T	C	0.063	-0.100	0.023	7.7E-06	<i>COG2</i>
2	161565599	rs190473377	C	T	0.015	-0.240	0.046	1.9E-07	<i>SLC4A10</i>
3	143888112	rs74521978	T	A	0.014	0.227	0.047	1.5E-06	<i>SLC9A9</i>
5	51832131	rs202180998	G	T	0.027	-0.180	0.035	3.0E-07	<i>ISL1</i>
6	23274391	rs34935015	A	T	0.019	-0.190	0.042	6.0E-06	<i>HDGFL1</i>
6	145461097	rs17075195	C	T	0.239	-0.060	0.013	1.3E-06	<i>EPM2A</i>
7	33698582	rs73308535	T	C	0.010	-0.250	0.057	7.6E-06	<i>BBS9</i>
7	85128231	rs4537247	A	G	0.374	0.053	0.012	6.6E-06	<i>SEMA3D</i>
12	47865599	rs7965360	A	G	0.041	0.127	0.029	8.8E-06	<i>VDR</i>
12	96039301	rs7966262	T	A	0.377	0.060	0.012	3.9E-07	<i>LTA4H</i>
15	92647833	rs12050771	A	C	0.240	-0.060	0.014	1.5E-06	<i>FAM174B</i>
18	75090996	rs1876486	A	G	0.304	-0.060	0.012	2.4E-06	<i>ZNF407</i>

3.2 主成分分析多性狀全基因組關聯分析(PC-GWAS)

3.2.1 eGFR, eGFR slope, albuminuria 主成分分析

在台灣人體生物資料庫同時具有 eGFR, eGFR slope, albuminuria 這三個性狀資料的一共有 13,904 人，經由初步主成分分析可發現在 PCA 散佈圖上有糖尿病患者分佈明顯偏離(如附錄-圖 1)。以排除糖尿病的 10,887 位受試者(基本資料如表 6)進行 eGFR, eGFR slope, albuminuria 的主成分分析，可以發現前三個主成分(PC1-PC3) 可以解釋所有的變異量，如表 7 所示。表 8 顯示各主成分的負荷向量 (loading vector)。各個原始性狀在 PC1 和 PC2 的向量顯示如圖 5。

表 6 主成分分析多性狀全基因組關聯分析的參與者的基本資料

分析人數, N	10,887
年齡, year	48.9 ± 10.3
男性, N (%)	3,950 (36.3%)
UACR*, mg/g	8.384 (4.839-9.587)
eGFR, ml/min/1.73m ²	106.5 ± 12.2
eGFR slope, ml/min/1.73m ² /year	-0.73 ± 1.24

*expressed as median (Q1-Q3)

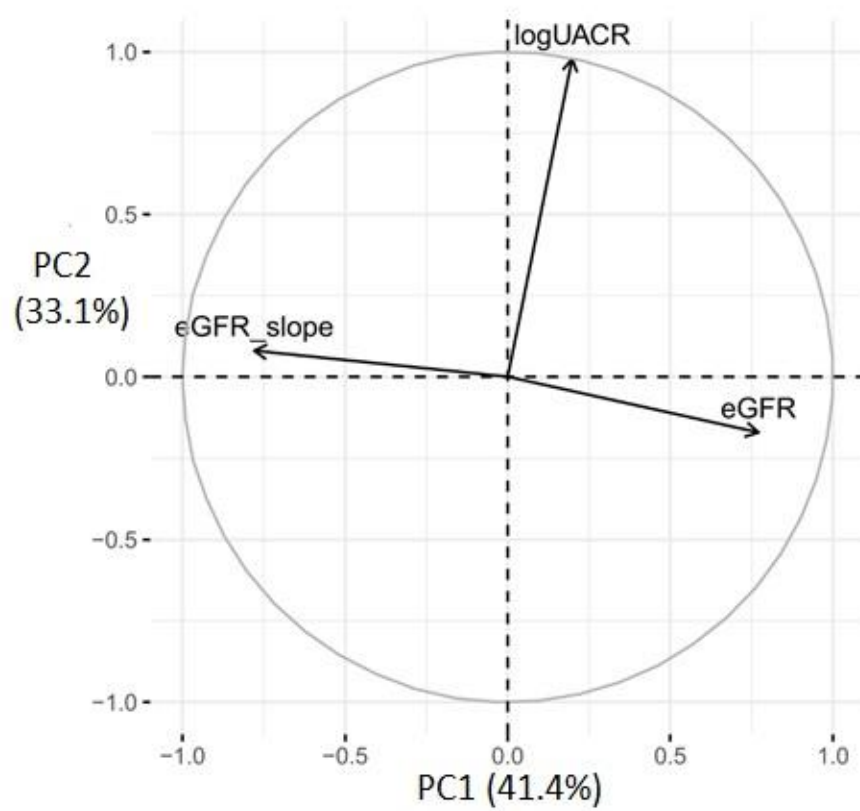
表 7 各主成分解釋 eGFR, eGFR slope, albuminuria 三項原始性狀之變異量

	解釋變異量 (%)	累積解釋變異量 (%)
PC1	41.4	41.4
PC2	33.1	74.5
PC3	25.5	100

表 8 各主成分由 eGFR, eGFR slope 和 albuminuria 的負荷向量組成

	PC1	PC2	PC3
logUACR	0.178286	0.981978	-0.06271
eGFR slope	-0.69961	0.081688	-0.70984
eGFR	0.691925	-0.17043	-0.70156

圖 5 原始性狀 eGFR, eGFR slope 和 albuminuria 在 PC1 和 PC2 的雙標圖



3.2.2 主成分進行 GWAS 之結果

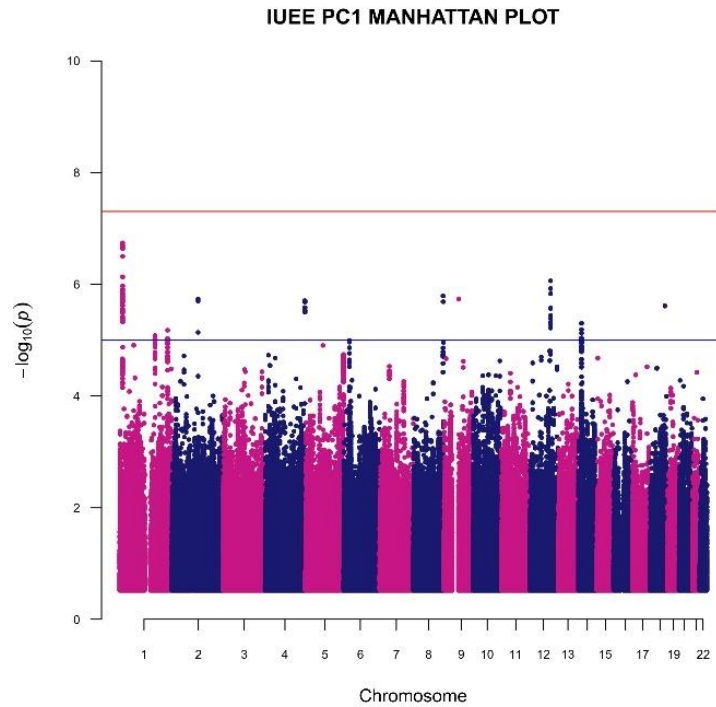
以三個主成分(PC1-PC3)分別進行 GWAS 的 manhattan plot 和 Q-Q plot 結果如圖 6-圖 8 所示。彙整顯著的結果如表 9 所示。





圖 6 主成分 PC1 GWAS 的(A)曼哈頓圖及(B) Q-Q 圖

(A)



(B)

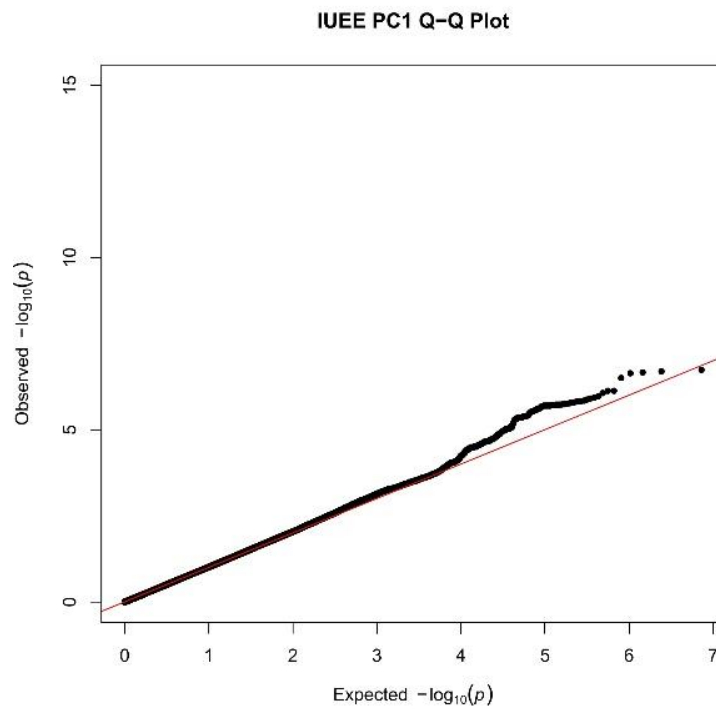
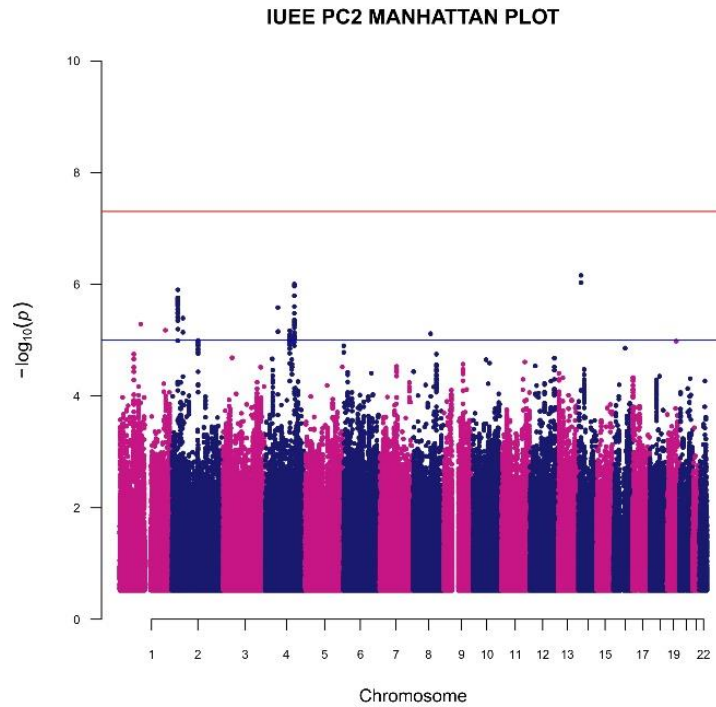




圖 7 主成分 PC2 GWAS 的(A)曼哈頓圖及(B) Q-Q 圖

(A)



(B)

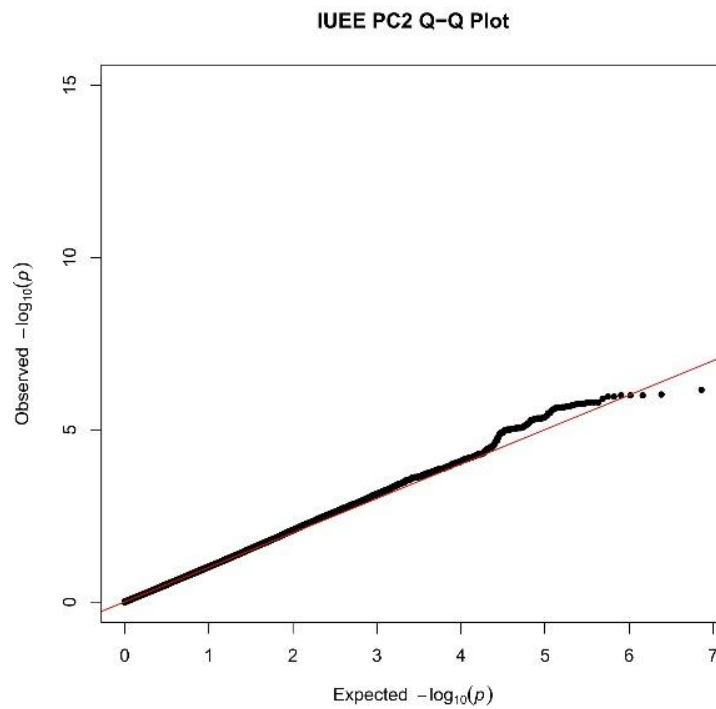
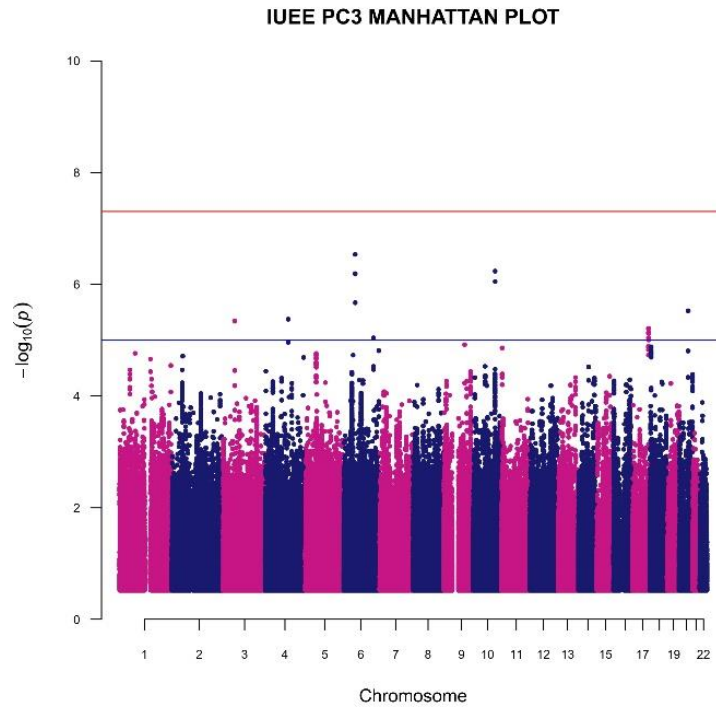


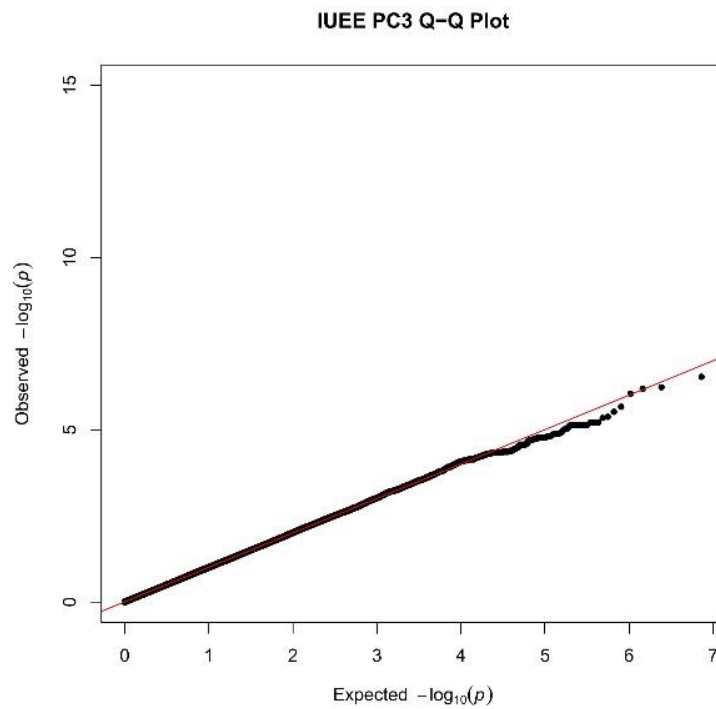


圖 8 主成分 PC3 GWAS 的(A)曼哈頓圖及(B) Q-Q 圖

(A)



(B)



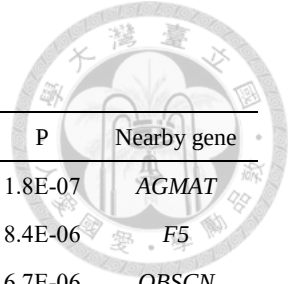


表 9 主成分分析多性狀全基因關聯分析顯著的 SNPs

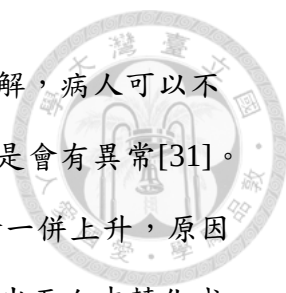
	Chr	BP	SNP	A1	A2	MAF	gnomAD	info score	P	Nearby gene
PC1	1	15586492	rs10159261	G	T	0.433	0.442	0.993	1.8E-07	AGMAT
	1	169524475	rs9332641	T	C	0.020	0.030	0.988	8.4E-06	F5
	1	228258445	rs10737429	A	G	0.434	0.416	0.995	6.7E-06	OBSCN
	2	123677383	rs10171475	C	T	0.227	0.211	0.985	1.8E-06	-
	4	188111676	rs67785655	T	C	0.188	0.180	0.976	2.0E-06	TRIML2
	8	140598947	rs4961274	T	C	0.087	0.085	0.972	1.6E-06	AGO2
	9	69531003	rs74445275	T	A	0.020	0.022	0.957	1.8E-06	APBA1
	12	97658889	rs4762397	T	C	0.399	0.401	0.994	8.6E-07	-
	14	34806498	rs7147621	A	G	0.217	0.217	0.997	5.0E-06	BAZ1A
	18	70501770	rs76982055	C	T	0.034	0.028	0.859	2.4E-06	-
PC2	1	101315274	rs77814797	A	G	0.056	0.079	0.971	5.1E-06	-
	1	218026147	rs193129422	G	T	0.011	0.015	0.936	6.7E-06	LOC105372922
	2	27512105	rs6547692	G	A	0.488	0.500	0.998	6.4E-06	GCKR
	2	27598615	rs12989678	T	C	0.465	0.486	0.996	1.3E-06	ZNF512
	2	51728241	rs190931819	G	C	0.051	0.046	0.945	4.1E-06	LOC730100
	4	60669935	rs56188710	T	C	0.022	0.021	0.930	2.6E-06	-
	4	116637209	rs10007743	G	A	0.384	0.393	1.000	6.7E-06	-
	4	126787643	rs139816356	C	T	0.022	0.021	0.999	8.7E-06	-
	4	138579033	rs1376095	A	G	0.209	0.218	0.998	1.0E-06	-
	4	138588684	rs10461249	C	G	0.361	0.354	0.998	7.3E-06	-
	8	82182829	rs79425178	T	C	0.031	0.030	0.938	7.7E-06	LOC105375930
	14	32571580	rs1155069	C	T	0.038	0.041	0.990	6.9E-07	AKAP6
PC3	3	54613722	rs76352818	T	C	0.022	0.027	1*	4.5E-06	CACNA2D3
	4	109504697	rs139161904	G	A	0.021	0.022	0.906	4.2E-06	SEC24B
	6	54486360	rs78828070	G	A	0.090	0.088	1*	6.5E-07	-
	6	54558447	rs146677015	G	C	0.015	0.011	0.910	2.9E-07	-
	6	141371205	rs117153205	G	A	0.069	0.061	1*	9.1E-06	-
	10	103868924	rs117231653	A	G	0.078	0.094	0.990	5.8E-07	-
	17	76196746	rs73360624	G	C	0.047	0.037	1.000	6.2E-06	RNF157
	20	42732473	rs8119626	T	C	0.013	0.016	0.982	3.0E-06	PTPRT

gnomAD, gnomAD v4.1 A1 allele frequency; info score, information score of imputed SNPs (*info score 1 denotes genotyped rather than imputed SNPs)

3.2.3 主成分分析多性狀全基因關聯分析(PC-GWAS)額外得到之位點

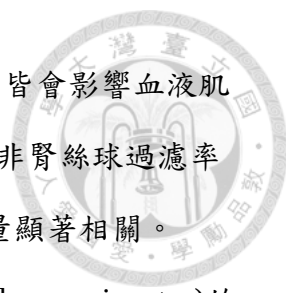
本研究的 PC-GWAS 得到 30 個位點。在原始單性狀 GWAS 以及 PC-GWAS 皆可找到的 10 個位點之中，有數個其鄰近的基因幾乎可以直接的認定和腎臟損傷或是疾病相關。AGMAT (Agmatinase)基因在不同類別的腎臟疾病其表現量顯著不同[23]。GCKR (Glucokinase Regulator)基因之多形性已經被確立和慢性腎臟病，末期腎臟疾病，以及血清尿酸值相關[24-26]。ZNF512 (Zinc Finger Protein 512)基因被發現和高尿酸血症以及痛風相關[27]。AKAP6 (A-Kinase Anchoring Protein 6)基因則已知和腎臟結石相關[28]。

有 20 個是相對於在原始單性狀 GWAS 額外得到的位點，而且也未曾出現在之前已經發表過的 albuminuria, cross-sectional eGFR or eGFR slope GWAS (表 10)。這 20 個新風險位點之區域圖(regional plots)如圖 9。除去 10 個 intergenic SNPs 之後，餘下的 10 個新位點之中有鄰近位於第一號染色體的 *F5* 基因(rs9332641; $P = 8.39 \times 10^{-6}$)，位於第一號染色體的 *OBSCN* 基因(rs10737429; $P = 6.66 \times 10^{-6}$)，位於第三號染色體的 *CACNA2D3* 基因(rs76352818; $P = 4.54 \times 10^{-6}$)，位於第四號染色體的 *SEC24B* 基因(rs139161904; $P = 4.21 \times 10^{-6}$)，位於第四號染色體的 *TRIML2* 基因(rs67785655; $P = 1.96 \times 10^{-6}$)，位於第八號染色體的 *AGO2* 基因(rs4961274; $P = 1.62 \times 10^{-6}$)，位於第九號染色體的 *APBA1* 基因(rs74445275; $P = 1.84 \times 10^{-6}$)，位於第十四號染色體的 *BAZ1A* 基因(rs7147621; $P = 5.02 \times 10^{-6}$)，位於第十七號染色體的 *RNF157* 基因(rs73360624; $P = 6.19 \times 10^{-6}$)以及位於第二十號染色體的 *PTPRP* 基因(rs8119626; $P = 2.99 \times 10^{-6}$)。藉由文獻回顧，可以發現一些以上位點相鄰的基因和腎臟損傷以及慢性腎臟病相關。*F5* 基因的產物是第五凝血因子，其在諸多的情境包括急性腎損傷時候會由巨噬細胞產生，所以和腎臟纖維化相關[29]。*OBSCN* 基因的產物 obscurin 被認為和肌漿網(sarcoplasmic reticulum)功能以及處理胞內鈣離子濃度有關，而 *OBSCN* 雙套對偶基因皆有功能缺失的突變時，這樣的人會重複發生橫紋肌溶解[30]。橫紋肌溶



解的臨床嚴重程度可以差異很大，若是發生亞臨床的橫紋肌溶解，病人可以不具典型症狀如肌肉痠痛、無力或是有深色尿液，但抽血檢驗還是會有異常[31]。除了肌酸激酶(creatine kinase)上升之外，肌酸酐(creatinine)亦會一併上升，原因部分為肌酸(creatine)和磷酸肌酸(phosphocreatine)從肌肉細胞釋出至血中轉化成肌酸酐，但是更大的原因為肌球蛋白(myoglobin)大量釋出由循環系統至腎臟造成腎小管壞死(acute tubular necrosis) [32]。腎小管壞死為急性腎損傷(acute kidney injury)常見原因，而為造成慢性腎臟疾病最重要因素之一[33]。AGO2 (Argonaute RISC Catalytic Component 2) 的產物是一種 RNA 結合蛋白，它藉由同時結合微小核糖核酸(microRNA)和目標的信使核糖核酸(messenger RNA)而聚合成 RNA 誘導沉默複合物 (RNA-induced silencing complex; RISC)。miR-429-3p 是一種微小核糖核酸，它會減少支鏈胺基酸在近端腎小管的分解代謝而導致鐵依賴性細胞死亡(ferroptosis) [34]；因此 AGO2 和腎損傷以及慢性腎臟病可能有相關。

在這些新位點之中有 4 個在 nephQTL[35]有顯著結果。rs9332641 顯著地影響 *SELL* 和 *ATP1B1* 的表現量。*SELL* 蛋白和白血球和活化內皮細胞之間初期的繫鏈(tethering)相關所以能夠調節不同器官的發炎反應，包括胰臟和腎臟[36]。*ATP1B1* 是鈉鉀幫浦(Na^+/K^+ -ATPase)的 $\beta 1$ 次單位，而鈉鉀幫浦是腎小管最重要的蛋白之一，為負責大部分腎小管在吸收鈉以及其他電解質和小分子的幫浦。miR-192-5p 是腎臟內含量最豐富的小分子核糖核酸(microRNA; miRNA)之一，其將 *ATP1B1* 的 mRNA 作為目標；將腎內 miR-192-5p knockdown 的大鼠再進行 *ATP1B1* 基因 knockdown，可以減少這些大鼠鈉鹽敏感性高血壓(salt-sensitive hypertension)的情形[37]。腎臟發炎反應愈久，則顯著造成腎臟纖維化以及慢性腎臟病[38]，而高血壓為慢性腎臟病及末期腎臟疾病的一大風險因子[39]；rs9332641 影響的兩個基因，從功能上來看和慢性腎臟病皆有關聯。rs7147621 和 *CFL2* 基因的表現顯著有關。*Cofilin-2* (*CFL2* 基因的產物) 是可以和肌動蛋白(actin)結合的一種蛋白，而 *CFL2* 突變會造成線狀體肌肉病變 (nemaline rod



myopathy) [40]。營養不良、肌肉病變等使肌肉量不足的情形，皆會影響血液肌酸酐的濃度，所以這樣的關聯可能是由於肌肉釋出量的改變而非腎絲球過濾率改變所致 [41]。rs73360624 和 *ST6GALNAC2* 及 *CYGB* 的表現量顯著相關。*ST6GALNAC2* 是一種可以將唾液酸(sialic acid)加入糖綴合物(glycoconjugates)的酵素，其和 IgA1 分子的唾液酸化(sialylation)有關；IgA1 分子的去唾液酸化(desialylation)是 A 型免疫球蛋白腎病變(IgA nephropathy; IgAN)的致病機轉之一，而有研究發現 *ST6GALNAC2* 啟動子(promoter)區域單套型(haplotype)為 ADG 的頻率在 IgAN 病人顯著較高[42]。IgAN 是全世界最常見的原發性腎絲球疾病，為慢性腎臟病及末期腎臟疾病的一個重要原因[43]。*CYGB* 在人類腎臟足狀細胞表現很豐富，而在 *Cygb* 基因剔除小鼠可以發現足狀細胞減少和腎功能下降且兩者呈正相關；分析 *CYGB*-dependent 的轉錄組(transcriptome) 發現到和氧化還原平衡、細胞凋亡以及慢性腎臟病相關的許多基因調控異常[44]。此外，有個 intergenic SNP rs117231653 和 *COL17A1* 在全血組織的表現量有顯著相關(GTEx [45])。*COL17A1* 為膠原蛋白 XVII 的 $\alpha 1$ 鍊。膠原蛋白 XVII 存在於人類以及小鼠腎臟的足狀細胞和內皮細胞，用電子顯微鏡檢可以更精細的確認它表現於足狀細胞的足突以及腎絲球基底膜，所以其可能和足狀細胞的成熟以及腎絲球過濾有關[46]。

表 10 主成分分析多性狀全基因關聯分析得到的 novel loci 與 eQTL 結果

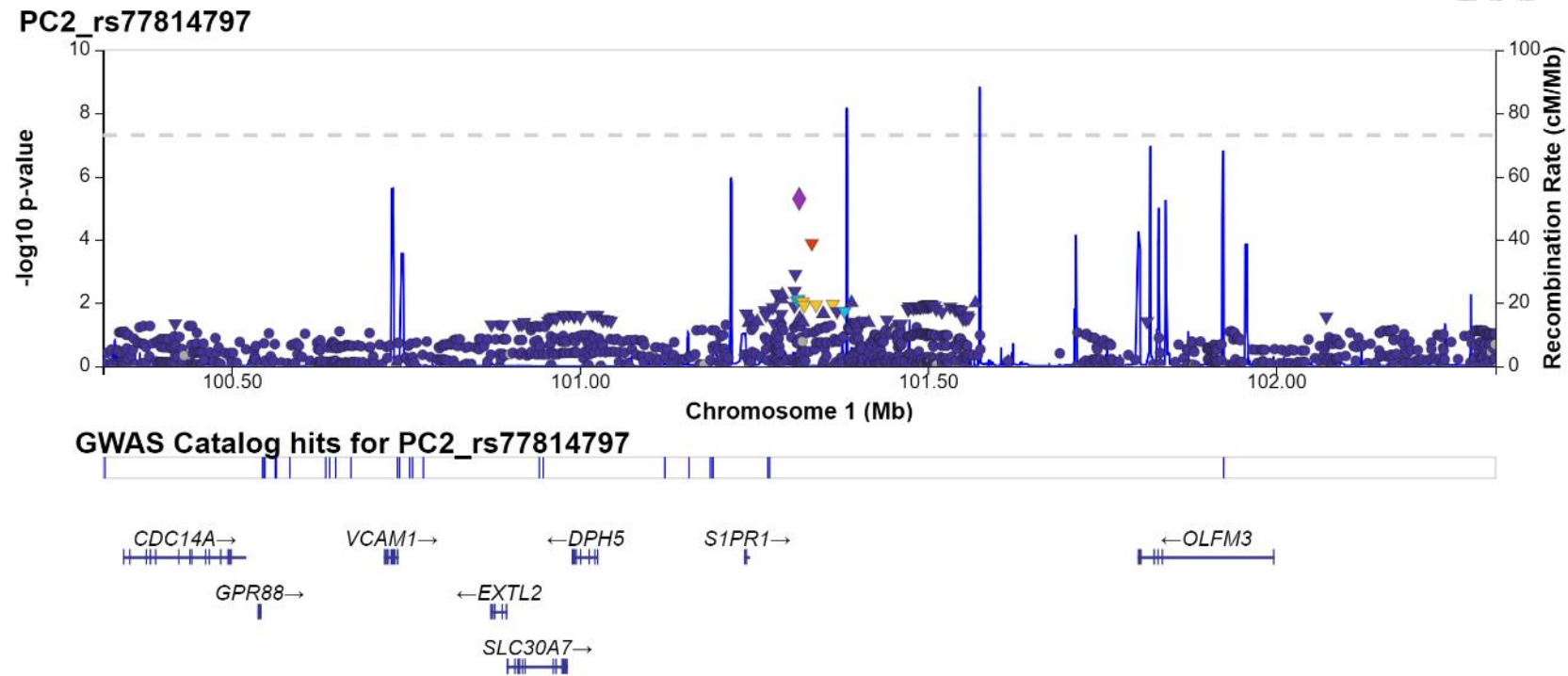
Chr	BP	SNP	A1	A2	MAF	P	Nearby gene	Involved gene expression in kidney tissue
1	101315274	rs77814797	A	G	0.056	5.14E-06	-	
1	169524475	rs9332641	T	C	0.02	8.39E-06	<i>F5</i>	<i>SELL</i> (glo); <i>ATP1B1</i> (ti)
1	228258445	rs10737429	A	G	0.434	6.66E-06	<i>OBSCN</i>	<i>TRIM17</i> , <i>OBSCN-AS1</i> , <i>IBA57-AS1</i> , <i>MRPL55</i> (ti)
2	123677383	rs10171475	C	T	0.227	1.82E-06	-	
3	54613722	rs76352818	T	C	0.022	4.54E-06	<i>CACNA2D3</i>	
4	60669935	rs56188710	T	C	0.022	2.61E-06	-	
4	109504697	rs139161904	G	A	0.021	4.21E-06	<i>SEC24B</i>	
4	116637209	rs10007743	G	A	0.384	6.73E-06	-	
4	126787643	rs139816356	C	T	0.022	8.72E-06	-	
4	138588684	rs10461249	C	G	0.361	7.29E-06	-	
4	188111676	rs67785655	T	C	0.188	1.96E-06	<i>TRIML2</i>	
6	141371205	rs117153205	G	A	0.069	9.14E-06	-	
8	140598947	rs4961274	T	C	0.087	1.62E-06	<i>AGO2</i>	
9	69531003	rs74445275	T	A	0.02	1.84E-06	<i>APBA1</i>	
10	103868924	rs117231653	A	G	0.078	5.82E-07	-	
12	97658889	rs4762397	T	C	0.399	8.64E-07	-	
14	34806498	rs7147621	A	G	0.217	5.02E-06	<i>BAZ1A</i>	<i>CFL2</i> (ti)
17	76196746	rs73360624	G	C	0.047	6.19E-06	<i>RNF157</i>	<i>ST6GALNAC2</i> , <i>UBE2O</i> (glo); <i>CYGB</i> (ti)
18	70501770	rs76982055	C	T	0.034	2.43E-06	-	
20	42732473	rs8119626	T	C	0.013	2.99E-06	<i>PTPRT</i>	

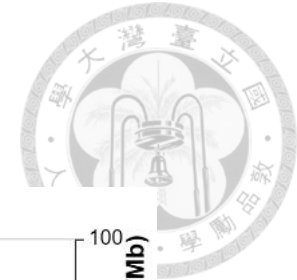
* glo, glomerulus; ti, tubulointerstitium



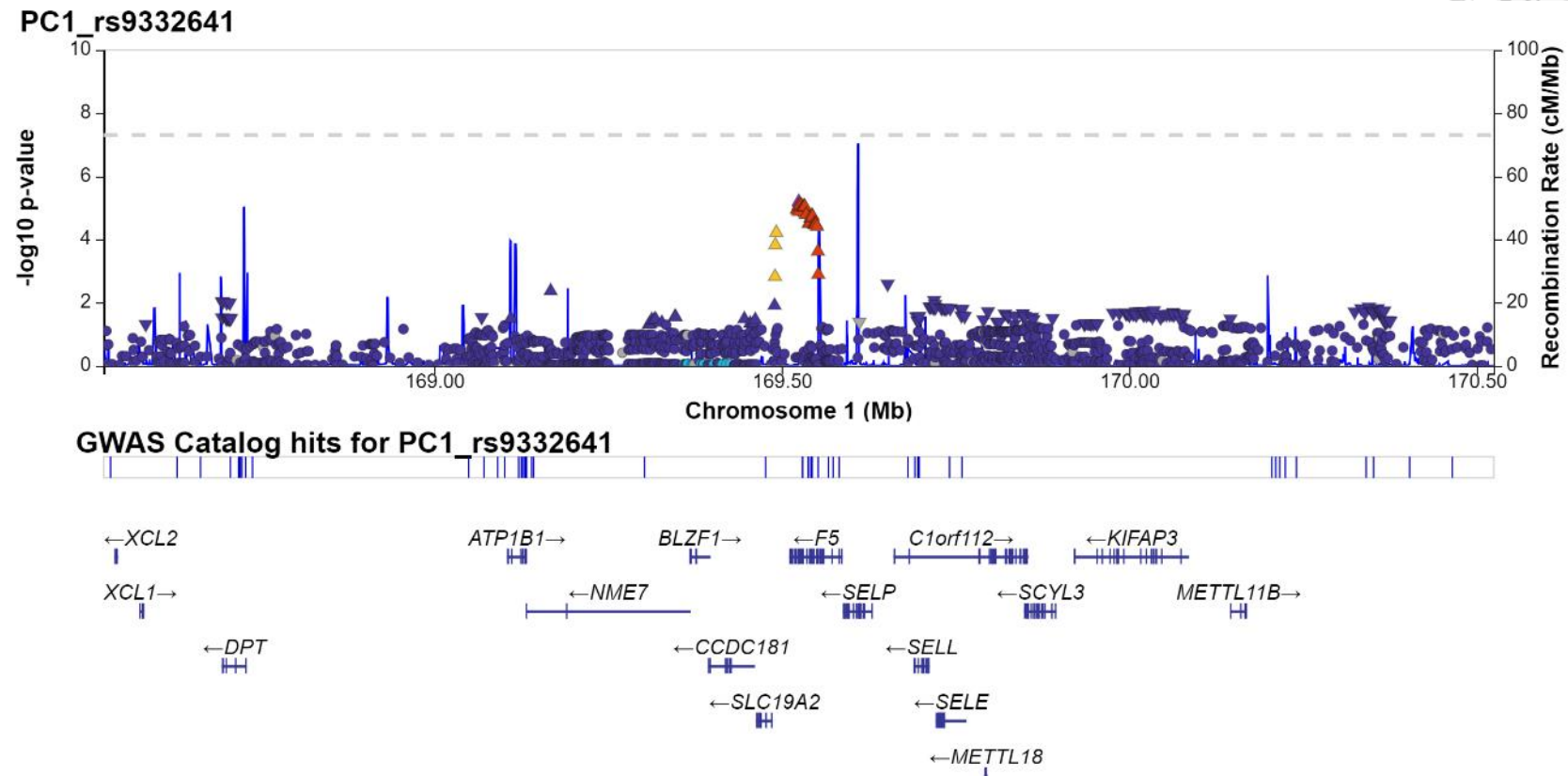
圖 9 主成分 GWAS 得到之新風險位點的區域圖

(A) rs77814797

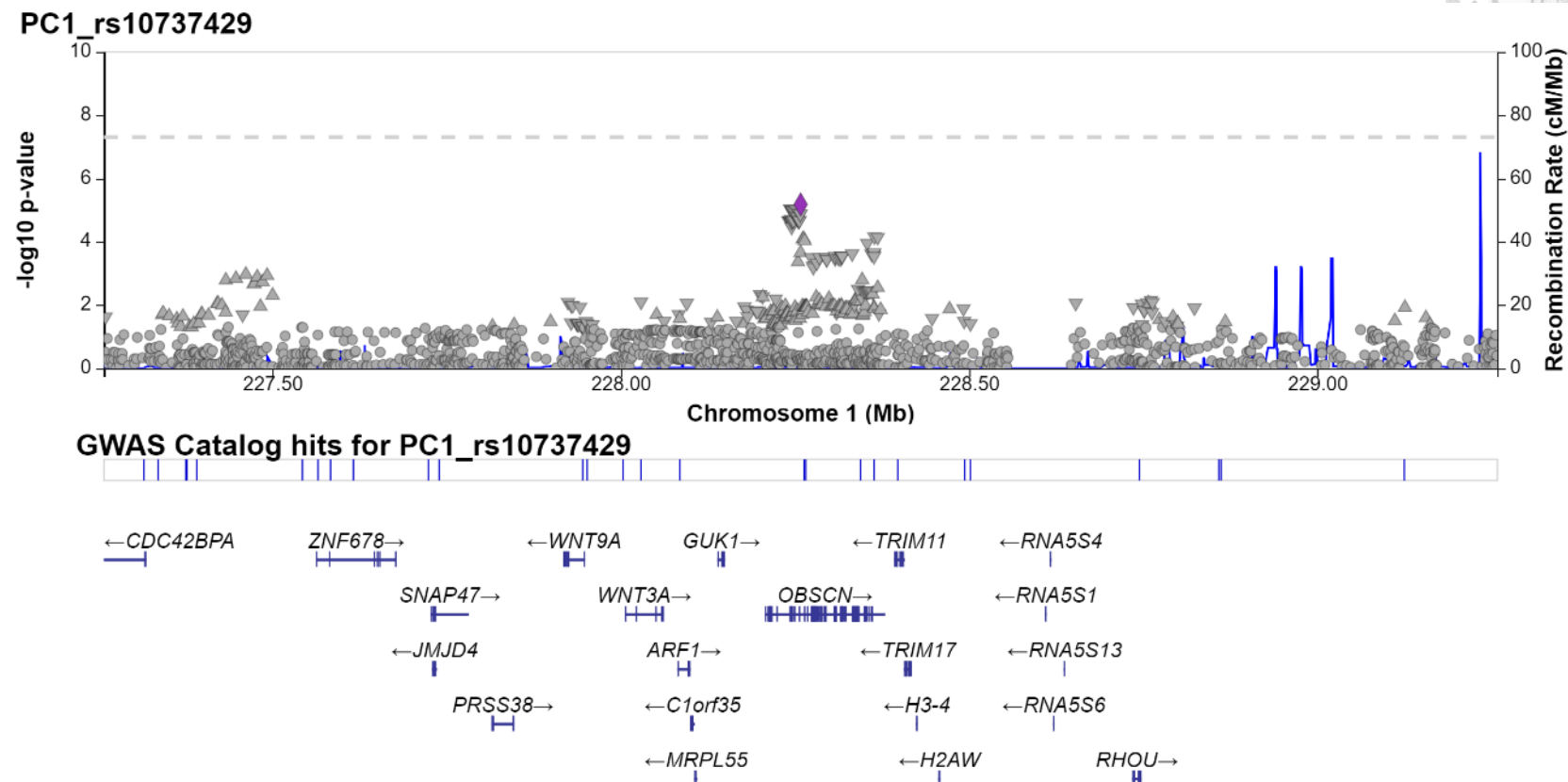




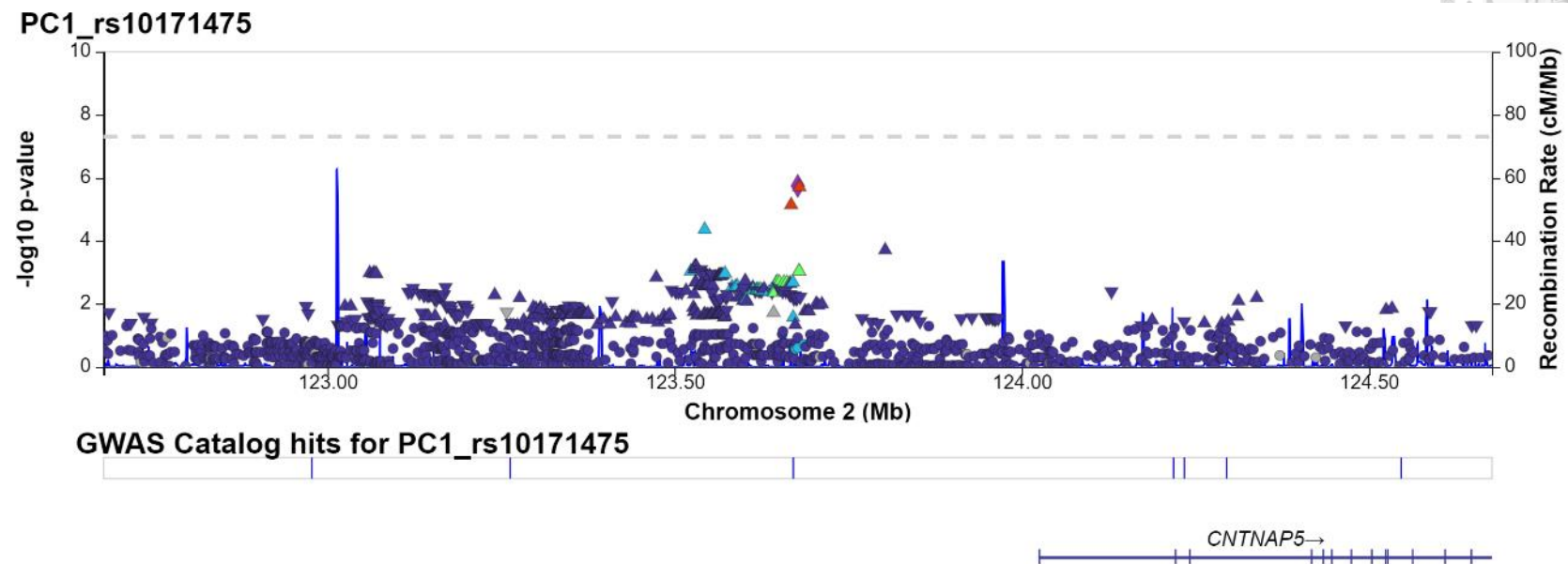
(B) rs9332641



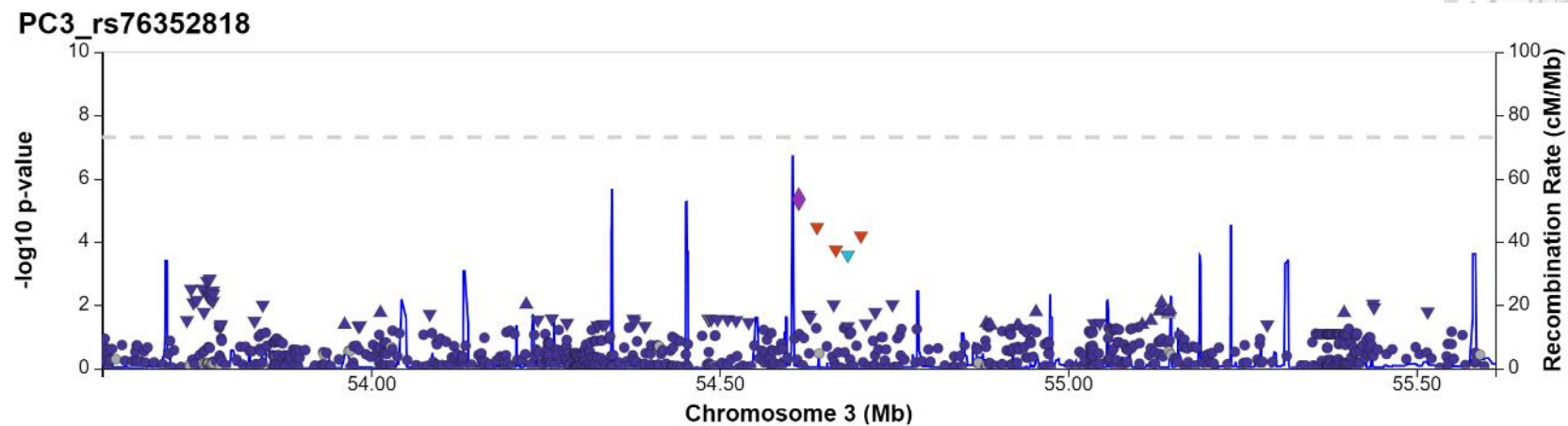
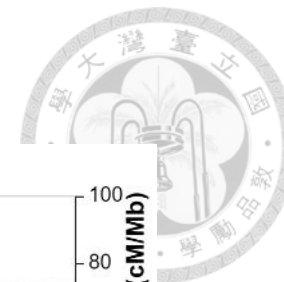
(C) rs10737429



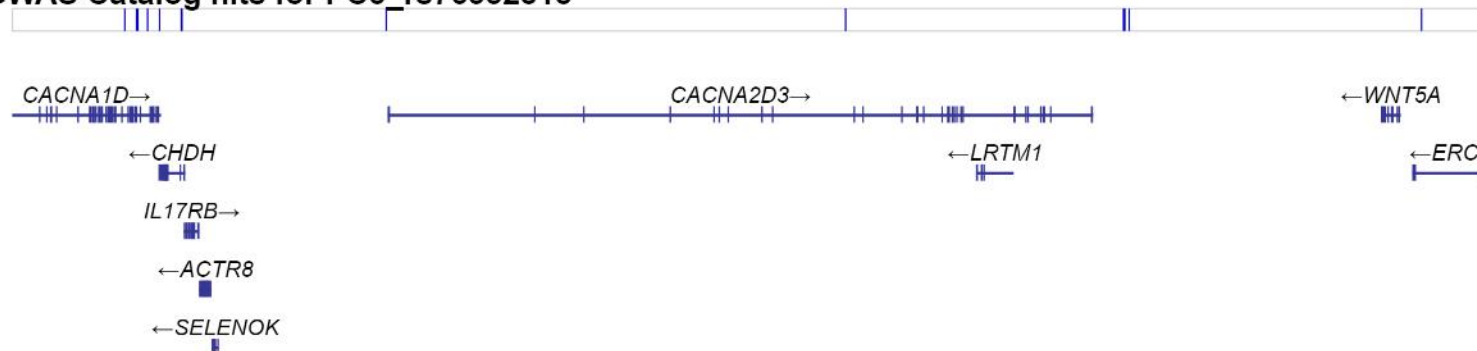
(D) rs10171475



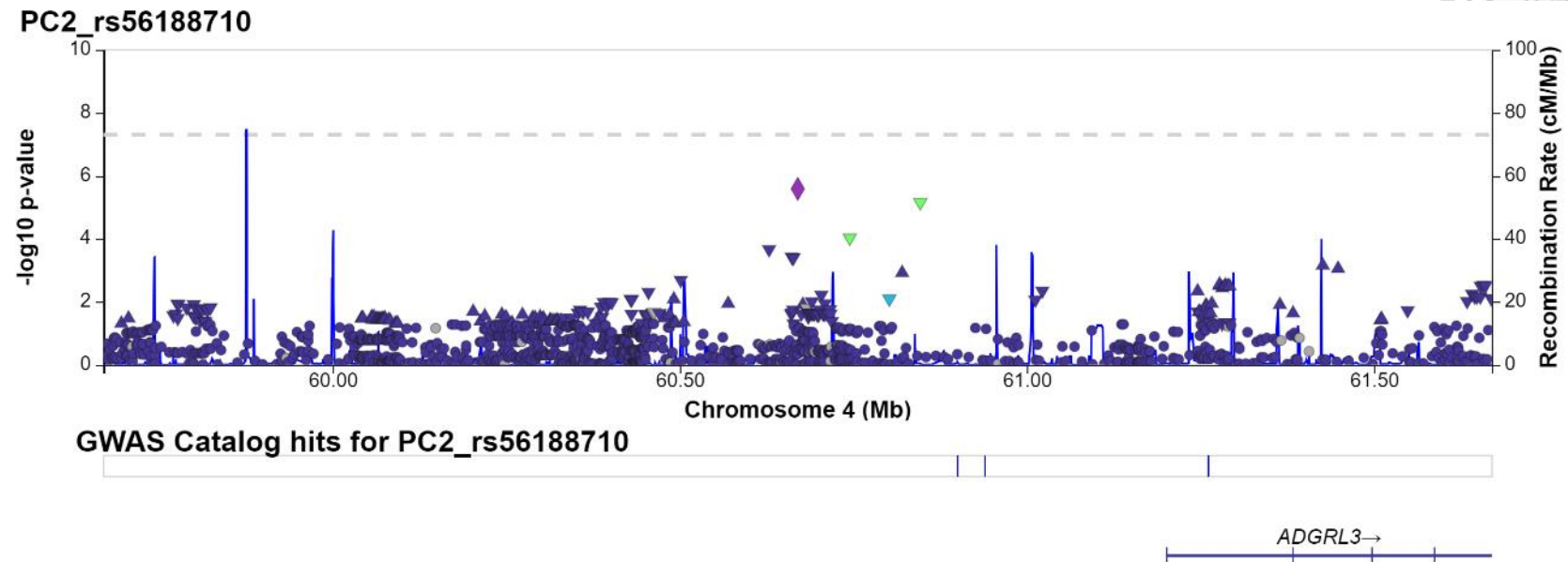
(E) rs76352818



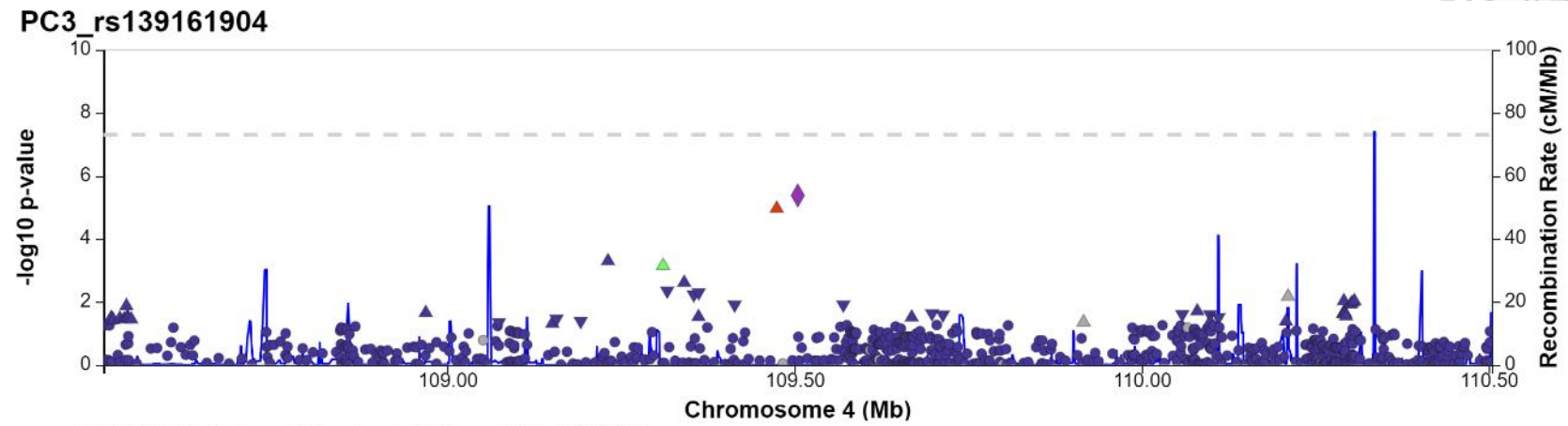
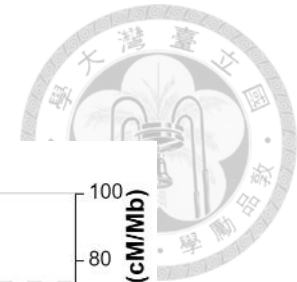
GWAS Catalog hits for PC3_rs76352818



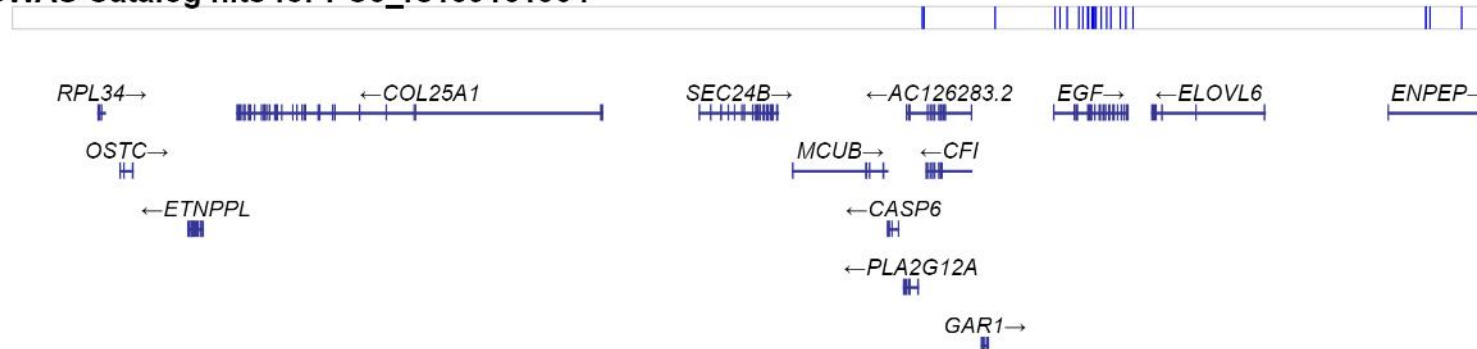
(F) rs56188710



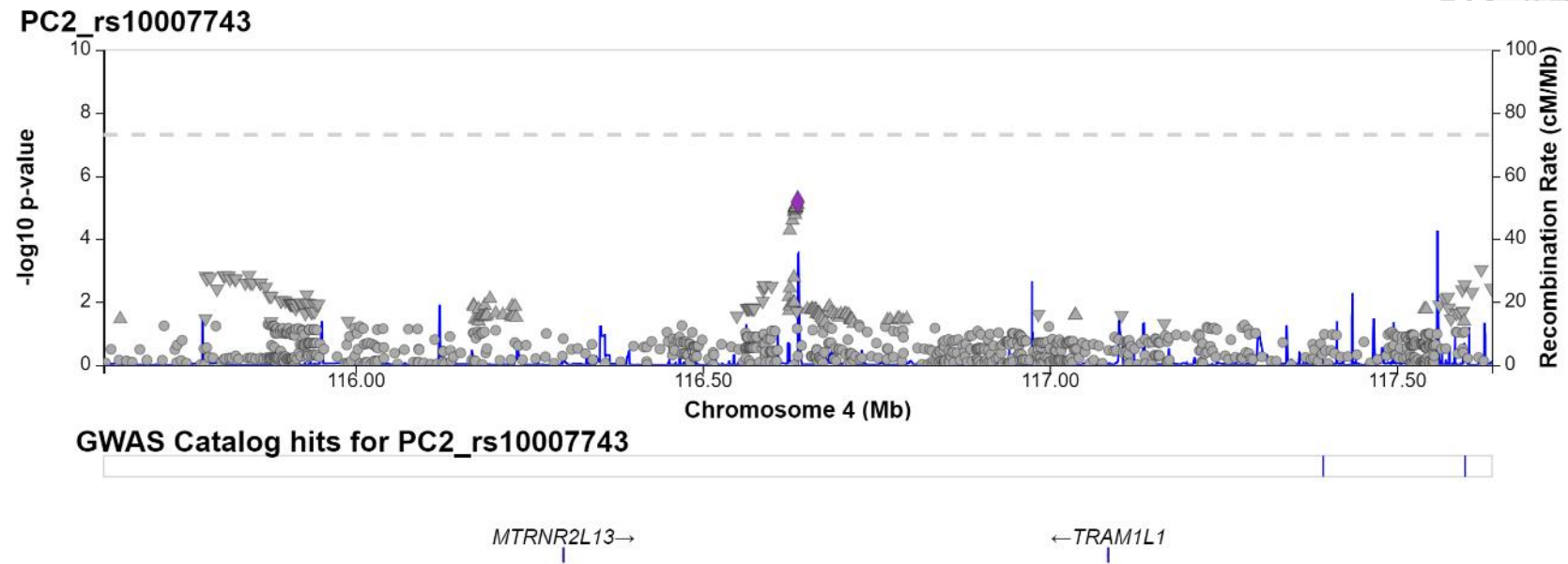
(G) rs139161904



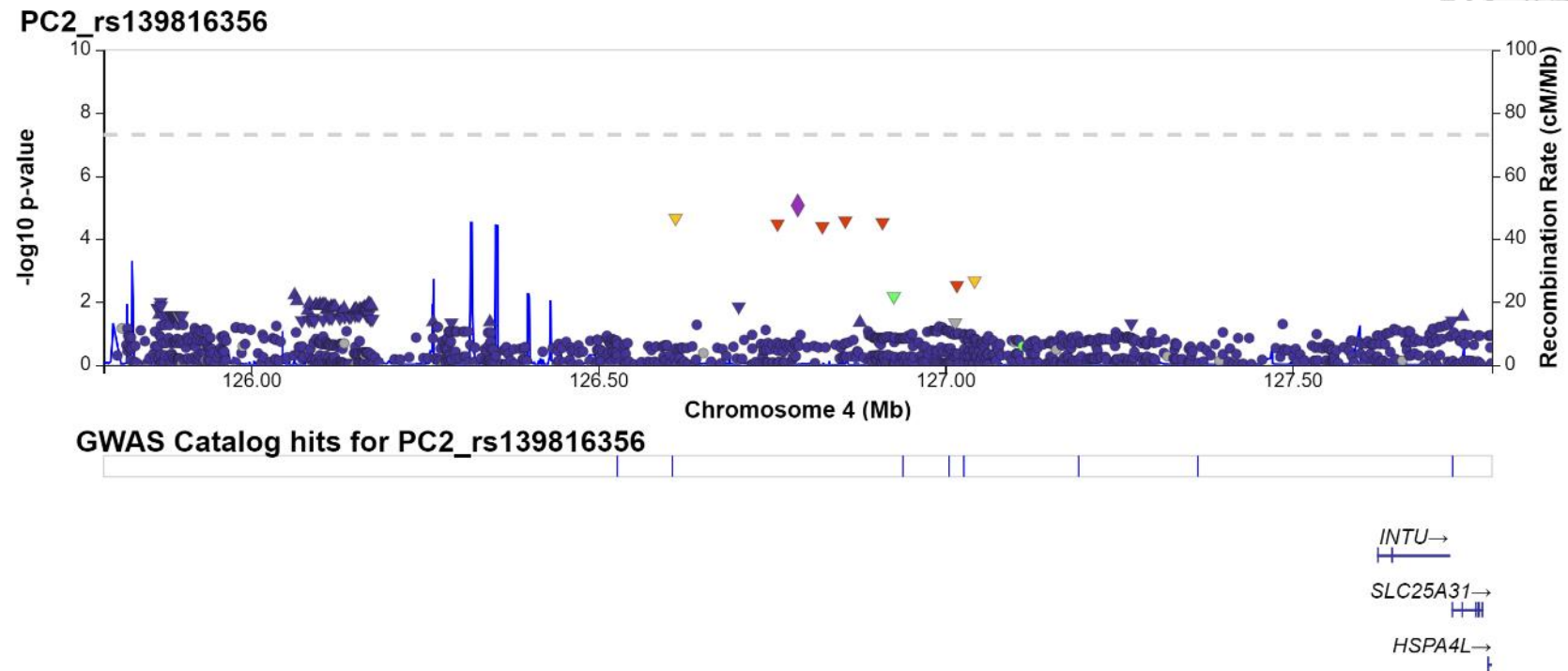
GWAS Catalog hits for PC3_rs139161904



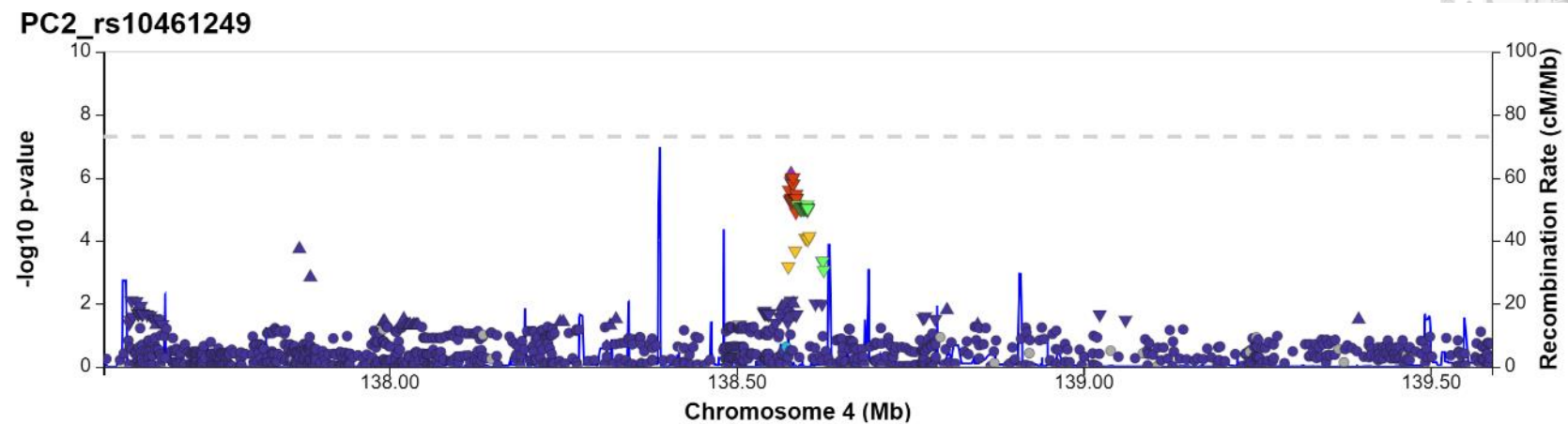
(H) rs10007743



(I) rs139816356



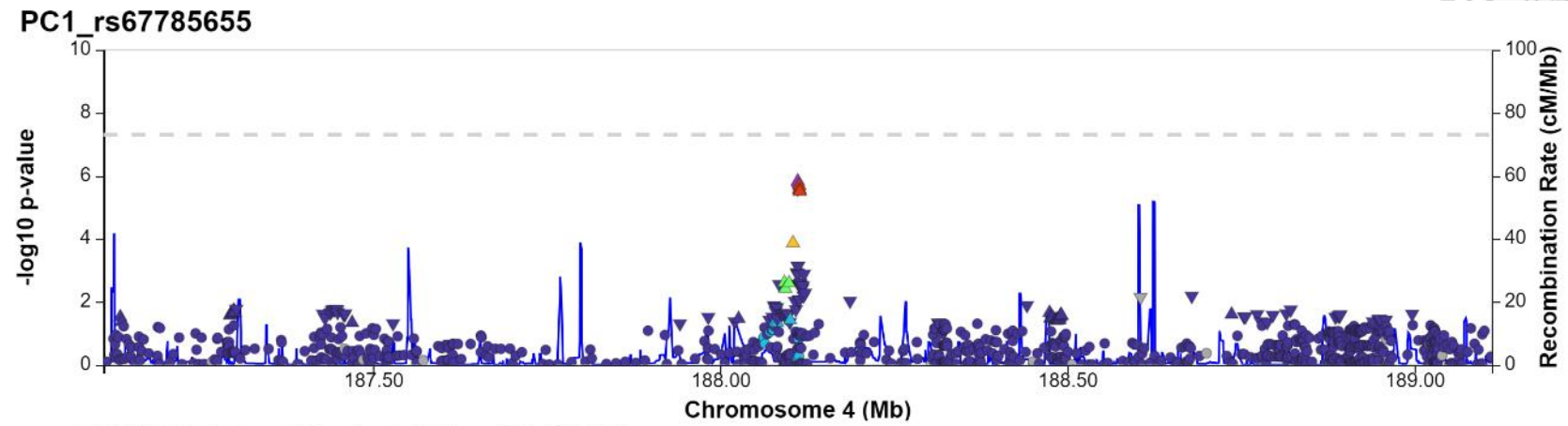
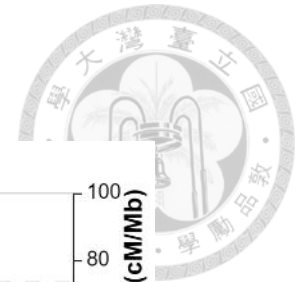
(J) rs10461249



GWAS Catalog hits for PC2_rs10461249



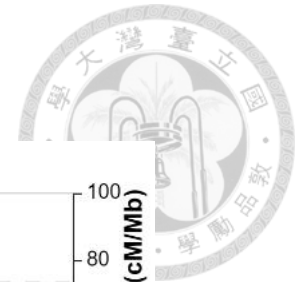
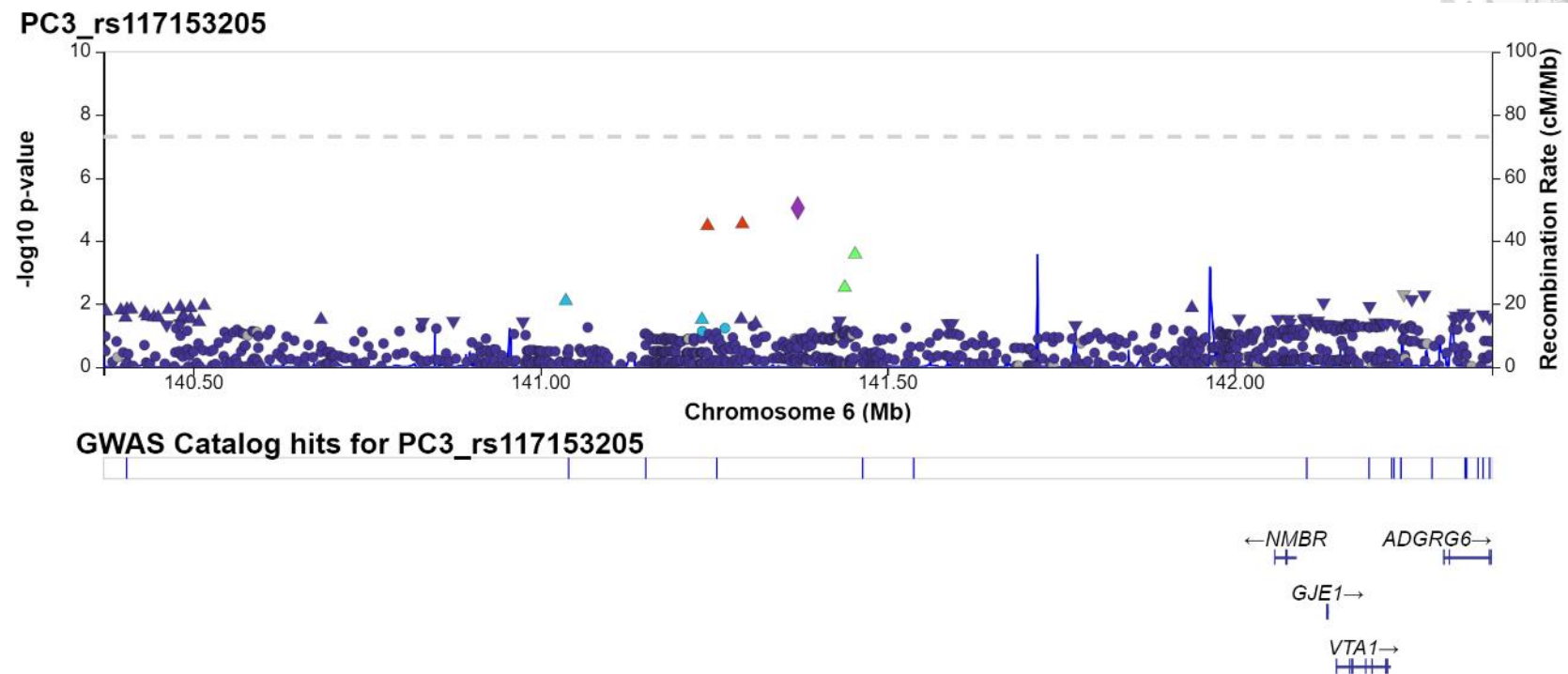
(K) rs67785655



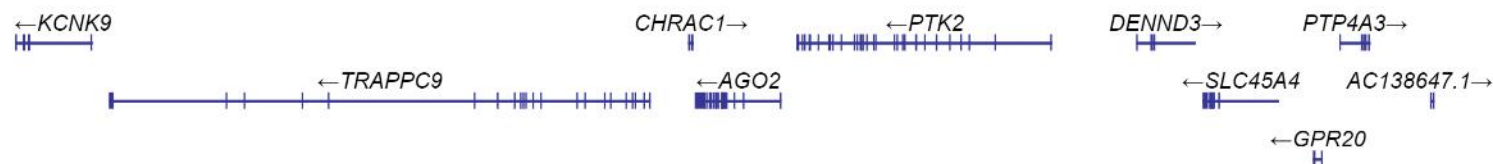
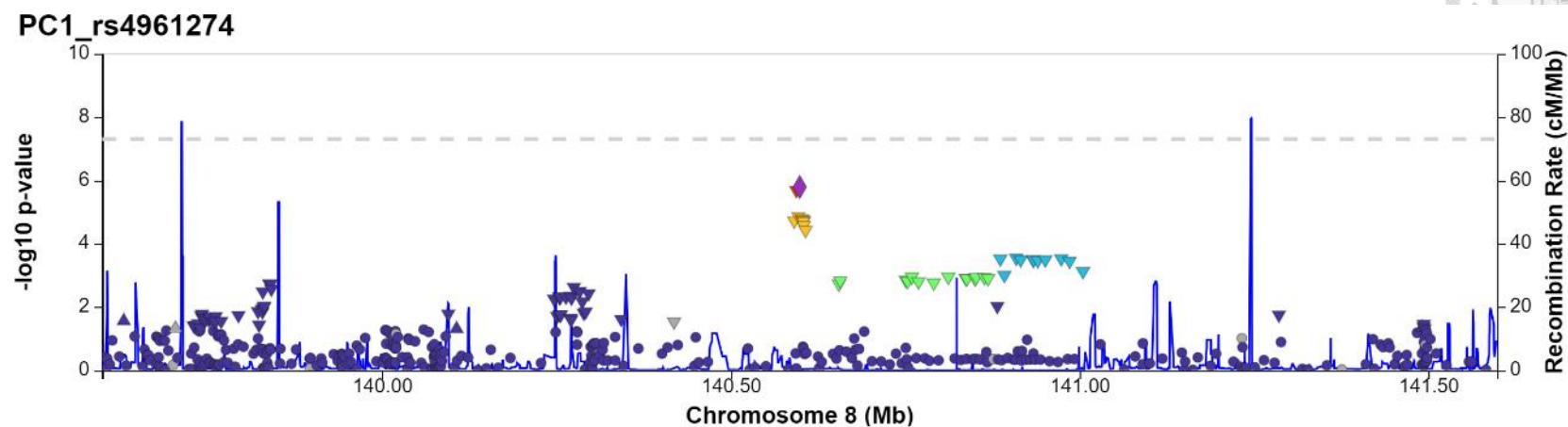
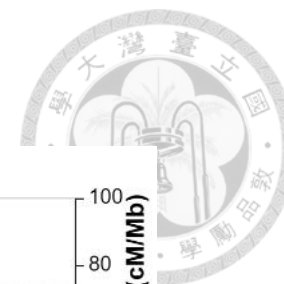
GWAS Catalog hits for PC1_rs67785655

ZFP42→
←TRIML2
TRIML1→

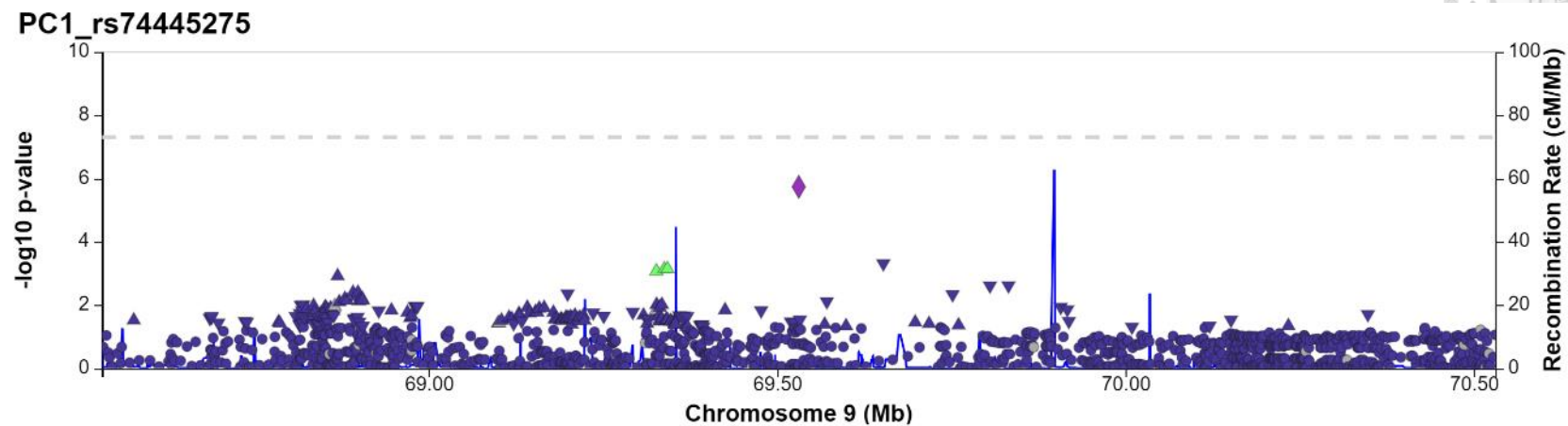
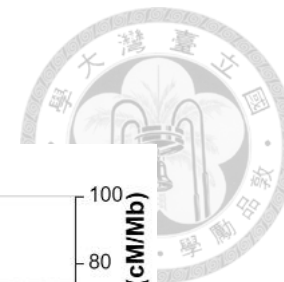
(L) rs117153205



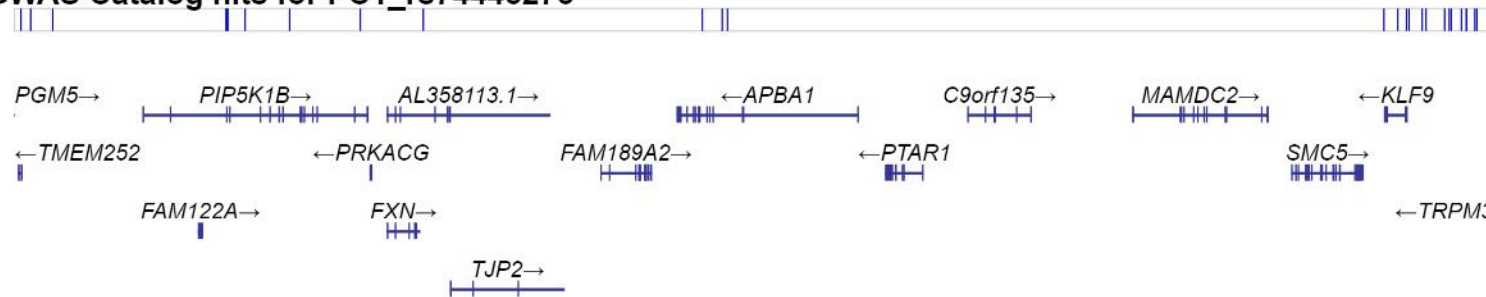
(M) rs4961274



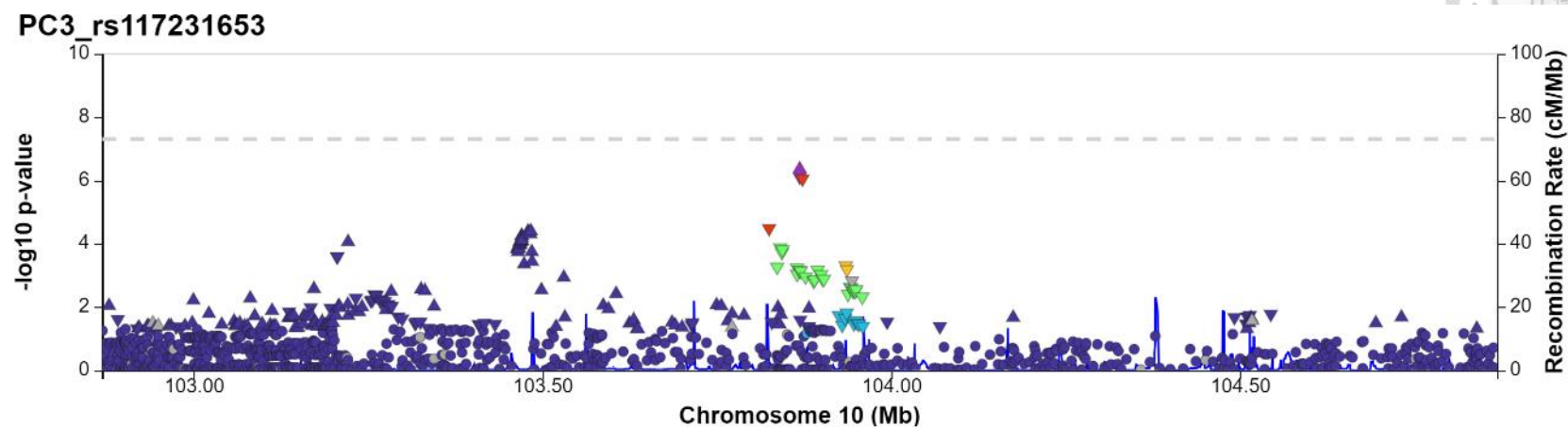
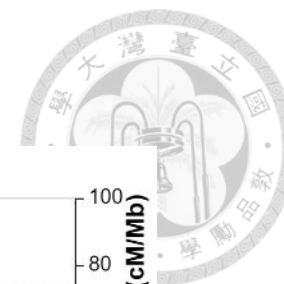
(N) rs74445275



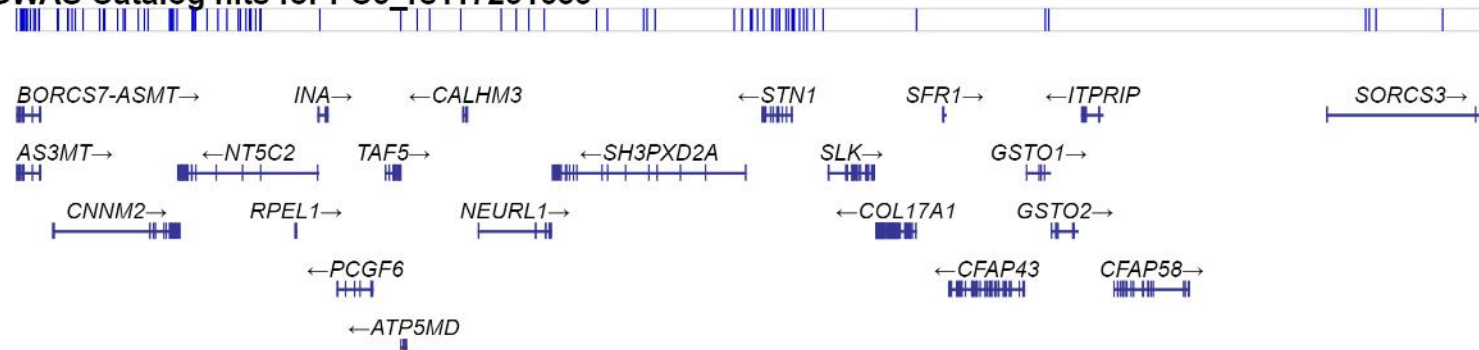
GWAS Catalog hits for PC1_rs74445275



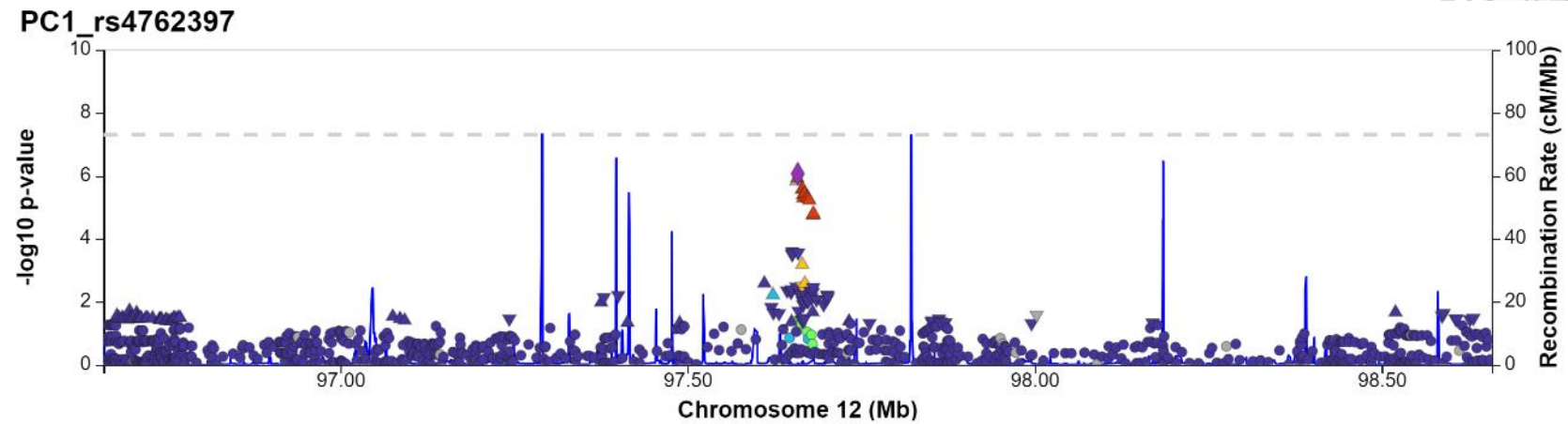
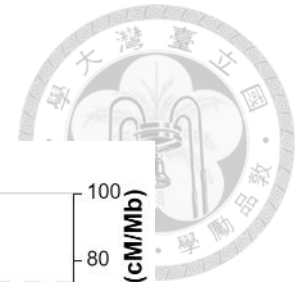
(O) rs117231653



GWAS Catalog hits for PC3_rs117231653



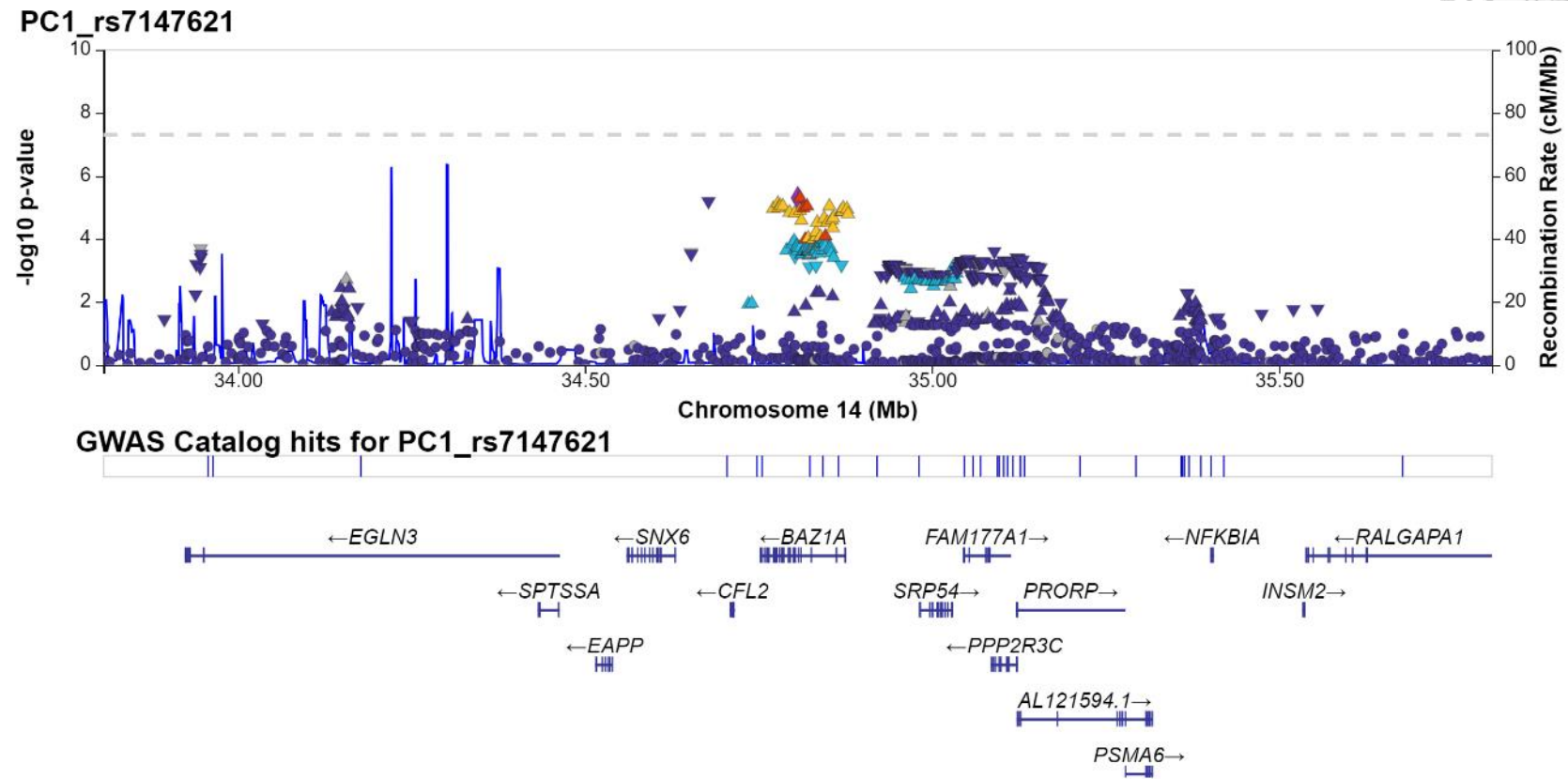
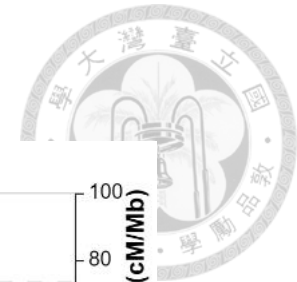
(P) rs4762397



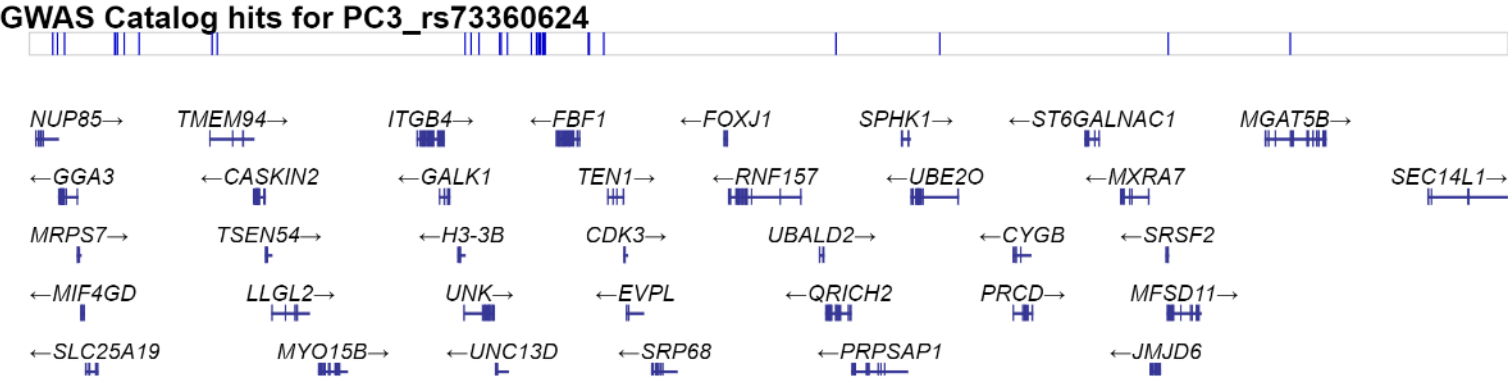
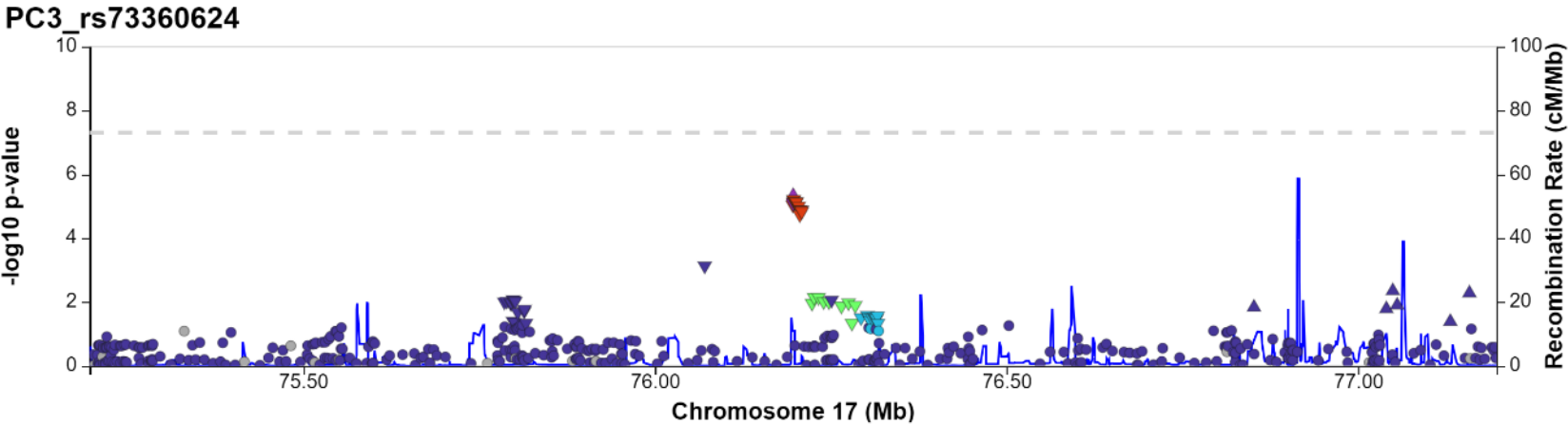
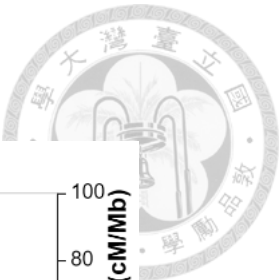
GWAS Catalog hits for PC1_rs4762397



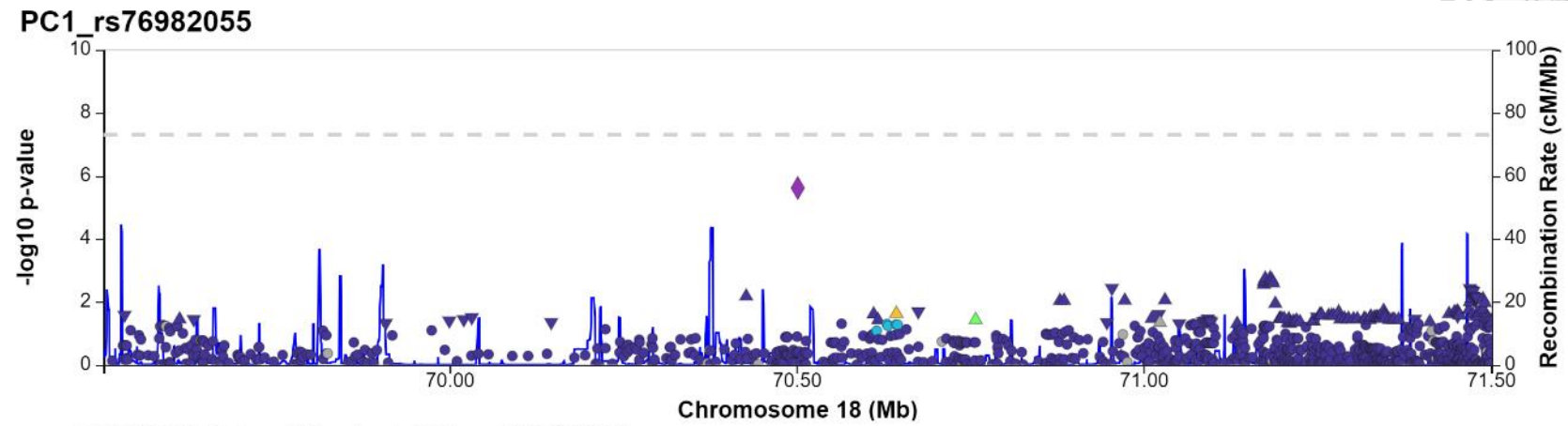
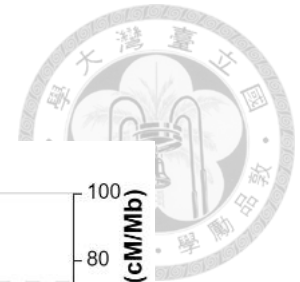
(Q) rs7147621



(R) rs73360624



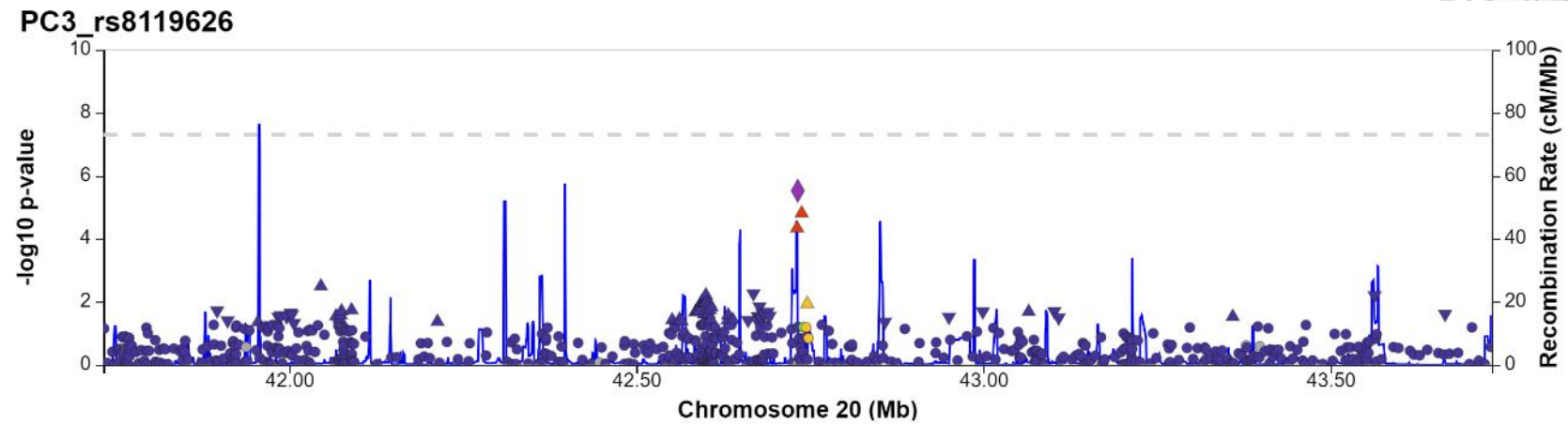
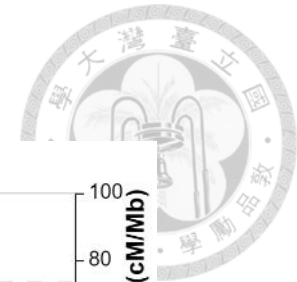
(S) rs76982055



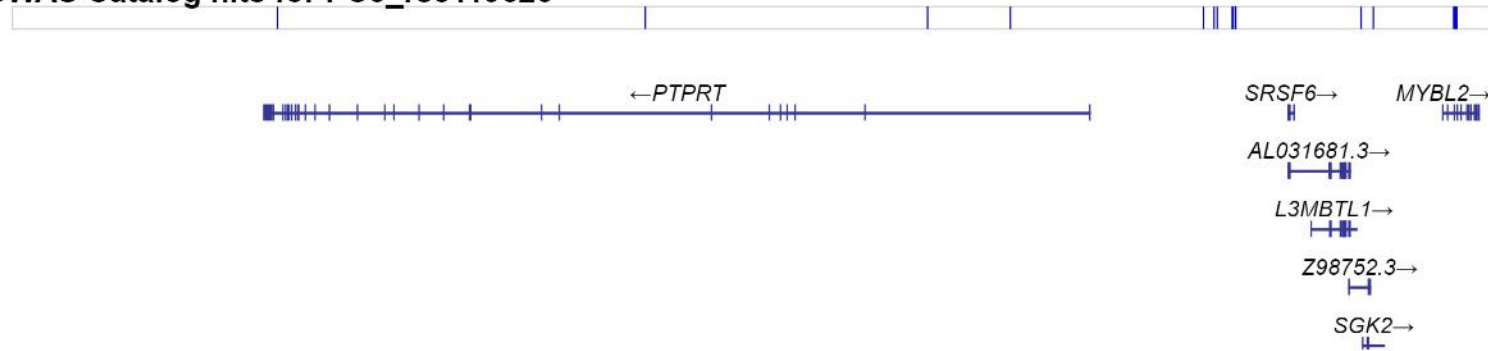
GWAS Catalog hits for PC1_rs76982055



(T) rs8119626



GWAS Catalog hits for PC3_rs8119626



第四章 討論與結論



4.1 結果討論

本研究對於 albuminuria、eGFR 以及 eGFR slope 這三個腎臟性狀，第一次利用 PC-GWAS 的方式發現了數個新的和腎功能變化/慢性腎臟病潛在相關的風險位點。藉由文獻回顧的方式，在功能上的探究可以找到以上位點的一部份其相鄰的基因和腎功能、腎病變以及慢性腎臟病相關。

一般來說在 GWAS 提到主成分分析，主要是對於基因型的應用來檢視是否有族群分層的問題(population stratification)，若有相當的程度則在最後相關分析的步驟時會將影響最大的前幾個基因型資料的 PC 拿來做共變項校正，如此可以去除來自族群分層造成的差異[47]。利用主成分分析應用於欲分析的性狀資料來進行 GWAS 則十分少見，因為欲研究的性狀必需是數量性狀，而且需要一群人同時擁有所有的性狀。PC-GWAS 優點為可以偵測到個別原始單性狀 GWAS 無法找到的位點；反之，個別原始單性狀 GWAS 找到的位點之中，可能許多只是和單面向的性狀相關但是並沒有和真正造成的疾病有關，所以等於是具有統計上的偽陽性 type 1 error [48]，而 PC-GWAS 找到的位點因為影響到各個原始性狀，比較具有代表性。

目前文獻應用 PC-GWAS 研究慢性腎臟病新風險位點的文章只有一篇由 Tran 等人著作[49]。Tran 等人的研究選取的原始主要性狀為血液肌酸酐、腎絲球過濾率預估值以及血液尿素氮；腎絲球過濾率預估值本身即為血液肌酸酐值的應變數，基本上為倒數相關，所以兩個性狀基本上算是重疊，分析時會有嚴重的共線性問題，而血液尿素氮對於腎功能的評估則是不具特異性的，常見腎功能正常者血液尿素氮也會高，因為其單純是反應體內水分不夠或是使用高蛋白飲食等等。基於上述原因，從原始性狀的選取不盡理想，其分析所得之結果則參考價值不大。



在很多 GWAS 會有遺傳性缺失(missing heritability)的議題[50]；藉由許多不同疾病和複雜性狀的研究，我們可以發現特定種族低頻率的變異點也佔有很重要的比重[51-56]，因此遺傳缺失這樣的現象至少部份是由於傳統 GWAS 主要都研究常見 SNP (MAF 超過 5%)，而忽視了相對不常見的變異點。本篇研究新找的 20 個位點裡面有一半其 MAF 介於 1-5%，因此有可能是此原因導致無法在別人較早發表的 GWAS 複製。不過 GWAS 未來的方向傾向會評估這些相對不常見的位點，也許本研究結果能在之後新的研究被複製。

糖尿病是全球慢性腎臟病的第一名病因，也就是影響 eGFR 變化的單一最重要後天因子[57]。在第二型糖尿病病人他們 eGFR 的變化極為複雜不一[58-60]；在糖尿病腎病變者，甚至有不少比例的人會有因為腎臟高度過濾(renal hyperfiltration)而導致 eGFR 在這階段不減反升[61]。相較於一些研究流程是保留有糖尿病的人而在後端分析再以後端共變項做校正，本篇研究是直接先將具糖尿病患者移除，如此才會有完整屏除糖尿病對於腎功能的影響。

本研究有一些限制。能夠同時具備原始三個性狀來進行主成分分析多性狀全基因組關聯分析者一共只有一萬多人，這在現今大規模基因體研究時代算是人數偏少。由於生物資料庫留的是單泡尿液，所以要評估白蛋白尿的程度如果只有白蛋白尿的濃度是不夠的，一般同時需要該泡尿液的肌酸酐值，才能做校正得到 UACR 而正確估計白蛋白尿的程度。台灣人體生物資料庫初期收集尿液時並沒有例行同時檢驗尿液肌酸酐值，而在某次與使用過資料庫的臨床醫師的會議，經由作者回饋此事，爾後尿液的檢驗皆包括測定肌酸酐值；截至作者分析的時候，擁有 UACR 資料的人數為一萬多人，相對於資料庫的總志願者人數仍偏少。eGFR 的取得，目前人體生物資料庫多半還是以檢驗血液肌酸酐為主，因為價格較便宜；但是用血液肌酸酐值推估腎功能會有些限制存在，所以目前趨勢是能夠同時有檢驗血液胱蛋白 C (cystatin C)來估計腎絲球過濾率會更精確[62]。目前台灣人體生物資料庫受試者追蹤的人數也持續增加當中，所以若未來

愈多人有追蹤的資料，具有 eGFR slope 資料的人也會增加。前述引用的 nephQTL 以及 GTEx 等 eQTL database 的資訊，由於建立的人種和台灣人不同，所以也許只能參考；未來若有建構台灣人腎臟的 transcriptomics database，將本篇研究找到的位點對照台灣人的 renal eQTL 也許有更好的結果。

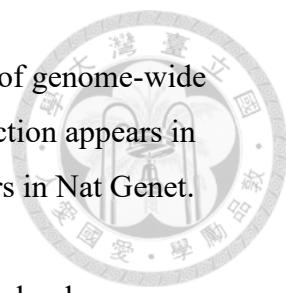
4.2 結論

本篇研究結果的部份位點已經在別的 GWAS 文章被發現，間接地重複驗證本研究的正確性，但是這些被發表過的位點就不是新穎的，另外也無法代表其它沒有被發表過的位點為真而仍然需要其他方式佐證。由於不同研究族群主成分分析的結果，即這些主成分 PCs 是獨一無二的，所以 PC-based multi-trait GWAS 的結果，並無法直接進行一般 GWAS 的複製，不過仍然可以去比對其他作者發表的各個原始性狀的全基因組關聯分析結果。另外也可以功能標註 (functional annotation) 的方式，如查詢表現數量性狀基因座資料庫，探索這些位點可能造成的影響。最終，之後仍需要嚴謹的功能研究，比如基因剔除小鼠實驗的結果來佐證目前的結果。藉由本研究的結果和經驗，PC-GWAS 的方式也值得未來在適合的疾病和族群應用和推廣。

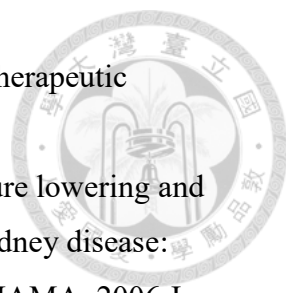
參考文獻

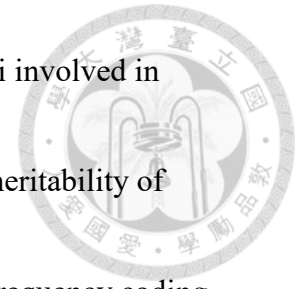


1. Hunter DJ. Gene-environment interactions in human diseases. *Nat Rev Genet.* 2005;6(4):287-298.
2. Dawn Teare M, Barrett JH. Genetic linkage studies. *Lancet.* 2005;366(9490):1036-1044.
3. Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-wide association studies. *Nat Rev Genet.* 2019;20(8):467-484.
4. Plomin R, Haworth CM, Davis OS. Common disorders are quantitative traits. *Nat Rev Genet.* 2009;10(12):872-878.
5. Chuang GT, Liu PH, Chyan TW, et al. Genome-wide association study for circulating fibroblast growth factor 21 and 23. *Sci Rep.* 2020;10(1):14578.
6. Liu PH, Chuang GT, Hsiung CN, et al. A genome-wide association study for melatonin secretion. *Sci Rep.* 2022;12(1):8025.
7. Yang WS, Chuang GT, Che TP, et al. Genome-Wide Association Studies for Albuminuria of Nondiabetic Taiwanese Population. *Am J Nephrol.* 2023;54(9-10):359-369.
8. Kovesdy CP. Epidemiology of chronic kidney disease: an update 2022. *Kidney Int Suppl* (2011). 2022;12(1):7-11.
9. Rosansky SJ. Renal function trajectory is more important than chronic kidney disease stage for managing patients with chronic kidney disease. *Am J Nephrol.* 2012;36(1):1-10.
10. Levey AS, Gansevoort RT, Coresh J, et al. Change in Albuminuria and GFR as End Points for Clinical Trials in Early Stages of CKD: A Scientific Workshop Sponsored by the National Kidney Foundation in Collaboration With the US Food and Drug Administration and European Medicines Agency. *Am J Kidney Dis.* 2020;75(1):84-104.
11. McCarthy MI, Abecasis GR, Cardon LR, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet.* 2008;9(5):356-369.
12. Yang J, Wray NR, Visscher PM. Comparing apples and oranges: equating the power of case-control and quantitative trait association studies. *Genet Epidemiol.* 2010;34(3):254-257.

- 
13. Turley P, Walters RK, Maghzian O, et al. Multi-trait analysis of genome-wide association summary statistics using MTAG [published correction appears in Nat Genet. 2019 Jul;51(7):1190] [published correction appears in Nat Genet. 2019 Aug;51(8):1295]. Nat Genet. 2018;50(2):229-237.
 14. Ringnér M. What is principal component analysis?. Nat Biotechnol. 2008;26(3):303-304.
 15. Fan CT, Lin JC, Lee CH. Taiwan Biobank: a project aiming to aid Taiwan's transition into a biomedical island. Pharmacogenomics. 2008;9(2):235-246.
 16. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559-575.
 17. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. Nat Methods. 2013;10(1):5-6.
 18. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet. 2012;44(8):955-959.
 19. Duggal P, Gillanders EM, Holmes TN, Bailey-Wilson JE. Establishing an adjusted p-value threshold to control the family-wide type 1 error in genome wide association studies. BMC Genomics. 2008;9:516.
 20. R Core Team. R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, 2021).
 21. Turner, S. qqman: An R package for visualizing GWAS results using Q-Q and Manhattan plots. J. Open Source Softw. 3, 731 (2018).
 22. Inker LA, Eneanya ND, Coresh J, et al. New Creatinine- and Cystatin C-Based Equations to Estimate GFR without Race. N Engl J Med. 2021;385(19):1737-1749.
 23. Tajti F, Kuppe C, Antoranz A, et al. A Functional Landscape of CKD Entities From Public Transcriptomic Data. Kidney Int Rep. 2019;5(2):211-224.
 24. Hishida A, Takashima N, Turin TC, et al. GCK, GCKR polymorphisms and risk of chronic kidney disease in Japanese individuals: data from the J-MICC Study. J Nephrol. 2014;27(2):143-149.
 25. Wang K, Shi M, Yang A, et al. GCKR and GCK polymorphisms are associated with increased risk of end-stage kidney disease in Chinese patients with type 2

- diabetes: The Hong Kong Diabetes Register (1995-2019). *Diabetes Res Clin Pract.* 2022;193:110118.
26. Ho LJ, Lu CH, Su RY, et al. Association between glucokinase regulator gene polymorphisms and serum uric acid levels in Taiwanese adolescents. *Sci Rep.* 2022;12(1):5519.
27. Matsuo H, Yamamoto K, Nakaoka H, et al. Genome-wide association study of clinically defined gout identifies multiple risk loci and its association with clinical subtypes. *Ann Rheum Dis.* 2016;75(4):652-659.
28. Yuan S, Larsson SC. Coffee and Caffeine Consumption and Risk of Kidney Stones: A Mendelian Randomization Study. *Am J Kidney Dis.* 2022;79(1):9-14.e1.
29. Oh H, Kwon O, Kong MJ, Park KM, Baek JH. Macrophages promote Fibrinogenesis during kidney injury. *Front Med (Lausanne).* 2023;10:1206362.
30. Cabrera-Serrano M, Caccavelli L, Savarese M, et al. Bi-allelic loss-of-function OBSCN variants predispose individuals to severe recurrent rhabdomyolysis. *Brain.* 2022;145(11):3985-3998.
31. Stahl K, Rastelli E, Schoser B. A systematic review on the definition of rhabdomyolysis. *J Neurol.* 2020;267(4):877-882.
32. Rawson ES, Clarkson PM, Tarnopolsky MA. Perspectives on Exertional Rhabdomyolysis. *Sports Med.* 2017;47(Suppl 1):33-49.
33. Levey AS, James MT. Acute Kidney Injury [published correction appears in *Ann Intern Med.* 2018 Jan 2;168(1):84]. *Ann Intern Med.* 2017;167(9):ITC66-ITC80.
34. Sone H, Lee TJ, Lee BR, Heo D, Oh S, Kwon SH. MicroRNA-mediated attenuation of branched-chain amino acid catabolism promotes ferroptosis in chronic kidney disease. *Nat Commun.* 2023;14(1):7814.
35. Gillies CE, Putler R, Menon R, et al. An eQTL Landscape of Kidney Tissue in Human Nephrotic Syndrome. *Am J Hum Genet.* 2018;103(2):232-244.
36. Stavarachi M, Panduru NM, Serafinceanu C, et al. Investigation of P213S SELL gene polymorphism in type 2 diabetes mellitus and related end stage renal disease. A case-control study. *Rom J Morphol Embryol.* 2011;52(3 Suppl):995-998.
37. Baker MA, Wang F, Liu Y, et al. MiR-192-5p in the Kidney Protects Against the Development of Hypertension. *Hypertension.* 2019;73(2):399-406.

- 
38. Huang R, Fu P, Ma L. Kidney fibrosis: from mechanisms to therapeutic medicines. *Signal Transduct Target Ther.* 2023;8(1):129.
 39. Wright JT Jr, Bakris G, Greene T, et al. Effect of blood pressure lowering and antihypertensive drug class on progression of hypertensive kidney disease: results from the AASK trial [published correction appears in JAMA. 2006 Jun 21;295(23):2726]. *JAMA.* 2002;288(19):2421-2431.
 40. Agrawal PB, Greenleaf RS, Tomczak KK, et al. Nemaline myopathy with minicores caused by mutation of the CFL2 gene encoding the skeletal muscle actin-binding protein, cofilin-2. *Am J Hum Genet.* 2007;80(1):162-167.
 41. Hari P, Bagga A, Mahajan P, Lakshmy R. Effect of malnutrition on serum creatinine and cystatin C levels. *Pediatr Nephrol.* 2007;22(10):1757-1761.
 42. Li GS, Zhu L, Zhang H, et al. Variants of the ST6GALNAC2 promoter influence transcriptional activity and contribute to genetic susceptibility to IgA nephropathy. *Hum Mutat.* 2007;28(10):950-957.
 43. Wyatt RJ, Julian BA. IgA nephropathy. *N Engl J Med.* 2013;368(25):2402-2414.
 44. Randi EB, Vervaet B, Tsachaki M, et al. The Antioxidative Role of Cytochrome in Podocytes: Implications for a Role in Chronic Kidney Disease. *Antioxid Redox Signal.* 2020;32(16):1155-1171.
 45. GTEx Consortium; Laboratory, Data Analysis & Coordinating Center (LDACC)—Analysis Working Group; Statistical Methods groups—Analysis Working Group; Genetic effects on gene expression across human tissues [published correction appears in Nature. 2017 Dec 20;:]. *Nature.* 2017;550(7675):204-213.
 46. Hurskainen T, Moilanen J, Sormunen R, et al. Transmembrane collagen XVII is a novel component of the glomerular filtration barrier. *Cell Tissue Res.* 2012;348(3):579-588.
 47. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006;38(8):904-909.
 48. Holberg CJ, Halonen M, Solomon S, et al. Factor analysis of asthma and atopy traits shows 2 major components, one of which is linked to markers on chromosome 5q. *J Allergy Clin Immunol.* 2001;108(5):772-780.
 49. Tran NK, Lea RA, Holland S, et al. Multi-phenotype genome-wide association



- studies of the Norfolk Island isolate implicate pleiotropic loci involved in chronic kidney disease. *Sci Rep.* 2021;11(1):19425.
50. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature.* 2009;461(7265):747-753.
 51. Marouli E, Graff M, Medina-Gomez C, et al. Rare and low-frequency coding variants alter human adult height. *Nature.* 2017;542(7640):186-190.
 52. Mola-Caminal M, Carrera C, Soriano-Tárraga C, et al. PATJ Low Frequency Variants Are Associated With Worse Ischemic Stroke Functional Outcome. *Circ Res.* 2019;124(1):114-120.
 53. Akiyama M, Ishigaki K, Sakaue S, et al. Characterizing rare and low-frequency height-associated variants in the Japanese population [published correction appears in *Nat Commun.* 2020 Mar 9;11(1):1350]. *Nat Commun.* 2019;10(1):4393.
 54. Pang H, Xia Y, Luo S, et al. Emerging roles of rare and low-frequency genetic variants in type 1 diabetes mellitus. *J Med Genet.* 2021;58(5):289-296.
 55. Chang X, Gurung RL, Wang L, et al. Low frequency variants associated with leukocyte telomere length in the Singapore Chinese population. *Commun Biol.* 2021;4(1):519.
 56. Clarelli F, Barizzzone N, Mangano E, et al. Contribution of Rare and Low-Frequency Variants to Multiple Sclerosis Susceptibility in the Italian Continental Population. *Front Genet.* 2022;12:800262.
 57. Alicic RZ, Rooney MT, Tuttle KR. Diabetic Kidney Disease: Challenges, Progress, and Possibilities. *Clin J Am Soc Nephrol.* 2017;12(12):2032-2045.
 58. Christensen PK, Gall MA, Parving HH. Course of glomerular filtration rate in albuminuric type 2 diabetic patients with or without diabetic glomerulopathy. *Diabetes Care.* 2000;23 Suppl 2:B14-B20.
 59. Hovind P, Rossing P, Tarnow L, Smidt UM, Parving HH. Progression of diabetic nephropathy. *Kidney Int.* 2001;59(2):702-709.
 60. Zoppini G, Targher G, Chonchol M, et al. Predictors of estimated GFR decline in patients with type 2 diabetes and preserved kidney function. *Clin J Am Soc Nephrol.* 2012;7(3):401-408.
 61. Tonneijck L, Muskiet MH, Smits MM, et al. Glomerular Hyperfiltration in Diabetes: Mechanisms, Clinical Significance, and Treatment. *J Am Soc Nephrol.*

- 2017;28(4):1023-1039.
62. Delgado C, Baweja M, Crews DC, et al. A Unifying Approach for GFR Estimation: Recommendations of the NKF-ASN Task Force on Reassessing the Inclusion of Race in Diagnosing Kidney Disease. *Am J Kidney Dis*. 2022;79(2):268-288.e1.



附 錄



表 1 腎絲球過濾率預估值 GWAS 顯著的 SNPs

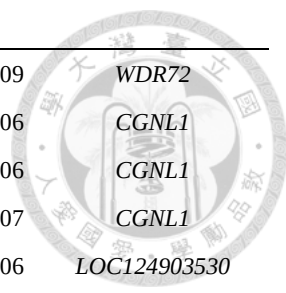
Chr	BP	SNP	A1	A2	MAF	BETA	SE	P	Nearby gene
1	15461250	rs7546774	C	T	0.24	0.256	0.056	5.2E-06	<i>CELA2A</i>
1	15471175	rs4661630	A	G	0.314	0.378	0.052	3.2E-13	<i>CELA2A</i>
1	15585452	rs45619934	G	T	0.44	0.544	0.048	1.9E-29	<i>AGMAT</i>
1	15987690	rs9287283	T	A	0.209	-0.28	0.059	2.6E-06	-
1	22133715	rs10917161	C	T	0.224	-0.28	0.057	7.1E-07	<i>WNT4</i>
1	23258863	rs59613632	A	G	0.265	0.241	0.054	9.3E-06	-
1	23313056	rs3765941	C	T	0.403	0.305	0.049	4E-10	<i>HNRNPR</i>
1	23382426	rs10737472	C	A	0.26	0.338	0.055	6.2E-10	<i>TCEA3</i>
1	77434854	rs1498403	G	T	0.238	-0.28	0.056	5E-07	<i>AK5</i>
1	77443300	rs6658302	G	A	0.424	-0.21	0.048	9.5E-06	<i>AK5</i>
1	77464358	rs12066994	T	C	0.433	0.241	0.048	6.2E-07	<i>AK5</i>
1	77604773	rs9787306	T	C	0.425	0.379	0.049	5.9E-15	<i>ZZZ3</i>
1	77609317	rs12026939	A	G	0.379	-0.31	0.049	5.5E-10	<i>ZZZ3</i>
1	78093465	rs586680	T	C	0.484	0.234	0.049	1.5E-06	<i>GIPC2</i>
1	87400026	rs151248243	A	G	0.01	-1.1	0.237	3.1E-06	<i>LOC105378833</i>
1	155179729	rs76872124	T	C	0.128	-0.58	0.071	3.2E-16	<i>TRIM46</i>
1	155185239	rs2070803	A	G	0.254	-0.44	0.055	1.2E-15	<i>TRIM46</i>
1	155248574	rs734073	T	C	0.293	-0.27	0.052	2.4E-07	<i>FAM189B</i>
2	27527678	rs2950835	A	G	0.487	0.413	0.048	1.4E-17	<i>GCKR</i>
2	27621734	rs1881396	G	T	0.372	-0.24	0.05	1.1E-06	<i>ZNF512</i>
2	86071465	rs311578	C	T	0.456	-0.23	0.048	1.2E-06	<i>POLR1A</i>
2	98816726	rs149569919	C	G	0.011	1.159	0.225	2.7E-07	<i>CRACDL</i>
2	121190731	rs4622731	G	A	0.343	0.226	0.051	8.3E-06	-
2	162206710	rs117309866	A	C	0.113	0.383	0.075	3.7E-07	<i>FAP</i>
2	169177199	rs2193196	T	C	0.026	-0.67	0.15	7.7E-06	<i>LRP2</i>
2	169298927	rs2244407	T	C	0.112	0.377	0.076	6.5E-07	<i>LRP2</i>
2	169335693	rs831027	T	C	0.111	0.612	0.076	1.2E-15	<i>LRP2</i>
2	169338731	rs141807916	G	A	0.031	0.674	0.139	1.2E-06	<i>LRP2</i>
2	169344929	rs831036	G	C	0.472	0.404	0.048	3.9E-17	<i>LRP2</i>
2	169354867	rs2389557	G	A	0.461	0.237	0.049	1E-06	<i>LRP2</i>

2	176141029	rs1348807	C	T	0.372	-0.28	0.05	1.6E-08	HOXD-AS2
2	176156673	rs4972505	G	A	0.288	-0.24	0.053	4.8E-06	HOXD3
2	178866928	rs11893004	A	G	0.077	-0.42	0.09	3.3E-06	CCDC141
2	216819113	rs2541381	G	T	0.488	0.306	0.048	1.4E-10	IGFBP-AS1
3	32841917	rs997519	T	C	0.13	0.317	0.071	8.5E-06	TRIM71
3	38393532	rs1069267	T	A	0.251	-0.26	0.056	3.9E-06	XYLB
3	38474878	rs762318	A	C	0.483	-0.29	0.048	9.2E-10	ACVR2B
3	38532255	rs7647657	G	A	0.396	0.227	0.049	4E-06	-
3	38550915	rs1805126	A	G	0.453	-0.25	0.048	3.1E-07	SCN5A
3	126190826	rs1992855	C	T	0.283	-0.26	0.053	5.4E-07	ALDH1L1-AS2
3	141376400	rs77316854	A	T	0.173	-0.31	0.063	6.1E-07	ZBTB38
3	141402972	rs2871960	C	A	0.339	-0.3	0.05	3.6E-09	ZBTB38
3	169382329	rs16853620	G	A	0.423	0.258	0.048	8.6E-08	MECOM
3	169396966	rs10804832	G	T	0.425	-0.31	0.048	1.6E-10	MECOM
3	169437688	rs62294340	A	G	0.343	0.245	0.05	1.1E-06	MECOM
3	186719129	rs5029975	T	C	0.336	-0.27	0.051	1.1E-07	KNG1
3	186727263	rs2304456	G	T	0.173	-0.36	0.063	1.6E-08	KNG1
4	10016355	rs12500805	C	T	0.19	-0.27	0.061	9.8E-06	SLC2A9
4	23721585	rs73260452	A	G	0.04	0.726	0.122	2.6E-09	LOC105374528
4	76273320	rs768981	A	G	0.386	-0.23	0.049	2.1E-06	FAM47E
4	76330768	rs117223770	A	G	0.059	-0.53	0.102	2E-07	CCDC158
4	76371648	rs6853408	C	T	0.296	-0.32	0.052	1.6E-09	CCDC158
4	76430579	rs80309664	T	C	0.022	-0.74	0.162	4.6E-06	SHROOM3
4	76446535	rs1398016	A	G	0.11	-0.6	0.076	3.2E-15	SHROOM3
4	76463939	rs76123685	A	G	0.026	-0.72	0.153	2.1E-06	SHROOM3
4	76472865	rs28394165	C	T	0.232	-0.57	0.056	7.7E-24	SHROOM3
4	76480042	rs6532414	G	A	0.474	0.245	0.048	3.3E-07	SHROOM3
4	76497920	rs4859683	C	T	0.442	0.248	0.048	2.4E-07	SHROOM3
4	80263187	rs16998073	T	A	0.418	0.309	0.048	1.6E-10	FGF5
4	88004244	rs72873442	T	G	0.446	0.219	0.048	5.2E-06	PKD2
4	88105276	rs2725263	A	C	0.462	0.245	0.048	3.4E-07	ABCG2
4	88108714	rs34472643	A	G	0.264	0.293	0.055	9.4E-08	ABCG2
4	88110434	rs11939579	G	T	0.334	0.323	0.051	2.6E-10	ABCG2
4	88123028	rs149027545	C	G	0.316	-0.31	0.051	2.3E-09	ABCG2
4	88146855	rs2725248	C	A	0.213	0.322	0.058	3.5E-08	ABCG2
4	88197859	rs10022955	G	A	0.336	0.245	0.051	1.8E-06	ABCG2
4	89341870	rs13108493	T	C	0.356	-0.25	0.05	6.8E-07	-

4	99901474	rs75911653	A	T	0.018	-0.84	0.182	4.5E-06	DNAJB14
4	102777629	rs223471	C	G	0.382	0.254	0.049	2.3E-07	LOC102723704
4	102827786	rs223389	A	G	0.206	0.298	0.059	5.4E-07	UBE2D3
4	103053903	rs7688014	T	C	0.472	-0.29	0.048	3.3E-09	SLC9B2
5	53491081	rs12518560	A	T	0.345	0.227	0.051	9.4E-06	FST
5	53822988	rs10072767	A	G	0.234	0.289	0.056	2.9E-07	LINC02105
5	177352688	rs34316179	A	G	0.068	-0.54	0.095	1.2E-08	LMAN2
5	177371305	rs11746443	A	G	0.135	-0.66	0.07	2.3E-21	RGS14
5	177388123	rs5030873	C	T	0.205	-0.55	0.06	3.3E-20	SLC34A1
6	25812564	rs141559354	A	G	0.02	-0.92	0.187	7.9E-07	SLC17A1
6	27530802	rs12208078	A	G	0.493	-0.28	0.05	1.2E-08	-
6	27820212	rs12196939	A	G	0.027	-0.78	0.163	2E-06	H4C11
6	27877387	rs9348780	T	C	0.296	0.253	0.055	4.2E-06	H4C13
6	27962858	rs839776	A	G	0.35	0.27	0.053	4.3E-07	OR2B6
6	28020989	rs140674470	G	C	0.192	0.349	0.063	2.8E-08	-
6	28148926	rs140323706	C	T	0.045	0.598	0.116	2.7E-07	ZKSCAN8
6	28240757	rs1150711	T	C	0.456	-0.26	0.05	1.7E-07	ZKSCAN4
6	28247668	rs1233710	T	C	0.372	0.363	0.053	8.9E-12	ZKSCAN4
6	28306874	rs75426794	G	A	0.374	0.282	0.052	5.4E-08	PGBD1
6	28417930	rs187072630	A	T	0.018	-0.83	0.188	9.4E-06	ZSCAN23
6	28509805	rs35403242	A	G	0.357	0.257	0.053	1.1E-06	GPX6
6	28515813	rs6921021	A	G	0.468	-0.25	0.052	1.8E-06	GPX6
6	28896337	rs144048344	A	T	0.046	0.612	0.116	1.5E-07	HCG14
6	29404107	rs10946988	A	T	0.238	0.257	0.058	8.4E-06	-
6	29472964	rs12206077	A	G	0.138	0.433	0.074	4.8E-09	LOC105375008
6	29516743	rs117442778	C	T	0.062	0.642	0.102	3.3E-10	-
6	29635267	rs62392944	T	C	0.121	0.43	0.082	1.8E-07	GABBR1
6	30008063	rs3765607	A	G	0.078	0.53	0.095	2.2E-08	ZNRD1ASP
6	31281350	rs2253487	A	G	0.412	-0.28	0.055	5.5E-07	-
6	31593738	rs9357135	T	C	0.118	0.501	0.089	2.1E-08	NCR3
6	31713383	rs385306	A	G	0.373	-0.27	0.055	9.3E-07	LY6G6D
6	31802019	rs9461726	T	C	0.043	-0.67	0.145	3.5E-06	LSM2
6	32320099	rs513095	G	T	0.095	-0.42	0.081	2.3E-07	TSBP1
6	32320913	rs9394085	A	G	0.144	0.397	0.076	1.6E-07	TSBP1
6	32363432	rs2223636	A	T	0.45	-0.22	0.05	9.2E-06	TSBP1
6	32372459	rs9268399	A	G	0.403	0.246	0.051	1.6E-06	TSBP1
6	32413684	rs3129967	T	C	0.316	-0.24	0.052	3.7E-06	-

6	32617972	rs75483390	A	G	0.025	-0.77	0.16	1.5E-06	-
6	32658953	rs34386495	T	C	0.049	-0.6	0.133	5.8E-06	<i>HLA-DQB1</i>
6	32666194	rs3830060	G	A	0.239	0.284	0.06	2.4E-06	<i>HLA-DQB1</i>
6	34268653	rs9469762	A	G	0.085	-0.43	0.086	5.9E-07	-
6	41690374	rs9471627	C	G	0.388	0.22	0.049	6.8E-06	<i>TFEB</i>
6	43837625	rs744103	T	A	0.218	0.335	0.058	6.6E-09	<i>LOC105375069</i>
6	54486360	rs78828070	G	A	0.089	-0.53	0.084	2.5E-10	-
6	54558447	rs146677015	G	C	0.014	-1.38	0.2	5.1E-12	-
6	55178438	rs4454139	A	C	0.257	-0.3	0.055	4.4E-08	<i>HCRTR2</i>
6	65644556	rs200330129	T	C	0.011	-1.03	0.226	5.5E-06	<i>EYS</i>
6	106724525	rs3827648	T	C	0.153	-0.35	0.066	1.4E-07	<i>LINC02532</i>
6	130181238	rs7356793	T	C	0.017	-0.87	0.185	2.8E-06	<i>SAMD3</i>
6	133628177	rs2636626	A	G	0.429	0.228	0.048	2.2E-06	<i>TARID</i>
6	160131975	rs4646275	T	C	0.435	-0.24	0.048	4.1E-07	<i>SLC22A1</i>
6	160134722	rs594709	G	A	0.259	0.282	0.054	2.1E-07	<i>SLC22A1</i>
6	160245769	rs9766047	T	G	0.079	-0.69	0.089	5.2E-15	<i>SLC22A2</i>
6	160253525	rs316010	C	T	0.032	0.716	0.136	1.3E-07	<i>SLC22A2</i>
6	160283880	rs3127586	C	T	0.084	-0.54	0.087	6.4E-10	<i>LOC105378088</i>
7	26901856	rs213519	T	A	0.294	-0.24	0.053	4.7E-06	-
7	41018175	rs17715435	T	C	0.166	-0.36	0.064	2.8E-08	<i>SUGCT</i>
7	46687365	rs856560	T	C	0.083	0.414	0.086	1.7E-06	<i>LOC730338</i>
7	46714086	rs700753	C	G	0.143	0.352	0.068	2.3E-07	-
7	77762950	rs6952760	C	G	0.312	-0.25	0.052	8.8E-07	<i>RSBN1L</i>
7	77922810	rs848486	G	A	0.206	-0.35	0.059	5E-09	<i>PHTF2</i>
7	155869487	rs28591492	A	G	0.075	-0.55	0.091	1.3E-09	-
7	156318624	rs10230170	T	C	0.413	0.23	0.049	2.8E-06	-
7	156482785	rs80153796	T	C	0.095	0.397	0.082	1.2E-06	<i>RNF32-DT</i>
8	23841808	rs310311	G	C	0.432	0.221	0.048	4.3E-06	<i>STC1</i>
8	23863584	rs1436148	C	T	0.293	-0.36	0.053	1.4E-11	-
8	23899888	rs1705689	G	C	0.409	-0.35	0.049	5E-13	-
8	32527258	rs6651145	T	C	0.263	-0.26	0.054	2.6E-06	<i>NRG1</i>
8	32578102	rs10102889	C	G	0.394	-0.25	0.049	4.7E-07	<i>NRG1</i>
8	86338721	rs7820234	A	C	0.447	-0.22	0.048	3E-06	<i>WWP1</i>
8	101620345	rs6992430	G	A	0.248	-0.25	0.055	6.6E-06	<i>GRHL2</i>
8	101663059	rs4734038	T	C	0.347	-0.23	0.05	3.2E-06	<i>GRHL2</i>
8	137598087	rs77766076	T	G	0.015	-0.89	0.199	7.5E-06	-
9	68816258	rs7042786	T	A	0.499	0.332	0.048	3.4E-12	<i>PIP5K1B</i>

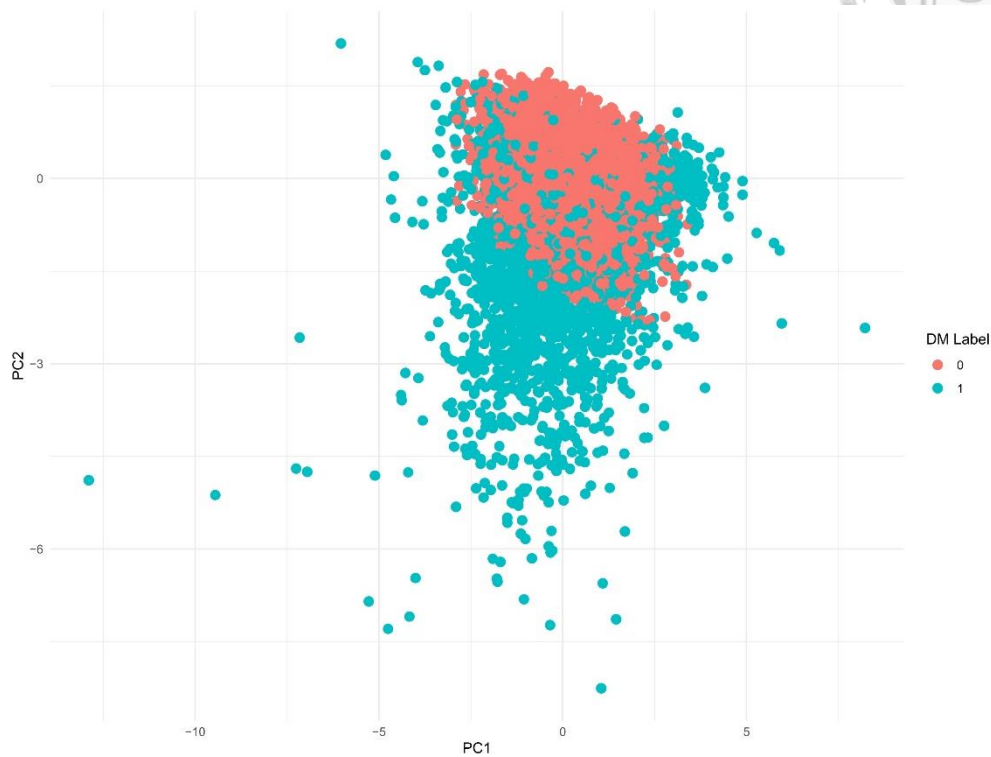
9	68836044	rs10869396	A	C	0.334	0.33	0.051	7.3E-11	PIP5K1B
9	68855195	rs7041004	T	C	0.373	-0.25	0.049	2.4E-07	PIP5K1B
9	94999231	rs10993449	G	A	0.22	-0.28	0.058	1.3E-06	AOPEP
9	95302247	rs356664	T	C	0.19	-0.27	0.061	8.7E-06	FANCC
9	99104830	rs10819636	T	G	0.435	-0.26	0.048	7.6E-08	TGFBR1
9	109353249	rs1857050	A	T	0.214	0.263	0.058	6.4E-06	-
10	823912	rs74876360	G	C	0.096	-0.47	0.081	9.6E-09	LARP4B
10	68192084	rs10823156	C	G	0.214	0.293	0.058	4.7E-07	MYPN
10	68205718	rs61854764	C	A	0.11	-0.35	0.077	5.7E-06	MYPN
10	87121419	rs9421590	G	A	0.371	-0.23	0.049	2.3E-06	SHLD2
10	124715572	rs4962687	G	A	0.379	0.222	0.049	6.1E-06	FAM53B
10	124784740	rs2947343	C	G	0.228	0.271	0.057	1.8E-06	EEF1AKMT2
11	2144346	rs35506085	G	A	0.439	-0.27	0.048	2.1E-08	IGF2
11	30747131	rs10767873	T	C	0.363	0.555	0.05	3.6E-29	LOC101928338
11	30948660	rs163874	T	C	0.44	0.226	0.048	2.8E-06	DCDC1
11	64639913	rs10400348	A	C	0.297	0.238	0.052	5E-06	NRXN2
11	78229029	rs2450128	A	G	0.378	0.276	0.049	1.8E-08	GAB2
11	78265429	rs4415799	T	C	0.291	-0.24	0.052	6.6E-06	GAB2
11	78429327	rs12277503	C	A	0.406	0.218	0.049	8.8E-06	LINC02728
12	236203	rs2080403	G	C	0.099	-0.38	0.08	2.6E-06	SLC6A13
12	246676	rs7969761	T	C	0.205	-0.35	0.059	2.8E-09	SLC6A13
12	3270736	rs3782768	A	G	0.076	-0.4	0.09	8.6E-06	TSPAN9
12	48002581	rs1635527	C	G	0.271	0.251	0.054	2.8E-06	COL2A1
12	65845505	rs61425523	A	G	0.225	-0.26	0.057	6.2E-06	HMGA2
12	65983657	rs61409219	C	T	0.074	0.457	0.092	6.3E-07	-
12	111269073	rs4766566	C	T	0.324	-0.24	0.051	3.2E-06	CUX2
12	112030402	rs11066132	T	C	0.279	-0.32	0.053	1.3E-09	NAA25
13	42128575	rs1170155	T	C	0.406	-0.23	0.049	2.4E-06	DGKH
13	42165097	rs9525588	G	A	0.358	-0.24	0.05	2.4E-06	DGKH
13	42201055	rs2324889	T	C	0.318	-0.26	0.052	4.5E-07	DGKH
13	42207571	rs67354544	T	C	0.41	0.276	0.048	1.2E-08	DGKH
13	44122774	rs147604452	T	C	0.015	-0.93	0.193	1.4E-06	SMIM2-AS1
13	47969823	rs12864691	C	T	0.277	-0.26	0.054	9.4E-07	SUCLA2
13	48093742	rs9567994	G	C	0.467	-0.24	0.048	6.2E-07	MED4
13	71774636	rs500830	C	T	0.123	-0.47	0.073	1.5E-10	DACH1
14	34778642	rs78657356	G	A	0.176	0.323	0.063	2.7E-07	BAZ1A
15	45348900	rs2467864	T	C	0.168	0.39	0.064	9.1E-10	-



15	53644221	rs10518731	T	A	0.482	0.277	0.048	5.5E-09	WDR72
15	57437769	rs7181770	C	A	0.066	0.443	0.096	3.9E-06	CGNL1
15	57495352	rs77319829	A	C	0.1	0.373	0.08	2.7E-06	CGNL1
15	57524384	rs6493932	A	G	0.051	0.572	0.109	1.4E-07	CGNL1
15	75131994	rs7172205	A	G	0.242	0.256	0.056	4.3E-06	LOC124903530
15	75922879	rs78653469	A	G	0.068	0.479	0.095	5.1E-07	FBXO22
15	75928941	rs75629444	T	C	0.155	0.499	0.066	3.8E-14	FBXO22
15	75946679	rs11636251	T	C	0.477	-0.35	0.048	1.3E-13	NRG4
15	82666511	rs182066759	G	A	0.026	-0.68	0.15	5.9E-06	AP3B2
15	83095928	rs118002544	A	G	0.047	-0.61	0.113	8.4E-08	LOC124903543
15	84771166	rs17534709	C	G	0.114	0.349	0.075	3.7E-06	ZNF592
15	98704789	rs4966014	C	T	0.24	-0.38	0.056	1.3E-11	IGF1R
15	98741451	rs4966020	G	A	0.451	-0.4	0.049	5.8E-16	IGF1R
15	98771392	rs12437561	T	C	0.29	-0.31	0.052	1.9E-09	IGF1R
16	20373822	rs34495073	C	T	0.139	0.38	0.069	3.8E-08	PDILT
16	51737193	rs7190448	C	G	0.388	-0.26	0.049	1.1E-07	LOC105371257
16	51743034	rs67157068	G	T	0.172	0.291	0.063	4.3E-06	LOC105371257
16	53283844	rs17303428	A	G	0.268	-0.25	0.054	3.2E-06	CHD9
17	36561039	rs8068351	C	T	0.311	0.242	0.051	2.4E-06	GGNBP2
17	39358417	rs12935933	G	T	0.248	-0.4	0.055	5.3E-13	FBXL20
17	61155004	rs11655024	T	C	0.089	0.384	0.085	5.7E-06	BCAS3
17	61161860	rs11868441	A	G	0.22	0.363	0.057	2.6E-10	BCAS3
17	61370008	rs11653176	T	C	0.489	-0.46	0.048	1.2E-21	BCAS3
17	61408032	rs2240736	T	C	0.415	0.49	0.048	4E-24	TBX2
17	61420889	rs9905385	A	G	0.281	-0.48	0.053	4.1E-19	-
18	5592780	rs1719939	C	T	0.242	-0.26	0.056	3.2E-06	EPB41L3
19	32789837	rs10409119	G	A	0.471	-0.21	0.048	9.3E-06	TDRD12
19	32869463	rs7252778	A	C	0.299	-0.26	0.052	9E-07	SLC7A9
19	32939910	rs12461782	T	G	0.413	0.242	0.048	5.7E-07	CEP89
20	10705728	rs2423512	C	A	0.482	-0.22	0.048	4.3E-06	-
20	58901754	rs8121252	T	C	0.272	-0.24	0.054	6.2E-06	GNAS
22	40489495	rs73167082	C	T	0.213	0.362	0.058	5.7E-10	MRTFA

圖 1 原始具有 eGFR, eGFR slope, albuminuria 資料之 13,904 人 PCA 散佈圖檢視有糖尿病者之分佈

(A) 以 PC1-PC2 為 X-Y 軸的散佈圖



(B) 以 PC2-PC3 為 X-Y 軸的散佈圖



OPEN

Genome-wide association study for circulating fibroblast growth factor 21 and 23

Gwo-Tsann Chuang^{1,2,14}, Pi-Hua Liu^{3,4,14}, Tsui-Wei Chyan⁵, Chen-Hao Huang⁵, Yu-Yao Huang^{4,6}, Chia-Hung Lin^{4,7}, Jou-Wei Lin⁸, Chih-Neng Hsu⁸, Ru-Yi Tsai⁸, Meng-Lun Hsieh⁹, Hsiao-Lin Lee⁹, Wei-shun Yang^{2,9}, Cassianne Robinson-Cohen¹⁰, Chia-Ni Hsiung¹¹, Chen-Yang Shen^{12,13} & Yi-Cheng Chang^{2,9,12}✉

Fibroblast growth factors (FGFs) 21 and 23 are recently identified hormones regulating metabolism of glucose, lipid, phosphate and vitamin D. Here we conducted a genome-wide association study (GWAS) for circulating FGF21 and FGF23 concentrations to identify their genetic determinants. We enrolled 5,000 participants from Taiwan Biobank for this GWAS. After excluding participants with diabetes mellitus and quality control, association of single nucleotide polymorphisms (SNPs) with log-transformed FGF21 and FGF23 serum concentrations adjusted for age, sex and principal components of ancestry were analyzed. A second model additionally adjusted for body mass index (BMI) and a third model additionally adjusted for BMI and estimated glomerular filtration rate (eGFR) were used. A total of 4,201 participants underwent GWAS analysis. rs67327215, located within *RGS6* (a gene involved in fatty acid synthesis), and two other SNPs (rs12565114 and rs9520257, located between *PHC2-ZSCAN20* and *ARGLU1-FAM155A* respectively) showed suggestive associations with serum FGF21 level ($P = 6.66 \times 10^{-7}$, 6.00×10^{-7} and 6.11×10^{-7} respectively). The SNPs rs17111495 and rs17843626 were significantly associated with FGF23 level, with the former near *PCSK9* gene and the latter near *HLA-DQA1* gene ($P = 1.04 \times 10^{-10}$ and 1.80×10^{-8} respectively). SNP rs2798631, located within the *TGFB2* gene, was suggestively associated with serum FGF23 level ($P = 4.97 \times 10^{-7}$). Additional adjustment for BMI yielded similar results. For FGF23, further adjustment for eGFR had similar results. We conducted the first GWAS of circulating FGF21 levels to date. Novel candidate genetic loci associated with circulating FGF21 or FGF23 levels were found. Further replication and functional studies are needed to support our findings.

Fibroblast growth factor (FGF) superfamily includes 22 family members. Except for the endocrine FGF19, FGF21, and FGF23, other family members act as autocrine or paracrine factors¹. Recent evidence showed that FGF21 exerts metabolic effects by acting both centrally and peripherally, mostly in the liver, adipose tissue and pancreas. In prolonged starvation, FGF21 is secreted by the liver to promote gluconeogenesis, increases hepatic

¹Department of Pediatrics, National Taiwan University Hospital, College of Medicine, National Taiwan University, Taipei, Taiwan, ROC. ²Graduate Institute of Medical Genomics and Proteomics, National Taiwan University, 5F, No.2, Xuzhou Rd., Zhongzheng Dist., Taipei 100, Taiwan, ROC. ³Clinical Informatics and Medical Statistics Research Center, College of Medicine, Chang Gung University, Taoyuan, Taiwan, ROC. ⁴Division of Endocrinology and Metabolism, Department of Internal Medicine, Chang Gung Memorial Hospital At Linkou, Taoyuan, Taiwan, ROC. ⁵Institute of Molecular Medicine, College of Medicine, National Taiwan University, Taipei, Taiwan, ROC. ⁶Department of Medical Nutrition Therapy, Chang Gung Memorial Hospital At Linkou, Taoyuan, Taiwan, ROC. ⁷Department of Chinese Medicine, College of Medicine, Chang Gung University, Taoyuan, Taiwan, ROC. ⁸Cardiovascular Center, National Taiwan University Hospital Yun-Lin Branch, Taipei, Taiwan, ROC. ⁹Department of Internal Medicine, College of Medicine, National Taiwan University, Taipei, Taiwan, ROC. ¹⁰Division of Nephrology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. ¹¹Data Science Statistical Cooperation Center, Institute of Statistical Science, Academia Sinica, Taipei, Taiwan, ROC. ¹²Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan, ROC. ¹³College of Public Health, China Medical University, Taichung, Taiwan, ROC. ¹⁴These authors contributed equally: Gwo-Tsann Chuang and Pi-Hua Liu. ✉email: b83401040@gmail.com



Characteristic	
Total participants (N)	4,201
Age, year	49.54 ± 10.82
Males (%)	1,446 (34.42)
BMI, kg/m ²	24.32 ± 6.67
Serum glucose, mg/dl	93.01 ± 8.26
eGFR (MDRD), ml/min/1.73 m ²	106.80 ± 24.98
FGF21, ng/ml	0.16 (0.10–0.27)
FGF23, ng/ml	0.36 (0.11–1.74)

Table 1. Study characteristics of participants. Data are mean ± SD, median (IQR) or number (%), as appropriate. Age is at specimen collection. BMI body mass index; MDRD modification of diet in renal disease equation.

SNP	Chr	Position	Nearest gene	FGF21 Increasing allele	Other allele	FGF21 increasing allele frequency	Model 1		Model 2	
							P value	Beta (SEM)	P value	Beta (SEM)
rs12565114	1	33,921,202	PHC2-ZSCAN20	G	A	0.69	6.00 × 10 ^{−7}	0.048 (0.010)	4.69 × 10 ^{−7}	0.049 (0.010)
rs9520257	13	107,631,394	ARGLU1-FAM155A	C	T	0.85	6.11 × 10 ^{−7}	0.062 (0.012)	3.96 × 10 ^{−7}	0.063 (0.012)
rs67327215	14	72,694,874	RGS6	G	A	0.62	6.66 × 10 ^{−7}	0.046 (0.009)	6.93 × 10 ^{−7}	0.046 (0.009)

Table 2. Top genetic polymorphisms associated with log-transformed FGF21 level. Model 1 was adjusted for age, sex and the first ten principal components of ancestry. Model 2 was additionally adjusted for BMI. SNP single nucleotide polymorphism; Chr chromosome. rs12565114 and rs9520257 are genotyped. rs67327215 is imputed with high info score (0.984).

fatty acid oxidation and ketogenesis², enhances secretion of glucocorticoid, and reduces behaviors that waste energy such as ovulation³. In contrast, in fed status, FGF21 promotes glucose uptake by fat cells and increase heat production of brown fat⁴. In the pancreatic islet, FGF21 preserves the survival of beta-cell and restores insulin synthesis under the stress of nutrient excess⁵. Of note, FGF21 acts on the brain to induce sympathetic nerve activity, thereby increasing energy expenditure and promoting weight loss⁶. FGF21 is regulated by both peroxisome proliferator-activated receptor (PPAR) γ and PPARα and is an important mediator of the downstream effects of PPARγ and PPARα^{6,7}. FGF21 is currently a drug target for treating lipid disorders and fatty liver⁸.

FGF23 is the physiological regulator of phosphate and vitamin D serum levels. FGF23 lowers apical membrane expression of sodium-dependent phosphate co-transporter 2A and sodium-dependent phosphate co-transporter 2C in the kidney, both of which primarily mediate renal tubular phosphate reabsorption. Furthermore, FGF23 lowers serum 1,25-dihydroxyvitamin D₃ (1,25(OH)₂D₃) levels, thereby preventing hyperphosphatemia and hypervitaminosis^{1,9}. FGF23 has structure and biological features similar to those of FGF19 and FGF21, which were known to regulate glucose and lipid metabolism¹⁰. FGF23 level was also found to be associated with obesity and dyslipidemia in studies of elderly cohorts¹¹. FGF23 is also an important drug target for treating phosphate disorder and bone disease¹². Robinson-Cohen et al. performed a genome-wide association study (GWAS) of circulating FGF23 concentrations among people of European ancestry, and found five genome-wide significant loci¹³.

To understand better the regulation of circulating FGF21 and FGF23 levels, we performed a GWAS to identify genetic determinants of circulating FGF21 and FGF23 levels in Taiwanese population.

Results

Of the 5,000 subjects enrolled from Taiwan Biobank, one withdrew from the study, 559 were excluded due to their having diabetes mellitus, and 239 were excluded by quality control procedures. A total of 617,073 genotyped autosomal SNPs remained. After imputation and post-imputation quality control, 7,897,704 SNPs remained. We performed GWAS analysis in the remaining 4,201 subjects, where log-transformed FGF21 or FGF23 level was the quantitative phenotype. Study characteristics are shown in Table 1.

The results of association testing for FGF21 and FGF23 are listed, showing SNP with strongest association in each region (Tables 2 and 3, respectively). Manhattan plots and quantile–quantile plots of log-transformed FGF21 and FGF23 are shown in Figs. 1 and 2, respectively. Regional GWAS plots of each SNP are shown in Figs 3 and 4. For FGF21, no significantly associated SNP is found. One suggestive locus is located at the intergenic region between PHC2 and ZSCAN20 genes on chromosome 1 (rs12565114; P=6.00 × 10^{−7}, Fig. 3a), another is between ARGLU1 and FAM155A genes on chromosome 13 (rs9520257; P=6.11 × 10^{−7}, Fig. 3b), and the other is within the RGS6 gene on chromosome 14 (rs67327215; P=6.66 × 10^{−7}, Fig. 3c). For FGF23, two significant loci are found. One locus is near gene PCSK9 on chromosome 1 (rs17111495; P=1.04 × 10^{−10}, Fig. 4a), and the other is near the HLA-DQA1 gene on chromosome 6 (rs17843626; P=1.80 × 10^{−8}, Fig. 4b). There is also a suggestive locus within the TGFB2 gene on chromosome 1 (rs2798631; P=4.97 × 10^{−7}, Fig. 4c). The array-based heritability estimates for FGF21 and FGF23 were 0.1171 and 0.1104 respectively. Proportion of variance in serum FGF21 level explained

SNP	Chr	Position	Nearest gene	FGF23 increasing allele	Other allele	FGF23 increasing allele frequency	Model 1		Model 2		Model 3	
							P value	Beta (SEM)	P value	Beta (SEM)	P value	Beta (SEM)
rs17111495	1	55,500,706	PCSK9	G	C	0.98	1.04×10^{-10}	0.463 (0.071)	1.57×10^{-10}	0.458 (0.071)	1.55×10^{-10}	0.458 (0.071)
rs17843626	6	32,621,013	HLA-DQA1	G	A	0.49	1.80×10^{-8}	0.117 (0.021)	1.70×10^{-8}	0.117 (0.021)	1.75×10^{-8}	0.117 (0.021)
rs2798631	1	218,611,878	TGFB2	G	A	0.75	4.97×10^{-7}	0.115 (0.023)	3.73×10^{-7}	0.116 (0.023)	3.71×10^{-7}	0.116 (0.023)

Table 3. Top genetic polymorphisms associated with log-transformed FGF23 level. Model 1 was adjusted for age, sex and the first ten principal components of ancestry. Model 2 was additionally adjusted for BMI. Model 3 was additionally adjusted for BMI and eGFR. SNP single nucleotide polymorphism; Chr chromosome. rs17111495, rs2798631 and rs17843626 are imputed with high info scores (0.809, 0.997 and 0.990 respectively).

by rs12565114, rs9520257 and rs67327215 were 0.5%, 0.6% and 0.5% respectively; for serum FGF23 level, proportion of variance explained by rs17111495, rs2798631 and rs17843626 were 1.0%, 0.6% and 0.7% respectively.

Further excluding two more subjects with severe renal impairment (eGFR < 30 ml/min/1.73m²) did not alter the results of top SNPs associated with FGF23 level (Supplementary Table S1).

We looked for the association of the five SNPs identified to be associated with serum FGF23 levels by Robinson-Cohen et al.¹³ in the European population but failed to found significant association with the FGF23 level in our cohort (Supplementary Table S2). In view of low minor allele frequency of rs17216707, we performed gene-based association analysis using MAGMA v1.07b¹⁴ and still found no significant association (*CYP24A1*, $P=0.18$). The regional plots regarding association with serum FGF23 levels in Taiwan Biobank at the five loci identified previously by Robinson-Cohen and colleagues are shown in Supplementary Fig. S1a–e. We also performed meta-analysis pooling our results and those reported by Robinson-Cohen et al.¹³ using METAL¹⁵ but the results were not significant as shown in Supplementary Table S3 (P values were 0.00070, 0.0010 and 0.00029 for rs17111495, rs2798631 and rs17843626, respectively).

Discussion

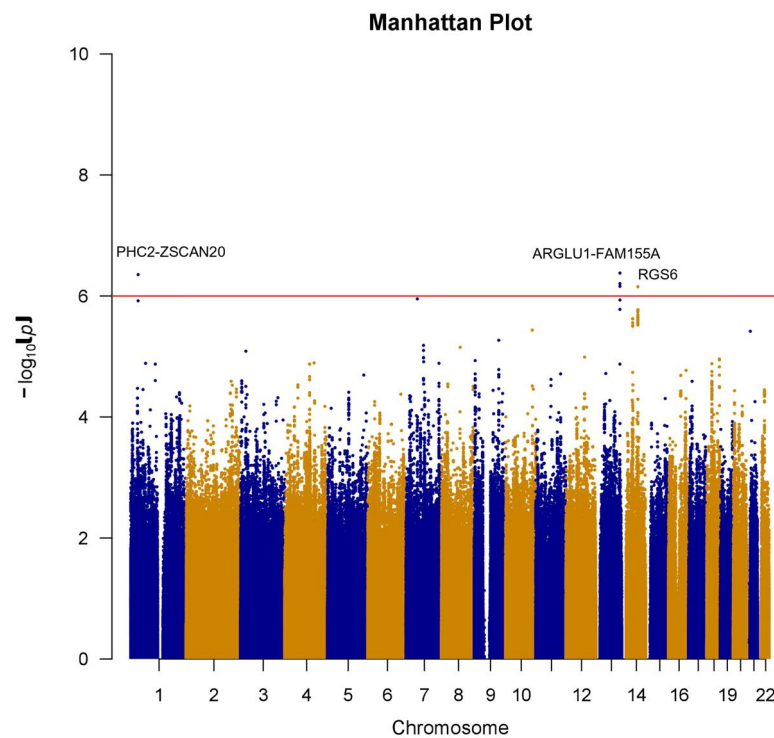
In this GWAS for evaluating both circulating FGF21 and FGF23 levels, we identified several loci for each trait that showed associations. SNP rs67327215, associated with FGF21 level, is located within *RGS6* (Regulator of G protein signaling 6). *RGS6* belongs to the R7 subfamily of RGS proteins¹⁶ and is expressed in a variety of organs and tissues, including the central nervous system and liver¹⁷. RGS proteins regulate G protein-coupled receptor-initiated signaling by acting as GTPase-activating proteins for Gα subunits, which hydrolyze GTP and restore the inactive Gα_{GDP} βγ heterotrimer¹⁸. A previous study has shown *Rgs16* knockout mice expressing increased *Fgf21* expression in the liver while overexpression of *Rgs16* decreased *Fgf21* expression. It was postulated that *RGS16* inhibits Gα_i/Gα_q-mediated fatty acid oxidation, thereby decreasing the induced *Fgf21* expression by peroxisome proliferator-activated receptor alpha (PPARα)¹⁹. *RGS6* knockout mice revealed lower mRNA levels of PPARα and PPARγ²⁰. These data support that *RGS6* may also act as a regulator of circulating FGF21 levels through fatty acid synthesis and oxidation.

The SNPs rs12565114 and rs9520257, also suggestively associated with FGF21 level, are intergenic SNPs on chromosome 1 and 13 respectively. rs12565114 is located between *PHC2* (Polyhomeotic Homolog 2) and *ZSCAN20* (Zinc Finger And SCAN Domain Containing 20) genes. *PHC2* is associated with conventional angiosarcoma²¹. *ZSCAN20* is associated with diabetic neuropathic pain²². rs9520257 is located between gene *ARGLU1* (Arginine And Glutamate Rich 1) and *FAM155A* (Family With Sequence Similarity 155 Member A) genes. *ARGLU1* acts in cooperation with MED1 (Mediator Complex Subunit 1) and is required for estrogen-dependent gene transcription and breast cancer cell growth²³. *FAM155A* is associated with diverticulitis²⁴. None of the genes above were related to FGF21 according to our current understanding. There might be microRNAs or long non-coding RNAs encoded in these intergenic regions affecting FGF21 levels.

SNP rs17111495, strongly associated with FGF23 levels, is located upstream of *PCSK9* (Proprotein convertase subtilisin/kexin type 9). *PCSK9* is a member of proprotein convertase family PCSK (subtilisin-like proprotein convertases previously) and is synthesized as a soluble zymogen that undergoes autocatalytic intramolecular processing in the endoplasmic reticulum²⁵. FGF23 is inactivated when being cleaved intracellularly by PCSK at the minimum consensus sequence RHTR¹⁷⁹ between Arg¹⁷⁹ and Ser¹⁸⁰^{26,27}. Autosomal dominant hypophosphatemic rickets is caused by gain of function mutations in FGF23 that renders it resistant to PCSK cleaving at site RHTR¹⁷⁹, thus resulting in elevated circulating FGF23 level and the consequent renal phosphate wasting, rickets, and osteomalacia²⁶. Hyperphosphatemic familial tumoral calcinosis due to GALNT3 mutation involves deficient O-glycosylation of the threonine residue in the RTHR¹⁷⁹ proprotein convertase-processing site, thus favoring intracellular degradation of intact FGF23 by PCSK and the resulting reduced phosphate renal excretion²⁸. Collectively, previous studies and our findings suggest that PCSK is a critical regulator of FGF23 secretion.

SNP rs17843626, significantly associated with circulating FGF23 levels, is located between *HLA-DQA1* and *HLA-DQB1* (Human Leukocyte Antigen, Class II, DQ Alpha 1 and Beta 1). In general, these HLA class II loci are known to be associated with autoimmune diseases such as type 1 diabetes, multiple sclerosis and rheumatoid arthritis²⁹. No literature has reported association between these loci and FGF23 level or its regulators, parathyroid hormone and vitamin D. This is also a new finding.

(a) Circulating levels of FGF21



(b) Circulating levels of FGF23

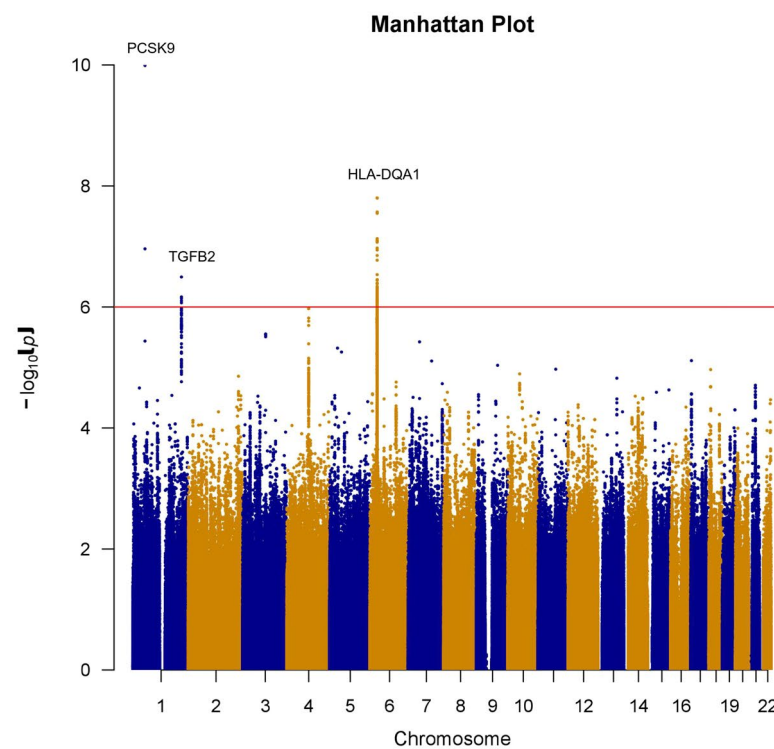
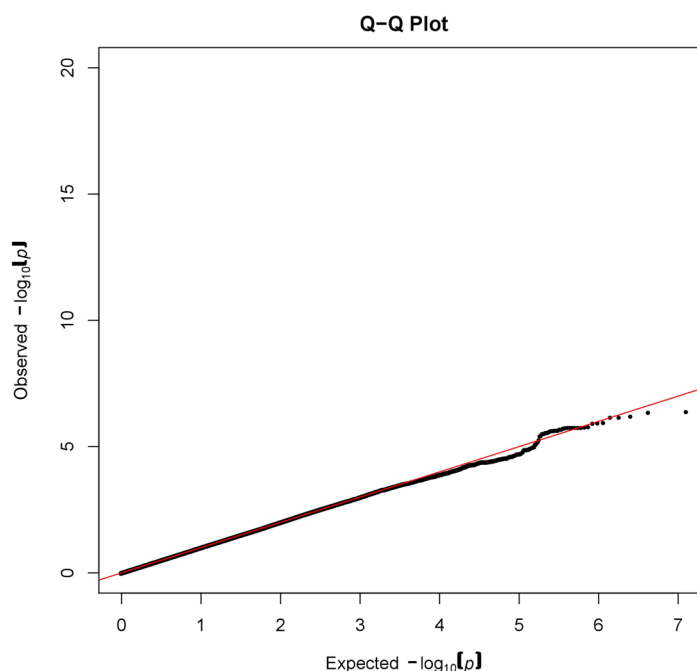
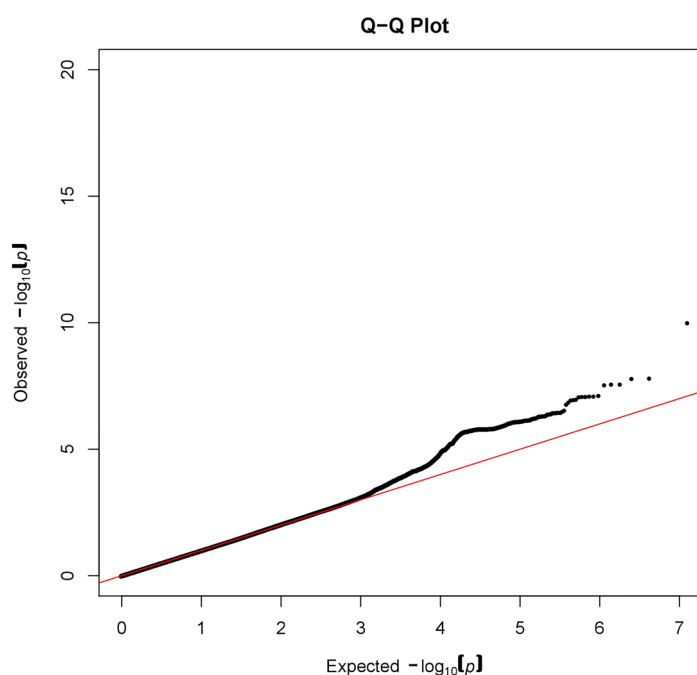


Figure 1. Manhattan plot of the GWAS results for FGF21. SNPs are plotted on the x axis according to their chromosome position against association with (a) log FGF21 or (b) log FGF23 on the y axis. The red horizontal line represents the suggestive threshold of $P = 1.0 \times 10^{-6}$.

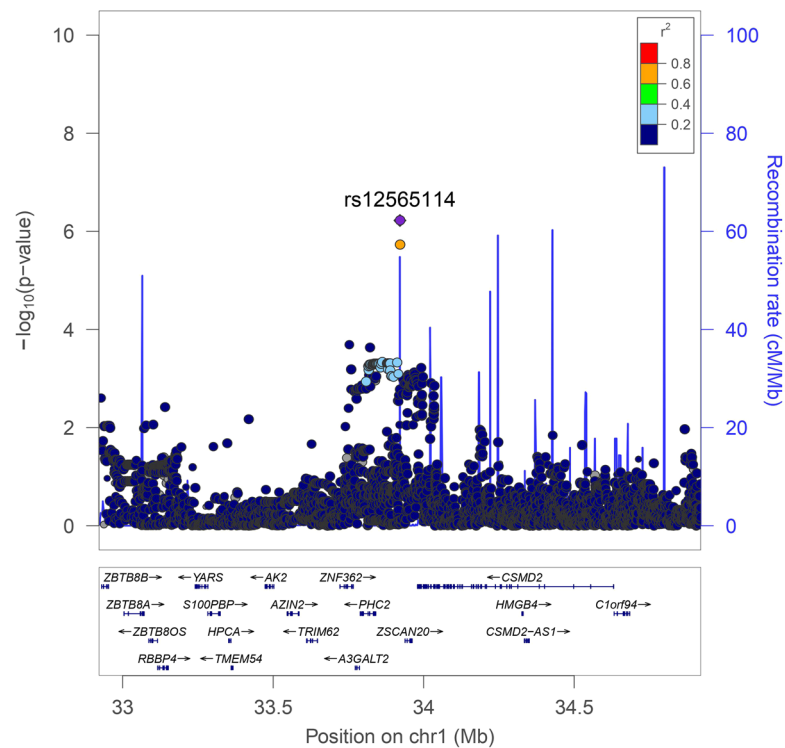
(a) Circulating FGF21 levels**(b) Circulating FGF23 levels****Figure 2.** Quantile–quantile plots of **(a)** log FGF21 and **(b)** log FGF23.

SNP rs2798631, suggestively associated with circulating FGF23 levels, is located within *TGFB2* (Transforming growth factor beta-2), which is involved in cell proliferation, differentiation, inflammation and apoptosis. A recent study has demonstrated that TGF- β 2 directly enhanced FGF23 production in rat osteoblast-like cells through increasing calcium entry³⁰. 1,25(OH) $_2$ D $_3$, the active vitamin D metabolite which induces FGF23 production, also inhibits TGF- β downstream signaling³¹. Downstream TGF- β and 1,25(OH) $_2$ D $_3$ signaling are characterized by intense crosstalk³². The expression of FGF23 by both modulators also seemed to be subjective to this crosstalk³³.

For more comprehensive bioinformatic annotations of these top SNPs, we used GRASP v2.0 to search for documented associations between these SNPs and phenotypes³⁴. One study showed association between rs12565114

doi:10.6342/NTU202401345

(a) rs12565114



(b) rs9520257

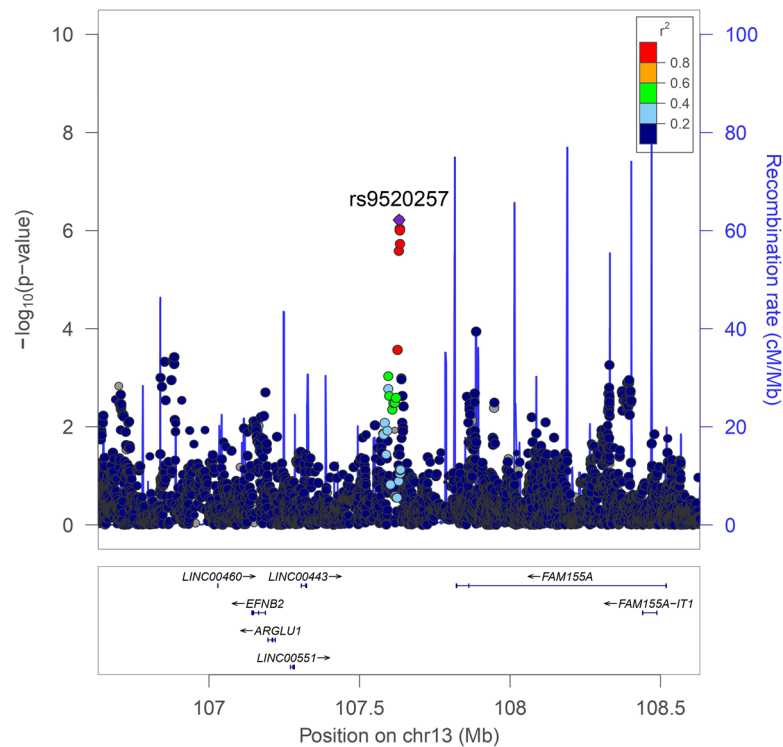
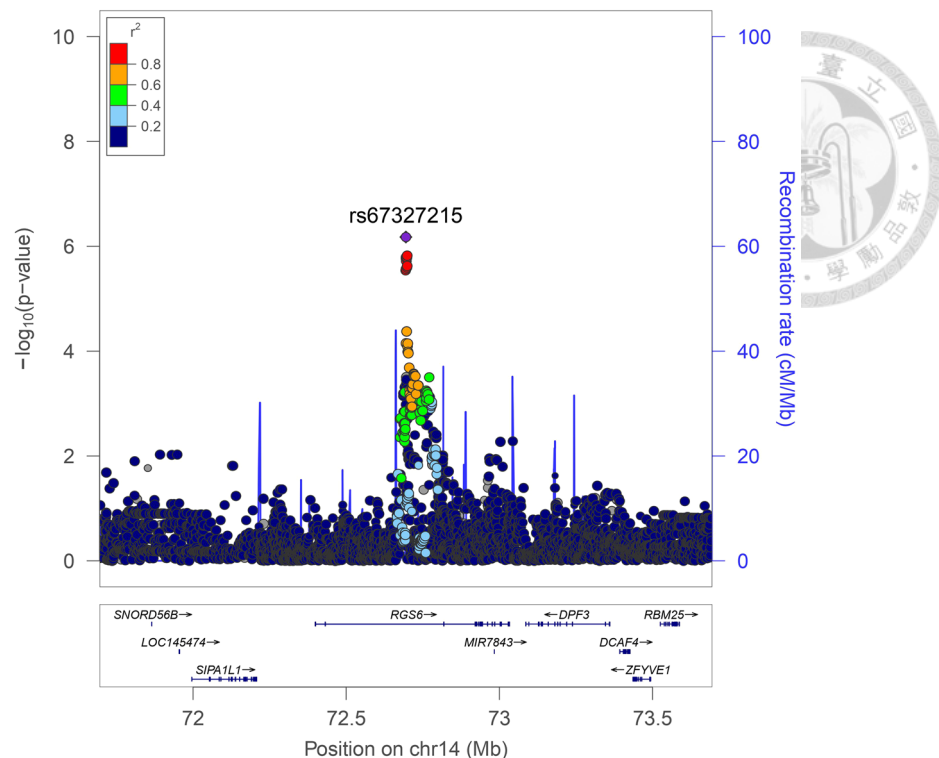


Figure 3. Regional association plots of log FGF21.

(c) rs67327215

**Figure 3.** (continued)

and schizophrenia³⁵. Several studies demonstrated association between rs17111495 and serum lipid levels^{36,37}. rs2798631 was associated with height³⁸, refractive error³⁹ and Parkinson's disease⁴⁰. There were no results for rs9520257, rs67327215 and rs17843626.

The absolute circulating FGF23 levels measured in this study were different from that obtained by Robinson-Cohen et al. This study used Duoset ELISA Development kit (R&D system, Inc.) to measure intact FGF23 levels; and the median intact FGF23 level in the studied population was 360 (110–1,740) pg/ml (expressed as median and interquartile range). For published researches using this ELISA kit, the median of intact FGF23 level was 269.9 (109.2–1,014.0) pg/ml in Chinese renal transplant donors⁴¹, and 150 (46–583) pg/ml in Mexicans with normal kidney function⁴². In contrary, Robinson-Cohen et al. used a different ELISA kit (Kainos Laboratories, Inc., Tokyo, Japan), and the population for their study was of European ancestry. Their mean intact FGF23 level was around 40–50 pg/ml¹³. The variations in FGF23 level obtained may be attributed to ethnicity of the studied populations and different ELISA kits used.

Our study has some unique strength and some limitations. It is the first known study identifying suggestive genetic determinants of circulating FGF21 level. We also evaluated SNPs associated with circulating FGF23 level from a population of ancestry different from those studied in previous GWAS, and novel candidate loci were found.

As mentioned above, our findings differed from those of Robinson-Cohen et al.¹³, possibly due to differences in ethnicity, differences in gene-environment and gene-gene interactions, and our relatively small sample size. As rs17216707 minor allele frequency was low in our cohort, gene-based association analysis was conducted as an alternative way in effort to replicate the findings but the result was not significant. A meta-analysis conducted on both this study and that of Robinson-Cohen et al. found no statistically significant results either.

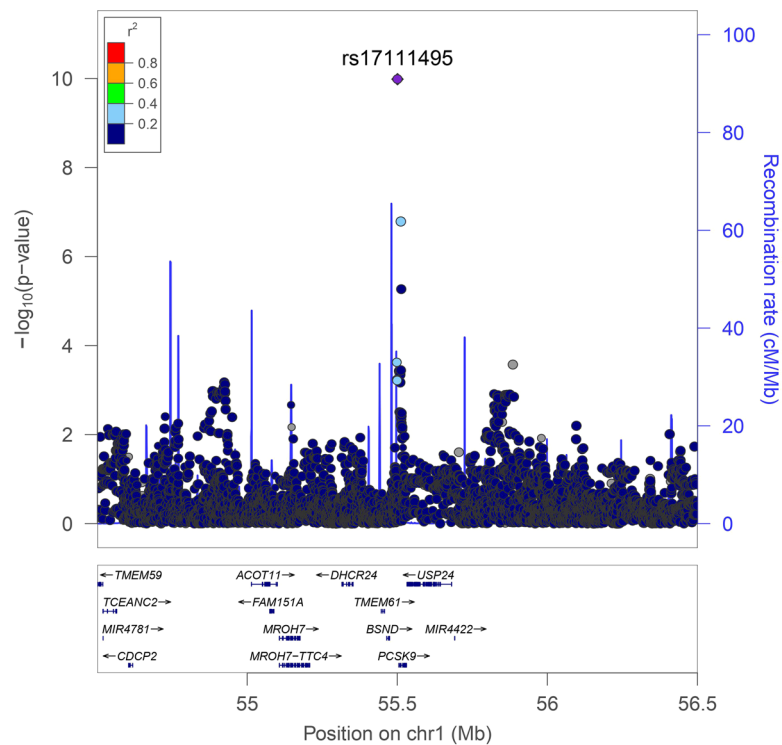
Regarding the lack of replication in an external cohort of the study findings, we searched for GWAS summary statistics of circulating FGF21 or FGF23 phenotypes in UK Biobank and Biobank Japan but found them unavailable. Further replication may be needed in the future.

In conclusion, this study is the first GWAS on circulating FGF21 level to date. Novel candidate genetic loci possibly related to circulating FGF21 and FGF23 levels were found. Further replication and functional studies are needed to support the present findings.

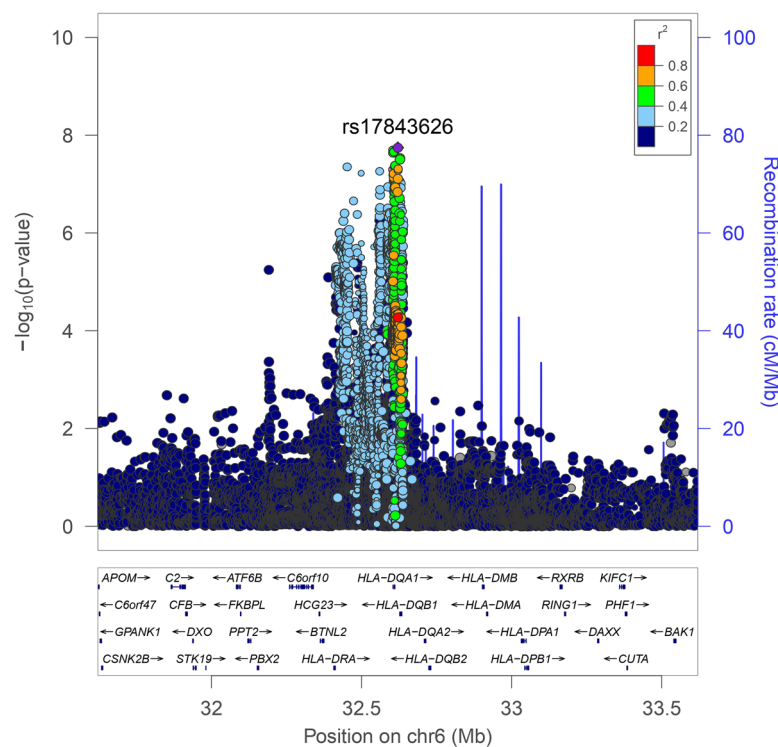
Materials and methods

Study population. Five thousand subjects were enrolled from Taiwan Biobank by random sampling as study population. Subjects are aged between 30 and 70 years, without cancer history and volunteered to join the Taiwan Biobank study population. Eating and drinking were refrained before blood samples were drawn. Delinked databases including biological specimens, personal data and clinical information were used in this investigation. The estimated glomerular filtration rate (eGFR) was calculated using the 4-variable equation from

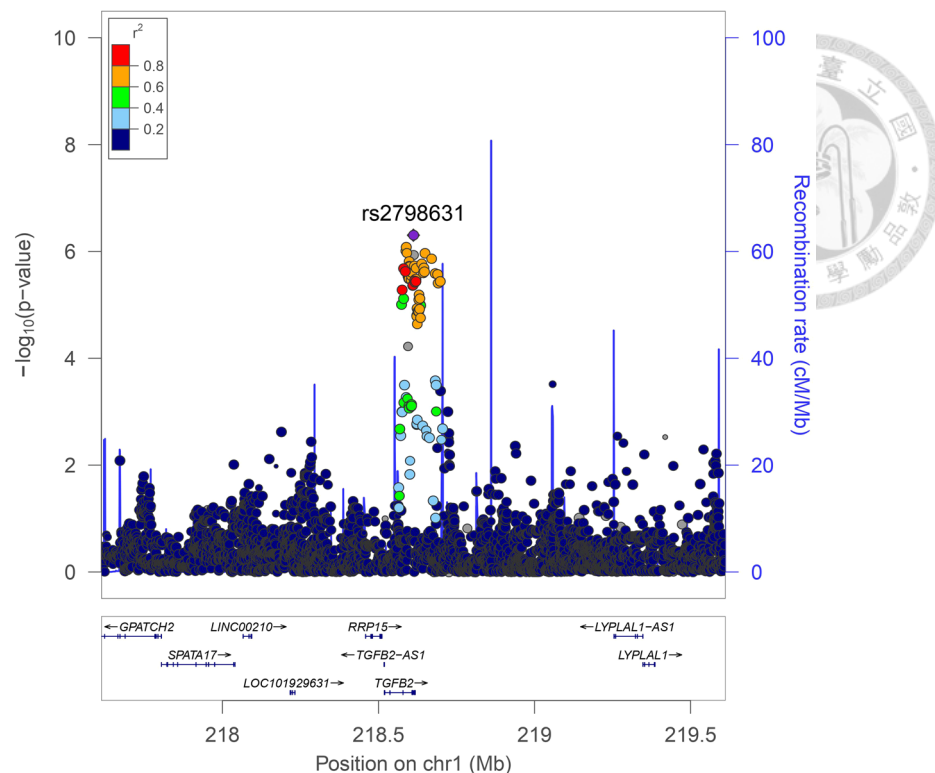
(a) rs17111495



(b) rs17843626

**Figure 4.** Regional association plots of log FGF23.

(c) rs2798631

**Figure 4.** (continued)

the Modification of Diet in Renal Disease (MDRD) Study⁴³. There were only two individuals with severe renal impairment (eGFR < 30 ml/min/1.73 m²), and their serum FGF23 levels were 0.259 and 0.709 ng/ml.

All subjects provided written informed consent. Individuals diagnosed with diabetes mellitus (either as stated in the questionnaire or HbA1c ≥ 6.5%) were excluded.

FGF21 and FGF23 measurements. Intact circulating FGF21 and FGF23 levels in serum of Taiwan Biobank subjects were measured using Enzyme-Linked Immunosorbent Assay (Duoset ELISA Development kit, R&D system, Inc.). The assays have high sensitivity and exhibit no interference or cross-reactivity with recombinant human FGF R1a/Fc Chimera, FGF R2a/Fc Chimera, FGF R3/Fc Chimera, FGF R4/Fc Chimera and Klotho. All standards by serial dilution were assayed in duplicates.

Genotyping, quality control and imputation. Genome-wide genotyping of all subjects was carried out at the National Center for Genome Medicine of Academia Sinica using the Axiom-Taiwan Biobank Array Plate (TWB chip; Affymetrix Inc, Santa Clara, California)⁴⁴. TWB chip consists of 653,291 SNPs, and was specifically customized for the Taiwanese population, who are mainly of Han-Chinese lineage, by including SNPs with detected polymorphisms in Taiwanese-based genotyping results from the Axiom Genome-Wide CHB 1 Array plate (Affymetrix Inc). SNPs from ancestry information panels, GWAS and cancer studies, as well as pharmacogenetics arrays were also incorporated into the TWB chip. PLINK (version 1.07), an open-source whole-genome association analysis toolset was used for quality control⁴⁵. Genotypes for SNPs with batch effect were set as missing. Individuals with high missing genotype rate (> 5%), with extreme heterozygosity rate (above or below 5 standard deviations of mean heterozygosity rate) or being closely related as assessed by Identity-By-Descent (IBD) estimation (IBD ≥ 0.1875) were excluded from analyses. SNPs with high missing genotype rate (> 5%), low frequency (minor allele frequency < 1%), or deviation from Hardy-Weinberg Equilibrium (HWE) (P value < 10⁻⁵) were excluded, with 617,073 SNPs remaining. Population structure was assessed using principal component analysis. Using PLINK, we computed the principal components on an LD-pruned (r^2 < 0.2) set of autosomal variants obtained by removing high-LD regions. Genome-wide genotype imputation was performed by SHAPEIT⁴⁶ and IMPUTE2⁴⁷ with 1,000 Genomes Project (1000GP) Phase 3 East Asian (EAS) population as reference panel. Quality control after imputation was as follows: We filtered the SNPs with imputation quality score of IMPUTE2 (info score⁴⁸) greater than 0.3 for further analysis. Indels were removed with VCFtools⁴⁹. SNPs with low frequency (minor allele frequency < 1%) were excluded.

Statistical analyses. Age, body mass index (BMI), glucose level and eGFR were expressed as mean and standard deviation. FGF21 and FGF23 levels were expressed as median and interquartile range. GWAS analyses were performed using an additive genetic model in PLINK v1.07. Skewed variables including serum FGF21

and FGF23 were logarithmically transformed to approximate normal distribution. We performed regression diagnostics plots to check the assumptions of linearity, normality, and homoscedasticity, as well as to identify the influential observations for each outcome of interest (logFGF21 and logFGF23) and SNPs, adjusting for covariates. We did not find any severe violations of assumptions. Further, variance inflation factor (VIF) was used to check for multicollinearity. No correlations between independent variables were found. Linear regression was employed to analyze association between SNPs and log-transformed FGF21 and FGF23 levels. Population stratification was not obvious ($\lambda = 1.002$ for FGF21, $\lambda = 0.999$ for FGF23). Age, sex and the first ten principal components of ancestry were adjusted in Model 1, and BMI was additionally adjusted in Model 2. For log FGF23, BMI and eGFR were additionally adjusted in Model 3. The threshold for genome-wide significance was set at $P = 5 \times 10^{-850}$. In view of the relatively small sample size and to avoid false negative results caused by too stringent threshold⁵¹, the P value threshold for suggestive results were set at 1.0×10^{-652} . Manhattan plots and quantile–quantile plots were drawn using the qqman R package^{53,54}. Regional association plots were generated by LocusZoom⁵⁵. 1,000 Genomes Project Phase 3 East Asian Ancestry was used for the reference population and Genome Reference Consortium Human Build 37 was used for gene annotation. The array-based heritability was calculated by linkage disequilibrium (LD) score regression⁵⁶. The proportion of phenotypic variance explained by single SNP was estimated by the following formula:

$$\frac{2\beta^2 * \text{MAF} * (1 - \text{MAF})}{2\beta^2 * \text{MAF} * (1 - \text{MAF}) + \text{SE}^2 * 2N * \text{MAF} * (1 - \text{MAF})}$$

where β , SE, N, and MAF are the effect size estimate of each minor allele on the relative concentration of FGF21/FGF23, standard error of the effect size, sample size, and minor allele frequency for the SNP, respectively⁵⁷. Gene-based association analysis was performed using MAGMA v1.07b¹⁴ with LD information retrieved from 1,000 Genomes Project Phase 3 East Asian Panel. Meta-analysis pooling our study and those reported by Robinson-Cohen et al. was performed using by METAL using a random effects model¹⁵.

Bioethics statement. All methods were carried out in accordance with relevant guidelines and regulations. The experimental protocols were approved by the Institutional Review Board of National Taiwan University Hospital.

Data availability

Individual researchers may request to use the data for specific projects on a collaborative basis.

Received: 24 November 2019; Accepted: 6 August 2020

Published online: 03 September 2020

References

1. Degirolamo, C., Sabba, C. & Moschetta, A. Therapeutic potential of the endocrine fibroblast growth factors FGF19, FGF21 and FGF23. *Nat. Rev. Drug Discov.* **15**, 51–69. <https://doi.org/10.1038/nrd.2015.9> (2016).
2. Badman, M. K. *et al.* Hepatic fibroblast growth factor 21 is regulated by PPARalpha and is a key mediator of hepatic lipid metabolism in ketotic states. *Cell Metab.* **5**, 426–437. <https://doi.org/10.1016/j.cmet.2007.05.002> (2007).
3. Bookout, A. L. *et al.* FGF21 regulates metabolism and circadian behavior by acting on the nervous system. *Nat. Med.* **19**, 1147–1152. <https://doi.org/10.1038/nm.3249> (2013).
4. Chartoumpekis, D. V. *et al.* Brown adipose tissue responds to cold and adrenergic stimulation by induction of FGF21. *Mol. Med.* **17**, 736–740. <https://doi.org/10.2119/molmed.2011.00075> (2011).
5. Wenthe, W. *et al.* Fibroblast growth factor-21 improves pancreatic beta-cell function and survival by activation of extracellular signal-regulated kinase 1/2 and Akt signaling pathways. *Diabetes* **55**, 2470–2478. <https://doi.org/10.2337/db05-1435> (2006).
6. Owen, B. M. *et al.* FGF21 acts centrally to induce sympathetic nerve activity, energy expenditure, and weight loss. *Cell Metab.* **20**, 670–677. <https://doi.org/10.1016/j.cmet.2014.07.012> (2014).
7. Dutchak, P. A. *et al.* Fibroblast growth factor-21 regulates PPARgamma activity and the antidiabetic actions of thiazolidinediones. *Cell* **148**, 556–567. <https://doi.org/10.1016/j.cell.2011.11.062> (2012).
8. Kharitononkov, A. & Adams, A. C. Inventing new medicines: The FGF21 story. *Mol. Metab.* **3**, 221–229. <https://doi.org/10.1016/j.molmet.2013.12.003> (2014).
9. Kanbay, M. *et al.* Novel faces of fibroblast growth factor 23 (FGF23): iron deficiency, inflammation, insulin resistance, left ventricular hypertrophy, proteinuria and acute kidney injury. *Calcif. Tissue Int.* **100**, 217–228. <https://doi.org/10.1007/s00223-016-0206-7> (2017).
10. Angelin, B., Larsson, T. E. & Rudling, M. Circulating fibroblast growth factors as metabolic regulators—a critical appraisal. *Cell Metab.* **16**, 693–705. <https://doi.org/10.1016/j.cmet.2012.11.001> (2012).
11. Mirza, M. A. *et al.* Circulating fibroblast growth factor-23 is associated with fat mass and dyslipidemia in two independent cohorts of elderly individuals. *Arterioscler. Thromb. Vasc. Biol.* **31**, 219–227. <https://doi.org/10.1161/ATVBAHA.110.214619> (2011).
12. Kinoshita, Y. & Fukumoto, S. X-linked hypophosphatemia and FGF23-related hypophosphatemic diseases: prospect for new treatment. *Endocr. Rev.* **39**, 274–291. <https://doi.org/10.1210/er.2017-00220> (2018).
13. Robinson-Cohen, C. *et al.* Genetic variants associated with circulating fibroblast growth factor 23. *J. Am. Soc. Nephrol.* **29**, 2583–2592. <https://doi.org/10.1681/ASN.2018020192> (2018).
14. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219. <https://doi.org/10.1371/journal.pcbi.1004219> (2015).
15. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191. <https://doi.org/10.1093/bioinformatics/btq340> (2010).
16. Chatterjee, T. K., Liu, Z. & Fisher, R. A. Human RGS6 gene structure, complex alternative splicing, and role of N terminus and G protein gamma-subunit-like (GGL) domain in subcellular localization of RGS6 splice variants. *J. Biol. Chem.* **278**, 30261–30271. <https://doi.org/10.1074/jbc.M212687200> (2003).
17. Uhlen, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419. <https://doi.org/10.1126/science.1260419> (2015).

18. Dohlman, H. G. & Thorner, J. RGS proteins and signaling by heterotrimeric G proteins. *J. Biol. Chem.* **272**, 3871–3874. <https://doi.org/10.1074/jbc.272.7.3871> (1997).
19. Pashkov, V. et al. Regulator of G protein signaling (RGS16) inhibits hepatic fatty acid oxidation in a carbohydrate response element-binding protein (ChREBP)-dependent manner. *J. Biol. Chem.* **286**, 15116–15125. <https://doi.org/10.1074/jbc.M110.216234> (2011).
20. Stewart, A. et al. Regulator of G protein signaling 6 is a critical mediator of both reward-related behavioral and pathological responses to alcohol. *Proc. Natl. Acad. Sci. USA* **112**, E786–795. <https://doi.org/10.1073/pnas.1418795112> (2015).
21. Perry, K. D. et al. Composite hemangioendothelioma with neuroendocrine marker expression: an aggressive variant. *Mod. Pathol.* **30**, 1589–1602. <https://doi.org/10.1038/modpathol.2017.83> (2017).
22. Meng, W., Deshmukh, H. A. & Donnelly, L. A. A genome-wide association study provides evidence of sex-specific involvement of Chr1p35.1 (ZSCAN20-TLR12P) and Chr8p23.1 (HMGB1P46) With Diabetic Neuropathic Pain. *EbioMedicine* **2**, 1386–1393. <https://doi.org/10.1016/j.ebiom.2015.08.001> (2015).
23. Zhang, D., Jiang, P., Xu, Q. & Zhang, X. Arginine and glutamate-rich 1 (ARGLU1) interacts with mediator subunit 1 (MED1) and is required for estrogen receptor-mediated gene transcription and breast cancer cell growth. *J. Biol. Chem.* **286**, 17746–17754. <https://doi.org/10.1074/jbc.M110.206029> (2011).
24. Sigurdsson, S. et al. Sequence variants in ARHGAP15, COLQ and FAM155A associate with diverticular disease and diverticulitis. *Nat. Commun.* **8**, 15789. <https://doi.org/10.1038/ncomms15789> (2017).
25. Lagace, T. A. PCSK9 and LDLR degradation: regulatory mechanisms in circulation and in cells. *Curr. Opin. Lipidol.* **25**, 387–393. <https://doi.org/10.1097/MOL.0000000000000114> (2014).
26. White, K. E. et al. Autosomal-dominant hypophosphatemic rickets (ADHR) mutations stabilize FGF-23. *Kidney Int.* **60**, 2079–2086. <https://doi.org/10.1046/j.1523-1755.2001.00064.x> (2001).
27. Benet-Pages, A. et al. FGF23 is processed by proprotein convertases but not by PHEX. *Bone* **35**, 455–462. <https://doi.org/10.1016/j.bone.2004.04.002> (2004).
28. Kato, K. et al. Polypeptide GalNAc-transferase T3 and familial tumoral calcinosis. Secretion of fibroblast growth factor 23 requires O-glycosylation. *J. Biol. Chem.* **281**, 18370–18377. <https://doi.org/10.1074/jbc.M602469200> (2006).
29. Jones, E. Y., Fugger, L., Strominger, J. L. & Siebold, C. MHC class II proteins and disease: a structural perspective. *Nat. Rev. Immunol.* **6**, 271–282. <https://doi.org/10.1038/nri1805> (2006).
30. Feger, M. et al. The production of fibroblast growth factor 23 is controlled by TGF-beta2. *Sci. Rep.* **7**, 4982. <https://doi.org/10.1038/s41598-017-05226-y> (2017).
31. Qiu, T. et al. TGF-beta type II receptor phosphorylates PTH receptor to integrate bone remodelling signalling. *Nat. Cell. Biol.* **12**, 224–234. <https://doi.org/10.1038/ncb2022> (2010).
32. Zerr, P. et al. Vitamin D receptor regulates TGF-beta signalling in systemic sclerosis. *Ann. Rheum. Dis.* **74**, e20. <https://doi.org/10.1136/annrheumdis-2013-204378> (2015).
33. Subramaniam, N. et al. Cross-talk between 1,25-dihydroxyvitamin D3 and transforming growth factor-beta signaling requires binding of VDR and Smad3 proteins to their cognate DNA recognition elements. *J. Biol. Chem.* **276**, 15741–15746. <https://doi.org/10.1074/jbc.M011033200> (2001).
34. Leslie, R., O'Donnell, C. J. & Johnson, A. D. GRASP: analysis of genotype-phenotype results from 1390 genome-wide association studies and corresponding open access database. *Bioinformatics* **30**, i185–i194. <https://doi.org/10.1093/bioinformatics/btu273> (2014).
35. Need, A. C. et al. A genome-wide investigation of SNPs and CNVs in schizophrenia. *PLoS Genet.* **5**, e1000373. <https://doi.org/10.1371/journal.pgen.1000373> (2009).
36. Kathiresan, S. et al. Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat. Genet.* **41**, 56–65. <https://doi.org/10.1038/ng.291> (2009).
37. Teslovich, T. M. et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713. <https://doi.org/10.1038/nature09270> (2010).
38. Lango Allen, H. et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832–838. <https://doi.org/10.1038/nature09410> (2010).
39. Stambolian, D. et al. Meta-analysis of genome-wide association studies in five cohorts reveals common variants in RBFOX1, a regulator of tissue-specific splicing, associated with refractive error. *Hum. Mol. Genet.* **22**, 2754–2764. <https://doi.org/10.1093/hmg/ddt116> (2013).
40. Pankratz, N. et al. Meta-analysis of Parkinson's disease: identification of a novel locus, RIT2. *Ann. Neurol.* **71**, 370–384. <https://doi.org/10.1002/ana.22687> (2012).
41. Deng, G. et al. The value of older donors' klotho level in predicting recipients' short-term renal function. *Med. Sci. Monit.* **24**, 7936–7943. <https://doi.org/10.12659/MSM.913274> (2018).
42. Farias-Basulto, A. et al. Circulating levels of soluble klotho and fibroblast growth factor 23 in diabetic patients and its association with early nephropathy. *Arch. Med. Res.* **49**, 451–455. <https://doi.org/10.1016/j.arcmed.2019.01.008> (2018).
43. Levey, A. S. et al. Using standardized serum creatinine values in the modification of diet in renal disease study equation for estimating glomerular filtration rate. *Ann. Intern. Med.* **145**, 247–254. <https://doi.org/10.7326/0003-4819-145-4-200608150-00004> (2006).
44. Chen, C. H. et al. Population structure of Han Chinese in the modern Taiwanese population based on 10,000 participants in the Taiwan Biobank project. *Hum. Mol. Genet.* **25**, 5321–5331. <https://doi.org/10.1093/hmg/ddw346> (2016).
45. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575. <https://doi.org/10.1086/519795> (2007).
46. Delaneau, O., Zagury, J. F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* **10**, 5–6. <https://doi.org/10.1038/nmeth.2307> (2013).
47. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* **44**, 955–959. <https://doi.org/10.1038/ng.2354> (2012).
48. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.* **11**, 499–511. <https://doi.org/10.1038/nrg2796> (2010).
49. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330> (2011).
50. Peér, I., Yelensky, R., Altshuler, D. & Daly, M. J. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet. Epidemiol.* **32**, 381–385. <https://doi.org/10.1002/gepi.20303> (2008).
51. Fallin, M. D. & Kao, W. H. Is, “X”-WAS the future for all of epidemiology? *Epidemiology* **22**, 457–459. <https://doi.org/10.1097/EDE.0b013e31821d3a9f> (2011).
52. Duggal, P., Gillanders, E. M., Holmes, T. N. & Bailey-Wilson, J. E. Establishing an adjusted p-value threshold to control the family-wide type 1 error in genome wide association studies. *BMC Genom.* **9**, 516. <https://doi.org/10.1186/1471-2164-9-516> (2008).
53. Turner, S. *qqman: An R Package for Visualizing GWAS Results Using Q-Q and Manhattan Plots*. (2014).
54. Team, R. D. C. R. *A Language and Environment for Statistical Computing*. (2008).
55. Pruim, R. J. et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337. <https://doi.org/10.1093/bioinformatics/btq419> (2010).

56. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295. <https://doi.org/10.1038/ng.3211> (2015).
57. Shim, H. *et al.* A multivariate genome-wide association analysis of 10 LDL subfractions, and their response to statin treatment, in 1868 Caucasians. *PLoS ONE* **10**, e0120758. <https://doi.org/10.1371/journal.pone.0120758> (2015).

Acknowledgements

This work is supported by grants from the Ministry of Science and Technology in Taiwan (MOST 105-2314-B-182-062, MOST 106-2314-B-182-043), the Translational Medical Research Program of Academia Sinica (AS-TM-108-01-04), National Taiwan University Hospital, Yunlin Branch (NTUHYL106.X003) and Chang Gung University, Taoyuan, Taiwan (NMRPD1F154, NMRPD1G0711, and BMRPD08). The authors would also like to thank the Data Science Statistical Cooperation Center of Academia Sinica (AS-CFII-108-117) for statistical support. The manuscript has been edited by the English editing service of K. T. Li Foundation for Development of Science and Technique, Taiwan (no. 2804_20200708121153).

Author contributions

G.T.C., P.H.L. and Y.C.C. conceived and designed the study. Analysis was done by G.T.C., P.H.L., C.R.C., C.N.H. and C.Y.S. Sample procurement and data generation was done by G.T.C., P.H.L., T.W.C., C.H.H., Y.Y.H., C.H.L., J.W.L., C.N.H., R.Y.T., M.L.H., H.L.L., W.S.Y. and Y.C.C. Manuscript writing was done by G.T.C., P.H.L. and Y.C.C.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-71569-8>.

Correspondence and requests for materials should be addressed to Y.-C.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020

OPEN

A genome-wide association study for melatonin secretion

Pi-Hua Liu^{1,2,16}, Gwo-Tsann Chuang^{3,4,16}, Chia-Ni Hsiung⁵, Wei-Shun Yang^{4,6}, Hsiao-Chia Ku⁷, Yi-Ching Lin⁷, Yi-Shun Chen⁷, Yu-Yao Huang^{2,8}, Chia-Hung Lin^{2,9}, Wen-Yi Li¹⁰, Jou-Wei Lin¹¹, Chih-Neng Hsu¹¹, Juey-Jen Hwang^{11,12}, Karen Chia-Wen Liao¹³, Meng-Lun Hsieh¹², Hsiao-Lin Lee¹², Chen-Yang Shen^{14,15} & Yi-Cheng Chang^{4,12,14}✉

Melatonin exerts a wide range of effects among various tissues and organs. However, there is currently no study to investigate the genetic determinants of melatonin secretion. Here, we conducted a genome-wide association study (GWAS) for melatonin secretion using morning urine 6-hydroxymelatonin sulfate-to-creatinine ratio (UMCR). We initially enrolled 5000 participants from Taiwan Biobank in this study. After excluding individuals that did not have their urine collected in the morning, those who had history of neurological or psychiatric disorder, and those who failed to pass quality control, association of single nucleotide polymorphisms with log-transformed UMCR adjusted for age, sex and principal components of ancestry were analyzed. A second model additionally adjusted for estimated glomerular filtration rate (eGFR). A total of 2373 participants underwent the genome-wide analysis. Five candidate loci associated with log UMCR (P value ranging from 6.83×10^{-7} to 3.44×10^{-6}) encompassing *ZFHX3*, *GALNT15*, *GALNT13*, *LDLRAD3* and intergenic between *SEPP1* and *FLJ32255* were identified. Similar results were yielded with further adjustment for eGFR. Interestingly, the identified genes are associated with circadian behavior, neuronal differentiation, motor disorders, anxiety, and neurodegenerative diseases. We conducted the first GWAS for melatonin secretion and identified five candidate genetic loci associated with melatonin level. Replication and functional studies are needed in the future.

Melatonin is a pleiotropic hormone primarily synthesized and secreted from the pineal gland. Many other tissues can also produce it, including leukocytes, bone marrow, gastrointestinal tract, neuronal cells, and gonads^{1–3}. Melatonin regulates various physiological processes, including circadian and seasonal rhythms, energy and glucose metabolism, antioxidant effects, anti-inflammatory actions, and immune function^{1,3–6}. There are many studies showing associations between melatonin and many disorders, including certain types of mental illness, cancer, cardiovascular disease, metabolic syndrome, type 2 diabetes, and obesity^{6–11}. Melatonin is secreted into the circulation following a circadian rhythm with peak levels at night¹². Aging was once thought to be directly associated with decreased melatonin secretion. However, there was no significant difference between circadian amplitude of the plasma melatonin between healthy elderlies and young adults¹³. Instead of aging, the degree of pineal calcification was associated with melatonin excretion amount¹⁴.

¹Clinical Informatics and Medical Statistics Research Center, College of Medicine, Chang Gung University, Taoyüan, Taiwan. ²Division of Endocrinology and Metabolism, Department of Internal Medicine, Chang Gung Memorial Hospital at Linkou, Taoyüan, Taiwan. ³Department of Pediatrics, College of Medicine, National Taiwan University Hospital, National Taiwan University, Taipei, Taiwan. ⁴Graduate Institute of Medical Genomics and Proteomics, National Taiwan University, 5F, No.2, Xuzhou Rd., Zhongzheng Dist., Taipei 100, Taiwan. ⁵Data Science Statistical Cooperation Center, Institute of Statistical Science, Academia Sinica, Taipei, Taiwan. ⁶Department of Internal Medicine, National Taiwan University Hospital Hsin-Chu Branch, Hsinchu, Taiwan. ⁷Department of Laboratory Medicine, National Taiwan University Hospital Hsin-Chu Branch, Hsinchu, Taiwan. ⁸Department of Medical Nutrition Therapy, Lin-Kou Chang-Gung Memorial Hospital, Chang-Gung Medical Foundation, Taoyüan, Taiwan. ⁹Department of Chinese Medicine, College of Medicine, Chang Gung University, Taoyüan, Taiwan. ¹⁰Department of Internal Medicine, National Taiwan University Hospital Yunlin Branch, Yunlin, Taiwan. ¹¹Cardiovascular Center, National Taiwan University Hospital Yun-Lin Branch, Yunlin, Taiwan. ¹²Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan. ¹³Biological Sciences Division, University of Chicago, Chicago, IL 60637, USA. ¹⁴Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan. ¹⁵College of Public Health, China Medical University, Taichung, Taiwan. ¹⁶These authors contributed equally: Pi-Hua Liu and Gwo-Tsann Chuang. ✉email: b83401040@gmail.com

doi:10.6342/NTU202401345

Substantial evidence suggests genetic factors also play a significant role in melatonin secretion^{15,16}. Genome-wide association study (GWAS) has been introduced as a powerful tool to identify common genetic variants of complex diseases or quantitative traits¹⁷. Currently, there is no published GWAS regarding melatonin levels. Here, we conducted the first GWAS of urine melatonin metabolite, 6-hydroxymelatonin sulfate (aMT6s), which surrogate the circulating melatonin level¹⁸.

Materials and methods

Study population. Five thousand individuals aged 30–70 years old and without cancer history were enrolled from Taiwan Biobank. Biological specimens, personal and clinical information as delinked data were used in this study. Individuals with a record of neurological disorders or psychiatric illnesses were excluded from this study as these conditions may affect melatonin secretion^{19,20}. This study was approved by the Institutional Review Board of Chang Gung Medical Foundation and the Institutional Review Board of National Taiwan University Hospital. All subjects have provided written informed consent, and all methods were carried out in accordance with relevant guidelines and regulations.

Urine aMT6s and creatinine measurement. It is infeasible to draw blood samples from volunteers in the middle of the night for serum melatonin levels. Urinary aMT6s is the major metabolite of melatonin excreted from the kidneys²¹. Thus, measuring morning urine aMT6s level is a practical alternative for serum melatonin level at night¹⁸. For better correlation, spot urine aMT6s level should be creatinine-corrected to adjust the effect of variable urinary dilution²². Urine aMT6s-to-creatinine ratio (UMCR) was calculated from urinary aMT6s divided by urine creatinine level. The concentration of aMT6s was measured in the urine of Taiwan Biobank subjects by an enzyme-linked immunosorbent assay (ELISA) kit using the manufacturer's protocol (Human Melatonin Sulfate ELISA kit, Elab science). No significant cross-reactivity or interference between melatonin sulfate and analogs was observed. All standards via serial dilution were assayed in duplicates. The urine creatinine level was measured using a chemistry analyzer (AU5800, Beckman Coulter) with compensated Jaffe method.

Genotyping, quality control and imputation. Genotyping with the Axiom-Taiwan Biobank Array Plate (TWB chip; Affymetrix Inc, Santa Clara, California) was performed at the National Center for Genome Medicine of Academia Sinica²³. We use PLINK (version 1.9), an open-source whole-genome data analysis tool-set, for quality control procedures²⁴. For SNPs with batch effect, their genotypes were set as missing. SNPs were excluded if missing genotype rate was high (>5%), minor allele frequency was low (<1%) or deviated from Hardy–Weinberg equilibrium (P value < 10^{-5}). Individuals with discordant sex (self-reported sex incongruent to genetic sex, where genetic male or female was defined by X chromosome homozygosity estimate above 0.8 or below 0.2), high missing genotyping rate (>5%), extreme heterozygosity rate (more than 5 standard deviations away from the mean) or high identity-by-descent score (≥ 0.1875) implying close relatedness were excluded from subsequent analyses. We computed the principal components on a linkage disequilibrium (LD)-pruned ($r^2 < 0.2$) set of autosomal variants obtained by removing high-LD regions via PLINK. Genotype imputation was carried out with SHAPEIT²⁵ and IMPUTE2²⁶. We applied 1000 Genomes Project Phase 3 East Asian Ancestry as the reference population. For gene annotation, Genome Reference Consortium Human Build 37 was used. Imputed SNPs with low-quality scores (info score²⁷ lower than 0.8) were excluded. Indels were removed by using VCFtools²⁸.

Statistical analyses. Age and estimated glomerular filtration rate (eGFR) were expressed as mean and standard deviation. Urine aMT6s and UMCR were expressed as median and interquartile range. Logarithmic transformation of UMCR was done to normalize the data. GWAS analysis was carried out via PLINK v1.9 with an additive genetic model. We applied linear regression for analyzing associations between SNPs and log UMCR. Covariate adjustment in Model 1 included age, sex, and the first ten principal components of ancestry. eGFR was additionally adjusted in Model 2. We used a genome-wide significance threshold of $P < 5.0 \times 10^{-8}$ ²⁹. Since this threshold is very conservative for small sample size, we set the level for suggestive significance at $P < 5 \times 10^{-6}$ ^{30,31}. The Manhattan plot and quantile–quantile plot were generated by the qqman R package^{32,33}. Regional association plots were made via LocusZoom³⁴. The proportion of phenotypic variance explained by SNP was calculated using the following items: effect size estimate of each minor allele on log UMCR, standard error of the effect size, sample size, and minor allele frequency for the SNP³⁵. The statistical power of this study was calculated using methods for quantitative GWAS³⁶.

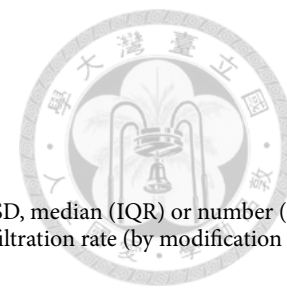
Bioethics statement. This study was approved by the Institutional Review Board of Chang Gung Medical Foundation and the Institutional Review Board of National Taiwan University Hospital. All subjects have provided written informed consent and all methods were carried out in accordance with relevant guidelines and regulations.

Results

Five thousand subjects were enrolled from Taiwan Biobank initially. One withdrew from the study. 2361 did not have their urine collected in the morning and were excluded. 128 people had documented neurological or psychiatric illness. 137 individuals did not pass quality control procedures. After imputation and quality control, 7,897,704 autosomal SNPs remained. After log transformation of UMCR, data is still not normalized, but the shape of the histogram is better. We performed a GWAS analysis for log UMCR in the remaining 2373 subjects. The characteristics of our study population are listed in Table 1. Age is not significantly associated with

Characteristics	
Total participants, N	2373
Age, year	50.75 ± 10.83
Males, N (%)	890 (37.51)
eGFR, ml/min/1.73 m ²	108.20 ± 27.83
Urine aMT6s, ng/ml	20.41 (11.88–30.19)
UMCR, ng/mg	16.98 (10.32–27.33)

Table 1. Descriptive characteristics of study subjects. Data are mean ± SD, median (IQR) or number (%), as appropriate. Age is at specimen collection. *eGFR* estimated glomerular filtration rate (by modification of diet in renal disease equation), *UMCR* urine aMT6s/creatinine ratio.



SNP	Chr	Position	Nearest gene	UMCR increasing allele	Other allele	UMCR increasing allele frequency	Model 1		Model 2	
							P value	Beta (SEM)	P value	Beta (SEM)
rs17681554	16	73,016,768	ZFHX3	A	C	0.804	6.86×10^{-7}	0.068 (0.014)	6.83×10^{-7}	0.068 (0.014)
rs142037747	3	16,121,712	GALNT15	G	A	0.989	7.82×10^{-7}	0.256 (0.052)	7.73×10^{-7}	0.256 (0.052)
rs7571016	2	155,166,873	GALNT13	A	G	0.617	1.53×10^{-6}	0.056 (0.011)	1.54×10^{-6}	0.056 (0.012)
rs9645614	11	36,159,947	LDLRAD3	A	G	0.953	2.90×10^{-6}	0.124 (0.026)	2.91×10^{-6}	0.124 (0.026)
rs6451653	5	42,915,584	SEPP1-FLJ32255	G	A	0.869	3.44×10^{-6}	0.080 (0.017)	3.42×10^{-6}	0.080 (0.017)

Table 2. Association of genetic loci with log UMCR in a Taiwan Han Chinese population. Model 1 was adjusted for age, sex and the first ten principal components of ancestry. Model 2 was additionally adjusted for eGFR. *SNP* single nucleotide polymorphism, *Chr* chromosome. SNPs are imputed with high info score (0.831, 0.988 and 0.946 for rs142037747, rs9645614 and rs6451653, respectively).

log UMCR (Pearson's $r = 0.031$; $P = 0.137$). There is no significant gender difference of log UMCR shown by T test (males 1.208, females 1.205; $P = 0.857$). eGFR is also not significantly associated with log UMCR (Pearson's $r = 0.001$; $P = 0.973$). Variants with the strongest association in each region regarding to log UMCR are shown in Table 2. Melatonin production is known to be decreased with advanced chronic kidney disease³⁷; thus, we adjusted eGFR additionally in Model 2. There was no evidence of liver disorders in these 2373 individuals, and thus no additional model was generated. Figures 1 and 2 are the Manhattan plot and quantile–quantile plot. The genomic inflation factor, defined as the median of the observed chi-squared test statistics divided by the expected median of the corresponding chi-squared distribution, was 1.006. Regional association plots of the top SNPs are shown in Fig. 3. Five loci showed suggestive significance, with one within the *ZFHX3* gene on chromosome 16 (rs17681554; $P = 6.83 \times 10^{-7}$, Fig. 3a), another near the *GALNT15* gene on chromosome 3 (rs142037747; $P = 7.82 \times 10^{-7}$, Fig. 3b), a third within the *GALNT13* gene on chromosome 2 (rs7571016; $P = 1.53 \times 10^{-6}$, Fig. 3c), a fourth within the *LDLRAD3* gene on chromosome 11 (rs9645614; $P = 2.90 \times 10^{-6}$, Fig. 3d), and the other between *SEPP1* and *FLJ32255* on chromosome 5 (rs6451653; $P = 3.44 \times 10^{-6}$, Fig. 3e).

Proportion of variance explained by the individual SNPs are 0.98%, 1.01%, 1.08%, 0.95% and 0.92% for rs17681554, rs142037747, rs7571016, rs9645614 and rs6451653 respectively. The calculated power of this GWAS was 56%.

Discussion

In this first GWAS on melatonin secretion, we identified five suggestive loci associated with variation in log UMCR. rs17681554 is located within *ZFHX3* (Zinc Finger Homeobox 3). *ZFHX3* is a transcriptional regulator which contains four homeodomains and seventeen zinc fingers³⁸. During neuronal differentiation, there is a preferential expression pattern of *ZFHX3* isoforms³⁹. In addition, circadian behavior alteration is shown in inducible conditional *Zfmx3* knockout adult mice⁴⁰. Further studies are needed to elucidate a direct linkage between *ZFHX3* and melatonin.

rs142037747 and rs7571016 are located near *GALNT15* (polypeptide N-acetylgalactosaminyltransferase 15) and within *GALNT13* (polypeptide N-acetylgalactosaminyltransferase 13), respectively. These two polypeptide N-acetylgalactosaminyltransferases of the same family catalyze initiation of mucin-type O-linked glycosylation by adding N-acetylgalactosamine to serine or threonine residues of the polypeptide chain^{41,42}. Glycosylation is associated with cell adhesion, signal transduction, molecular trafficking, and differentiation in central nervous system development⁴³. Whether and how *GALNT15* or *GALNT13* significantly affects melatonin levels remains determined.

rs9645614 is located within *LDLRAD3* (low density lipoprotein receptor class A domain containing 3). *LDLRAD3* alters the proteolysis of amyloid precursor protein and increases the production of amyloid beta-peptide (Aβ)⁴⁴. The primary pathogenesis of Alzheimer's disease (AD) has been attributed to the extracellular aggregation of Aβ⁴⁵. Melatonin levels are altered in AD patients, possibly due to decreased suprachiasmatic nucleus cell number and functional pineal gland volume⁴⁶. Patients with neurodegenerative disorder such as Alzheimer's disease exhibit reduced serum and cerebrospinal fluid melatonin levels comparing to age-matched controls^{47,48}.

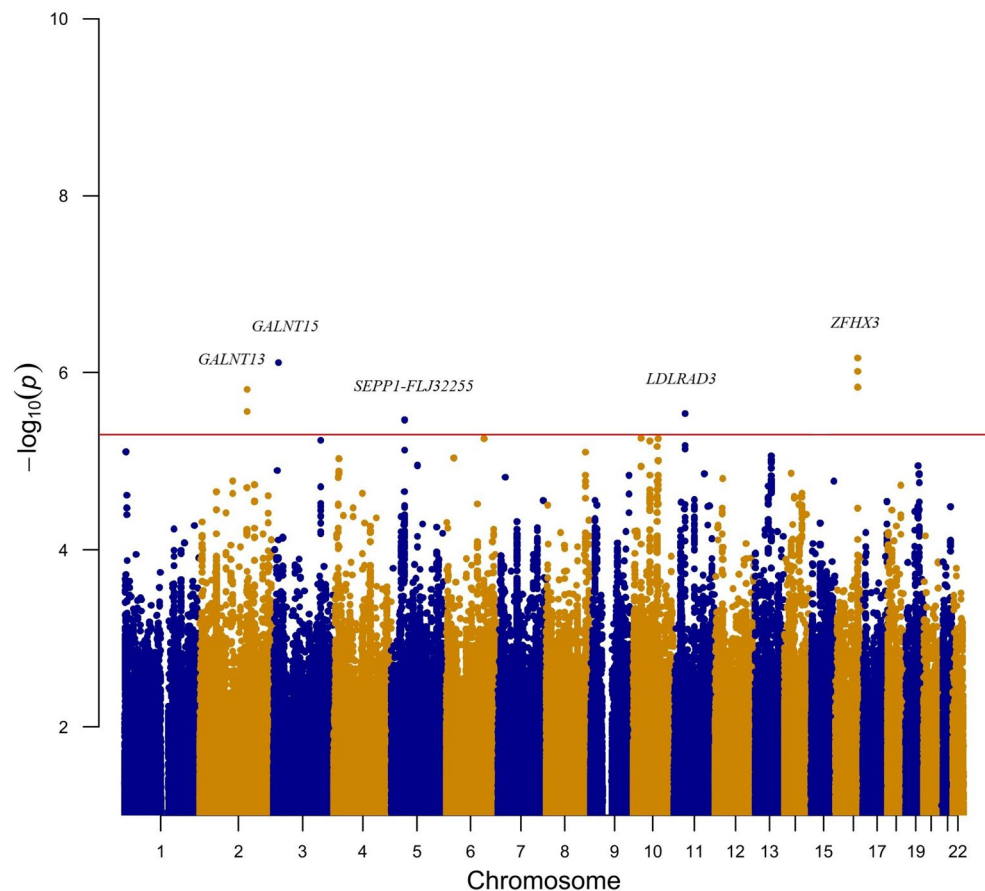


Figure 1. Manhattan plot of the GWAS results for log UMCR. SNPs are plotted on the x axis according to their chromosome position against association with log UMCR on the y axis. The red horizontal line represents the suggestive association threshold of $P = 5.0 \times 10^{-6}$.

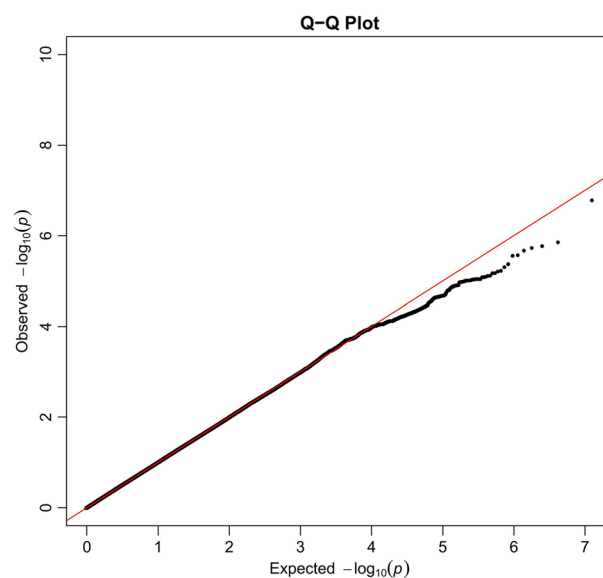


Figure 2. Quantile-quantile plots of log UMCR.

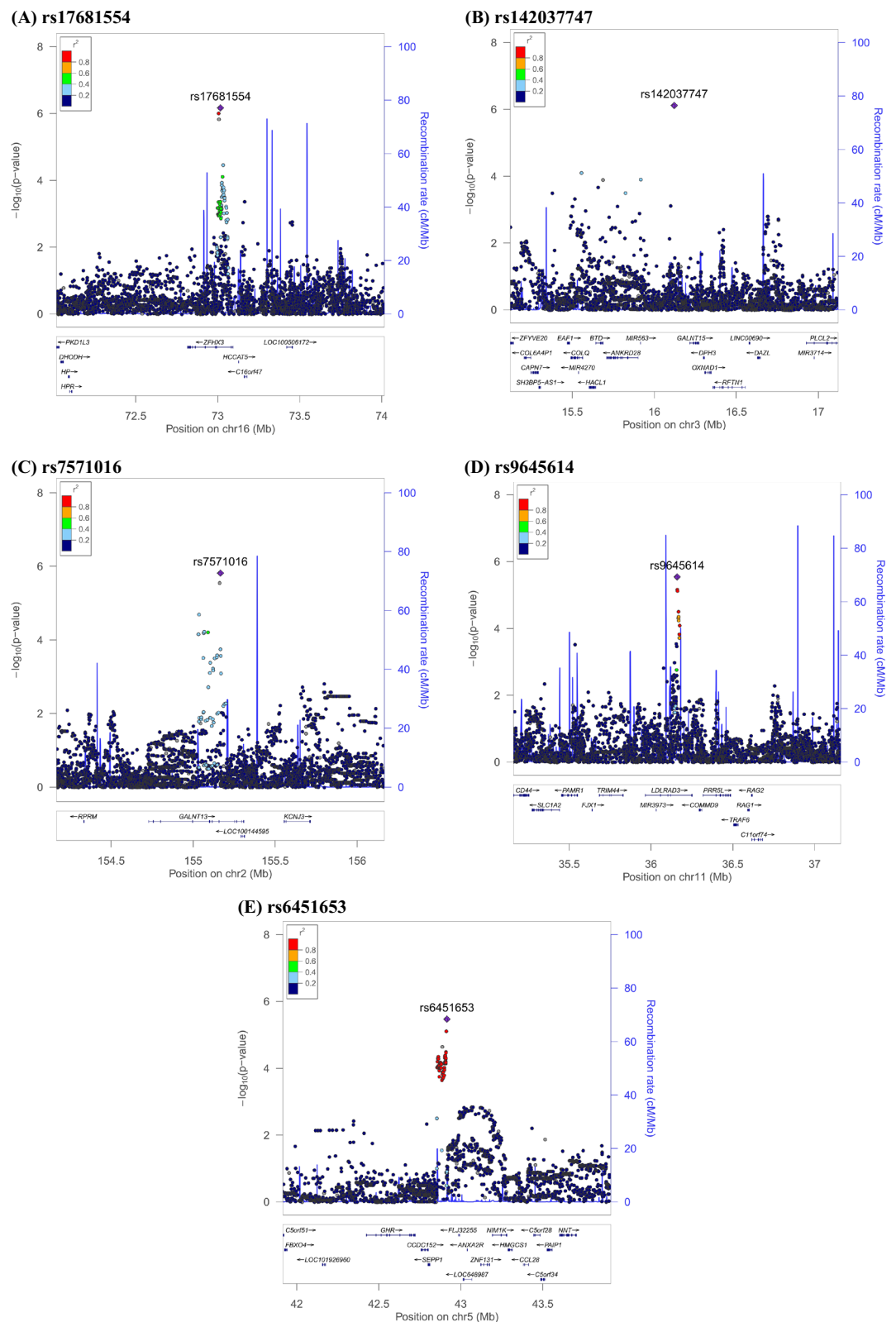


Figure 3. Regional association plots of log UMCR. (A) rs17681554, (B) rs142037747, (C) rs7571016, (D) rs9645614, (E) rs6451653.

Our present unbiased genetic study, revealing the *LDLRAD3* variant associated with melatonin secretion from pineal gland, provides additional evidence for potential mechanistic explanation in AD patients with altered melatonin levels.

rs6451653 is located between *SEPP1* (selenoprotein P, or SELENOP) and pseudogene *FLJ32255*. *SEPP1* serves as a phospholipid hydroperoxide glutathione peroxidase and thus protect the plasma membrane from oxidative damage and is expressed in all brain tissues⁴⁹. *SEPP1* is secreted from astrocytes to neurons for prevention of oxidative damage⁵⁰. Several studies demonstrated that *Sepp1* knockout mice displayed cerebellar ataxia, anxiety, impaired spatial memory, and widespread neurodegeneration in various studies^{51–55}. Also, deletion of *SEPP1* in dogs resulted in central nervous system atrophy and cerebellar ataxia⁵⁶. It is convincing that the *SEPP1* variant is associated with melatonin levels.

This study also showed borderline significance regarding the positive correlation between age and log UMCr. Since aging causes sarcopenia, the subsequently decreased creatinine excretion from urine increases the substance-to-creatinine ratio. Our results support the current concept that aging itself will not cause a decrease in melatonin secretion or excretion.

There was a concern that aMT6s excretion may be altered when renal function declines. A previous study enrolling 20 elderly demonstrated that 24-h urine aMT6s was a reliable surrogate for plasma melatonin level, at least among individuals with GFR 24.6 ml/min or above⁵⁷. Our study confirmed that morning UMCr is not significantly correlated with eGFR, and adjusting eGFR in GWAS analysis essentially did not influence the results.

There are limitations to our study. First, it lacks replication of the result in another cohort. We searched in the UK Biobank, but melatonin as phenotype does not exist in the database. Moreover, the sample size is relatively small; thus, for the time being, these SNPs can only be seen as suggestive signals. The statistical power of this GWAS is only 56%, and therefore there are true loci that remain to be identified and validated. Also, there might be individuals receiving medications that can affect melatonin levels but not documented in the Taiwan Biobank data due to the imprecise nature of the questionnaire survey. However, production or selling of melatonin pills is illegal in Taiwan, and thus this important confounding factor may not be significant in our study.

In summary, we have performed the first GWAS regarding melatonin secretion to date. This GWAS identified five highly suggestive genetic loci encompassing genes that demonstrated potential functional connectivity between the genes-associated melatonin level and circadian behavior, neuronal differentiation, cerebellar ataxia, neurodegeneration and Alzheimer's disease. Replication and functional studies of these genetic variations are warranted to understand better the regulation of melatonin secretion and related clinical disorders.

Data availability

Individual researchers may request to use the data for specific projects on a collaborative basis. Our data has been submitted to the NHGRI-EBI GWAS Catalog (accession ID: GCST90101875).

Received: 5 July 2021; Accepted: 4 May 2022

Published online: 16 May 2022

References

- Galano, A., Tan, D. X. & Reiter, R. J. Melatonin as a natural ally against oxidative stress: A physicochemical examination. *J. Pineal Res.* **51**, 1–16 (2011).
- Hardeland, R. *et al.* Melatonin—A pleiotropic, orchestrating regulator molecule. *Prog. Neurobiol.* **93**, 350–384 (2011).
- Favero, G., Rodella, L. F., Reiter, R. J. & Rezzani, R. Melatonin and its atheroprotective effects: A review. *Mol. Cell. Endocrinol.* **382**, 926–937 (2014).
- Reiter, R. J. Melatonin: Clinical relevance. *Best Pract. Res. Clin. Endocrinol. Metab.* **17**, 273–285 (2003).
- Claustrat, B., Brun, J. & Chazot, G. The basic physiology and pathophysiology of melatonin. *Sleep Med. Rev.* **9**, 11–24 (2005).
- Nduhirabandi, F. & Lochner, A. Melatonin and the metabolic syndrome. In *Melatonin and Melatonergic Drugs in Clinical Practice*, Chapter 6 1st edn (eds Srinivasan, V. *et al.*) 71–95 (Springer, 2014).
- Peschke, E. Melatonin, endocrine pancreas and diabetes. *J. Pineal Res.* **44**, 26–40 (2008).
- Tengattini, S. *et al.* Cardiovascular diseases: Protective effects of melatonin. *J. Pineal Res.* **44**, 16–25 (2008).
- Espino, J., Pariente, J. A. & Rodríguez, A. B. Role of melatonin on diabetes-related metabolic disorders. *World J. Diabetes* **2**, 82–91 (2011).
- Hardeland, R., Madrid, J. A., Tan, D. X. & Reiter, R. J. Melatonin, the circadian multioscillator system and health: The need for detailed analyses of peripheral melatonin signaling. *J. Pineal Res.* **52**, 139–166 (2012).
- Sigurdardottir, L. G. *et al.* Urinary melatonin levels, sleep disruption, and risk of prostate cancer in elderly men. *Eur. Urol.* **67**, 191–194 (2015).
- Arendt, J. Melatonin: Characteristics, concerns, and prospects. *J. Biol. Rhythms* **20**, 291–303 (2005).
- Zeitler, J. M. *et al.* Do plasma melatonin concentrations decline with age?. *Am. J. Med.* **107**, 432–436 (1999).
- Kunz, D. *et al.* A new concept for melatonin deficit: On pineal calcification and melatonin excretion. *Neuropsychopharmacol. Off. Publ. Am. Coll. Neuropsychopharmacol.* **21**, 765–772 (1999).
- Wetterberg, L., Iselius, L. & Lindsten, J. Genetic regulation of melatonin excretion in urine. A preliminary report. *Clin. Genet.* **24**, 399–402 (1983).
- Hallam, K. T. *et al.* The heritability of melatonin secretion and sensitivity to bright nocturnal light in twins. *Psychoneuroendocrinology* **31**, 867–875 (2006).
- McCarthy, M. I. *et al.* Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. *Nat. Rev. Genet.* **9**, 356–369 (2008).
- Graham, C., Cook, M. R., Kavet, R., Sastre, A. & Smith, D. K. Prediction of nocturnal plasma melatonin from morning urinary measures [published correction appears in *Journal of Pineal Research* 1999 **26** 128]. *J. Pineal Res.* **24**, 230–238 (1998).
- Meng, H., Liu, T., Borjigin, J. & Wang, M. M. Ischemic stroke destabilizes circadian rhythms. *J. Circadian Rhythms* **6**, 9 (2008).
- Pacchierotti, C., Iapichino, S., Bossini, L., Pieraccini, F. & Castrogiovanni, P. Melatonin in psychiatric disorders: A review on the melatonin involvement in psychiatry. *Front. Neuroendocrinol.* **22**, 18–32 (2001).
- Kopin, I. J., Pare, C. M., Axelrod, J. & Weissbach, H. The fate of melatonin in animals. *J. Biol. Chem.* **236**, 3072–3075 (1961).

22. Klante, G., Brinschwitz, T., Secci, K., Wollnik, F. & Steinlechner, S. Creatinine is an appropriate reference for urinary sulphatoxy-melatonin of laboratory animals and humans. *J. Pineal Res.* **23**, 191–197 (1997).
23. Chen, C. H. *et al.* Population structure of Han Chinese in the modern Taiwanese population based on 10,000 participants in the Taiwan Biobank project. *Hum. Mol. Genet.* **25**, 5321–5331 (2016).
24. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
25. Delaneau, O., Zagury, J. F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* **10**, 5–6 (2013).
26. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* **44**, 955–959 (2012).
27. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.* **11**, 499–511 (2010).
28. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics (Oxford, England)* **27**, 2156–2158 (2011).
29. Peér, I., Yelensky, R., Altshuler, D. & Daly, M. J. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet. Epidemiol.* **32**, 381–385 (2008).
30. Fallin, M. D. & Kao, W. H. Is “X”-WAS the future for all of epidemiology? [Published correction appears in *Epidemiology* 2011 **22** 881]. *Epidemiology* **22**, 457–468 (2011).
31. Duggal, P., Gillanders, E. M., Holmes, T. N. & Bailey-Wilson, J. E. Establishing an adjusted p-value threshold to control the family-wide type 1 error in genome wide association studies. *BMC Genom.* **9**, 516 (2008).
32. Turner, S. qqman: An R package for visualizing GWAS results using Q-Q and Manhattan plots. *J. Open Source Softw.* **3**, 731 (2018).
33. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2018).
34. Pruim, R. J. *et al.* LocusZoom: Regional visualization of genome-wide association scan results. *Bioinformatics (Oxford, England)* **26**, 2336–2337 (2010).
35. Shim, H. *et al.* A multivariate genome-wide association analysis of 10 LDL subfractions, and their response to statin treatment, in 1868 Caucasians. *PLoS ONE* **10**, e0120758 (2015).
36. Delongchamp, R., Faramawi, M. F., Feingold, E., Chung, D. & Abouelenein, S. The association between SNPs and a quantitative trait: Power calculation. *Eur. J. Environ. Public Health* **2**, 10 (2018).
37. Koch, B. C. *et al.* Impairment of endogenous melatonin rhythm is related to the degree of chronic kidney disease (CREAM study). *Nephrol. Dial. Transplant. Off. Publ. Eur. Dial. Transplant Assoc. Eur. Ren. Assoc.* **25**, 513–519 (2010).
38. Morinaga, T., Yasuda, H., Hashimoto, T., Higashio, K. & Tamaoki, T. A human alpha-fetoprotein enhancer-binding protein, ATBF1, contains four homeodomains and seventeen zinc fingers. *Mol. Cell. Biol.* **11**, 6041–6049 (1991).
39. Miura, Y. *et al.* Cloning and characterization of an ATBF1 isoform that expresses in a neuronal differentiation-dependent manner. *J. Biol. Chem.* **270**, 26840–26848 (1995).
40. Wilcox, A. G., Vizor, L., Parsons, M. J., Banks, G. & Nolan, P. M. Inducible knockout of mouse *Zfh3* emphasizes its key role in setting the pace and amplitude of the adult circadian clock. *J. Biol. Rhythms* **32**, 433–443 (2017).
41. Cheng, L. *et al.* Characterization of a novel human UDP-GalNAc transferase, pp-GalNAc-T15. *FEBS Lett.* **566**, 17–24 (2004).
42. Zhang, Y. *et al.* Cloning and characterization of a new human UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase, designated pp-GalNAc-T13, that is specifically expressed in neurons and synthesizes GalNAc alpha-serine/threonine antigen. *J. Biol. Chem.* **278**, 573–584 (2003).
43. Iqbal, S., Ghanimi Fard, M., Everest-Dass, A., Packer, N. H. & Parker, L. M. Understanding cellular glycan surfaces in the central nervous system. *Biochem. Soc. Trans.* **47**, 89–100 (2019).
44. Ranganathan, S. *et al.* LRAD3, a novel low-density lipoprotein receptor family member that modulates amyloid precursor protein trafficking. *J. Neurosci.* **31**, 10836–10846 (2011).
45. Hardy, J. & Selkoe, D. J. The amyloid hypothesis of Alzheimer’s disease: Progress and problems on the road to therapeutics [published correction appears in *Science* 2002 **297** 2209]. *Science* **297**, 353–356 (2002).
46. Nous, A., Engelborghs, S. & Smolders, I. Melatonin levels in the Alzheimer’s disease continuum: A systematic review. *Alzheimer’s Res. Ther.* **13**, 52 (2021).
47. Wu, Y. H. & Swaab, D. F. The human pineal gland and melatonin in aging and Alzheimer’s disease. *J. Pineal Res.* **38**, 145–152 (2005).
48. Baskett, J. J., Cockrem, J. F. & Antunovich, T. A. Sulphatoxy-melatonin excretion in older people: Relationship to plasma melatonin and renal function. *J. Pineal Res.* **24**, 58–61 (1998).
49. Saito, Y. *et al.* Selenoprotein P in human plasma as an extracellular phospholipid hydroperoxide glutathione peroxidase. Isolation and enzymatic characterization of human selenoprotein p. *J. Biol. Chem.* **274**, 2866–2871 (1999).
50. Steinbrenner, H. & Sies, H. Selenium homeostasis and antioxidant selenoproteins in brain: Implications for disorders in the central nervous system. *Arch. Biochem. Biophys.* **536**, 152–157 (2013).
51. Peters, M. M., Hill, K. E., Burk, R. F. & Weeber, E. J. Altered hippocampus synaptic function in selenoprotein P deficient mice. *Mol. Neurodegener.* **1**, 12 (2006).
52. Caito, S. W. *et al.* Progression of neurodegeneration and morphologic changes in the brains of juvenile mice with selenoprotein P deleted. *Brain Res.* **1398**, 1–12 (2011).
53. Raman, A. V. *et al.* Absence of selenoprotein P but not selenocysteine lyase results in severe neurological dysfunction. *Genes Brain Behav.* **11**, 601–613 (2012).
54. Byrns, C. N., Pitts, M. W., Gilman, C. A., Hashimoto, A. C. & Berry, M. J. Mice lacking selenoprotein P and selenocysteine lyase exhibit severe neurological dysfunction, neurodegeneration, and audiogenic seizures. *J. Biol. Chem.* **289**, 9662–9674 (2014).
55. Burk, R. F. *et al.* Selenoprotein P and apolipoprotein E receptor-2 interact at the blood–brain barrier and also within the brain to maintain an essential selenium pool that protects against neurodegeneration. *FASEB J.* **28**, 3579–3588 (2014).
56. Christen, M. *et al.* Deletion of the SELENOP gene leads to CNS atrophy with cerebellar ataxia in dogs. *PLoS Genet.* **17**, e1009716 (2021).
57. Zhou, J. N., Liu, R. Y., Kamphorst, W., Hofman, M. A. & Swaab, D. F. Early neuropathological Alzheimer’s changes in aged individuals are accompanied by decreased cerebrospinal fluid melatonin levels. *J. Pineal Res.* **35**, 125–130 (2003).

Acknowledgements

This work is supported by Grants from the Ministry of Science and Technology in Taiwan (MOST 105-2314-B-182-062, MOST 106-2314-B-182-043), the Translational Medical Research Program of Academia Sinica (ASTM-108-01-04), National Taiwan University Hospital, Yunlin Branch Intramural Grant (NTUHYL106.X003, NTUHYL107S004, NTUHYL 111.X019, NTUHYL110.X017) and Chang Gung University, Taoyuan, Taiwan (NMRPD1F154, NMRPD1G0711, and BMRPD08). We would like to thank Taiwan Biobank for providing the biological specimens and information for our research, and the staff of the Core Lab, Department of Medical Research, National Taiwan University Hospital for technical support during the study. We would also like to thank Professor Lee-Ming Chuang at the Department of Internal Medicine, National Taiwan University Medical College for the critical review and valuable discussion about this manuscript. Lastly, we thank the Data Science Statistical Cooperation Center of Academia Sinica (AS-CFII-108-117) for statistical support.

doi:10.6342/NTU202401345

Author contributions

P.-H.L., G.-T.C. and Y.-C.C. contributed to the experimental design. Y.-S.W., H.-C.K., Y.-C.L., Y.-S.C., Y.-Y.H., C.-H.L., W.-Y.L., J.-W.L., Chih-Neng H., J.-J.H., M.-L.H., H.-L.L. and Y.-C.C. contributed to the funding, sample procurement and data generation. G.-T.C. was responsible for data analysis; P.-H.L., Chia-Ni H., and C.-Y.S. provided statistical consultation. K.C.-W.L. wrote the revised discussion. Manuscript writing was done by G.-T.C., P.-H.L. and Y.-C.C.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Y.-C.C.

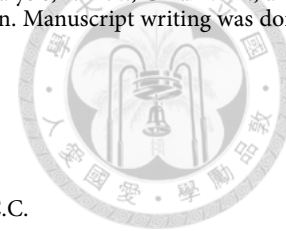
Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022





Genome-Wide Association Studies for Albuminuria of Nondiabetic Taiwanese Population

Wei-Shun Yang^{a,b} Gwo-Tsann Chuang^{b,c} Tony Pan-Hou Che^b
Li-Yun Chueh^b Wen-Yi Li^d Chih-Neng Hsu^e Chia-Ni Hsiung^f
Hsiao-Chia Ku^g Yi-Ching Lin^g Yi-Shun Chen^g Siow-Wey Hee^h
Tien-Jyun Chang^{h,i} Shiau-Mei Chen^h Meng-Lun Hsieh^j Hsiao-Lin Lee^h
Karen Chia-Wen Liao^k Chen-Yang Shen^l Yi-Cheng Chang^{b,h,i,l}

^aDepartment of Internal Medicine, Division of Nephrology, National Taiwan University Hospital Hsin-Chu Branch, Hsinchu, Taiwan; ^bGraduate Institute of Medical Genomics and Proteomics, National Taiwan University, Taipei, Taiwan; ^cDivision of Nephrology, Department of Pediatrics, National Taiwan University Children's Hospital, Taipei, Taiwan; ^dDepartment of Internal Medicine, Division of Nephrology, National Taiwan University Hospital Yunlin Branch, Yunlin, Taiwan; ^eCardiovascular Center, National Taiwan University Hospital Yun-Lin Branch, Yunlin, Taiwan; ^fData Science Statistical Cooperation Center, Institute of Statistical Science, Academia Sinica, Taipei, Taiwan; ^gDepartment of Laboratory Medicine, National Taiwan University Hospital Hsin-Chu Branch, Hsinchu, Taiwan; ^hDepartment of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan; ⁱDepartment of Medicine, College of Medicine, National Taiwan University Hospital, Taipei, Taiwan; ^jDepartment of Medicinal Chemistry, College of Pharmacy, University of Florida, Gainesville, FL, USA; ^kBiological Sciences Division, University of Chicago, Chicago, IL, USA; ^lInstitute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan

Keywords

Genome-wide association studies · Albuminuria · Nondiabetics

Abstract

Introduction: Chronic kidney disease, which is defined by a reduced estimated glomerular filtration rate and albuminuria, imposes a large health burden worldwide. Ethnicity-specific associations are frequently observed in genome-wide association studies (GWAS). This study conducts a GWAS of albuminuria in the nondiabetic population of Taiwan. **Methods:** Nondiabetic individuals aged 30–70 years without a history of cancer were enrolled from the Taiwan

Biobank. A total of 6,768 subjects were subjected to a spot urine examination. After quality control using PLINK and imputation using SHAPEIT and IMPUTE2, a total of 3,638,350 single-nucleotide polymorphisms (SNPs) remained for testing. SNPs with a minor allele frequency of less than 0.1% were excluded. Linear regression was used to determine the relationship between SNPs and log urine albumin-to-creatinine ratio. **Results:** Six suggestive loci are identified in or near the *FCRL3* ($p = 2.56 \times 10^{-6}$), *TMEM161* ($p = 4.43 \times 10^{-6}$), *EFCAB1* ($p = 2.03 \times 10^{-6}$), *ELMOD1*

Wei-Shun Yang and Gwo-Tsann Chuang contributed equally to this work

($p = 2.97 \times 10^{-6}$), *RYR3* ($p = 1.34 \times 10^{-6}$), and *PIEZO2* ($p = 2.19 \times 10^{-7}$). Genetic variants in the *FCRL3* gene that encode a secretory IgA receptor are found to be associated with IgA nephropathy, which can manifest as proteinuria. The *PIEZO2* gene encodes a sensor for mechanical forces in mesangial cells and renin-producing cells. Five SNPs with a p -value between 5×10^{-6} and 5×10^{-5} are also identified in five genes that may have a biological role in the development of albuminuria. **Conclusion:** Five new loci and one known suggestive locus for albuminuria are identified in the non-diabetic Taiwanese population.

© 2023 S. Karger AG, Basel

Introduction

Chronic kidney disease (CKD) is a significant health burden worldwide and increases the risk of other comorbidities including cardiovascular disease (CVD), infection, and malignancy. CKD, which is defined by a reduced estimated glomerular filtration rate (eGFR) and/or albuminuria, affects 8–16% of the global population [1]. Patients with the lowest eGFR and highest albuminuria are at greatest risk of developing CVD or mortality. Albuminuria is a significant predictor of a decline in renal function, end-stage renal disease, CVD, or all-cause mortality [2–5], and the risk is proportionate to the amount of total urinary protein that is excreted [6, 7].

Genome-wide association studies (GWAS) are used to detect disease variants and can account for different ethnic backgrounds. Several ethnicity-specific variants have been identified. Genetic variants in the *PAX4* gene (R121W) are associated with early-onset diabetes, but only for East Asians [8]. The East Asian-specific genetic variant rs671 in the *ALDH2* gene is associated with coronary artery disease [9] and elevated blood pressure [10]. Genetic variants rs9565164 in the *TBC1D4* gene are associated with muscle insulin resistance and postprandial glucose in Greenlandic Inuits [11, 12]. Genetic variant rs7903146 in the *TCF7L2* gene is strongly associated with type 2 diabetes in populations with European ancestry [13] but not in East Asian populations [14, 15]. Genetic variant rs2237892 in the *KCNQ1* gene is strongly associated with type 2 diabetes in North Asia but not in Europe [16]. These examples demonstrate the necessity to identify ethnicity-specific variants, especially in the East Asian population, which has a relatively isolated genetic variant [17, 18].

Previous large-scale GWAS have identified hundreds of loci that are significantly associated with kidney disease traits in databases that include *UMOD* and serum creatinine, and eGFR [19–21] and *CUBN* are associated with

urine albumin-to-creatinine ratio (UACR) [22]. Some of these findings were validated using in vitro or in vivo experiments. A transgenic mouse model with noncoding *UMOD* variants develops salt-sensitive hypertension and kidney damage [23]. However, these results are not replicated in other ethnic populations, such as North American Indians [24] and East Asians [25–27]. Genetic background, lifestyle, diet, and environmental factors are possible causes of these discrepancies.

This study conducted a GWAS of albuminuria in 6,768 participants and of eGFR in 101,073 nondiabetic Taiwanese participants that were enrolled from the Taiwan Biobank. Five new genetic loci for *TMEM161B*, *EFCAB1*, *ELMOD1*, *RYR3*, and *PIEZO2* genes and one locus near gene *FCRL3* are identified.

Methods and Materials

Study Population

The Taiwan Biobank enrolls volunteers aged 30–70 years who have no history of cancer [28]. Urine was collected in the morning from a total of 6,768 subjects. Biological specimens and questionnaires about personal and clinical information were analyzed as de-linked data.

Serum, Urine Creatinine, and Urine Albumin Measurements

Serum and urine creatinine levels were measured using the compensated Jaffe method and a chemistry analyzer (Hitachi LST008). Urine albumin was measured using an automatic quantitative immunoturbidimetric assay. Albuminuria is presented as the UACR in spot urine and is log-transformed because it is not normally distributed.

Genotyping, Quality Control, and Imputation

Genotyping using the Axiom-Taiwan Biobank Array Plate (TWB chip; Affymetrix Inc., Santa Clara, CA, USA) was performed at the National Center for Genome Medicine of Academia Sinica [28]. PLINK (version 1.9), which is an open-source whole-genome data analysis toolset, was used for quality control procedures [29]. Genotypes of SNPs with batch effect were set as missing. SNPs were excluded if there were a large number of missing genotypes (>5%), minor allele frequency was low (<0.1%), or if the values deviated from the Hardy-Weinberg equilibrium (p value < 10^{-5}).

Individuals with discordant sex (self-reported sex incongruent with genetic sex, where genetic male or female is defined as an X chromosome homozygosity estimate of more than 0.8 or less than 0.2), a large number of missing genotypes (>5%), an extreme incidence of heterozygosity (more than 5 standard deviations away from the mean), or a high identity-by-descent score (>0.1875) so close relatedness is excluded from subsequent analyses. The principal components on a linkage disequilibrium-pruned ($r^2 < 0.2$) set of autosomal variants were identified by removing high-linkage disequilibrium regions using PLINK.

Genotype imputation was performed with SHAPEIT [30] and IMPUTE2 [31] with a reference panel, using combined whole-genome sequencing data from the 1,000 Genome Project East

Table 1. Significant loci ($p < 5 \times 10^{-6}$) associated with albuminuria in 6,768 Taiwanese

Chr	BP	SNP	Albuminuria-increasing allele	Other allele	Albuminuria-increasing allele frequency	Beta	SE	p value	Nearest gene
Known									
1	157,675,915	rs141851930	A	T	0.0275	0.0733	0.0156	2.56×10^{-6}	FCRL3
Unknown									
5	88,363,221	rs140005329	G	C	0.0150	0.0957	0.0208	4.43×10^{-6}	TMEM161B (TMEM161B-AS1) ₂
8	48,593,061	rs72639070	A	G	0.0780	0.0344	0.0072	2.03×10^{-6}	EFCAB1
11	107,623,478	rs1939894	C	T	0.1715	0.0245	0.0052	2.97×10^{-6}	ELMOD1
15	33,323,086	rs12101967	G	A	0.3250	0.0204	0.0042	1.34×10^{-6}	RYR3
18	10,721,279	rs4796904	T	A	0.1045	0.0344	0.0066	2.19×10^{-7}	PIEZO2

Chr, chromosome; BP, base pair position in build GRCh38, albuminuria-increasing allele frequency in Taiwan Biobank; beta, regression coefficient; SE, standard error of regression coefficient.

Asian population and the Taiwan Biobank. Imputed SNPs with low-quality scores (info score [32] less than 0.3), a missing genotype rate >5%, or a low frequency (minor allele frequency <1%) were excluded. Indels were removed using VCF tools [33].

Validation

A GWAS using data from CKDgen and the UK Biobank was used to validate the results of this study. Subjects with or without DM and with different ancestry backgrounds were considered separately, if possible.

MR of UACR Using the eGFR

To determine the relationship between the six suggestive loci of UACR and the eGFR, we performed a Mendelian randomization analysis using R package Mendelian Randomization. The relationship between all six SNPs and eGFR was determined. A GRS was generated using these SNPs, which considered the effect size and the estimated variation. GRS is defined as $GRS = \sum_{i=0}^n \beta_i S_i$. The β_i value is the individual UACR-level regression coefficient for the i th SNP and S_i indicates the allele dosage for the i th SNP. The correlation between GRS and UACR level is determined, and then MR analysis, including the proportion of UACR changes that is explained, uses the coefficient that is calculated using the GRS regression. A MR-Egger regression is used to determine possible pleiotropic effects. MR also assumes that the SNPs/GRS are not associated with confounding variables, all of which are defined with respect to the statistical significance of the association between GRS and these confounding factors. To allow the assumption that SNPs/GRS are not associated with confounding variables, the relationship between GRS and confounding factors, including age, sex, body mass index, systolic blood pressure (SBP), and diastolic blood pressure, is determined. A logistic regression is used for sex and linear regression for the remainder of the confounding factors.

Statistical Analyses

Age and eGFR are expressed as mean and standard deviation. The UACR is expressed as a median and an interquartile range. The results of the GWAS are analyzed using PLINK v1.9 with an additive genetic model. Linear regression is used to determine the relationship between SNPs and logUACR. Covariate adjustments include age, sex, and the first ten principal components of ancestry.

The conventional genome-wide significance threshold of $p < 5.0 \times 10^{-8}$ [34] is very conservative for a small sample size, so the level for suggestive significance for this study is $p < 5.0 \times 10^{-6}$ [35, 36]. The threshold for the validation study is 0.05 divided by 6, according to the Bonferroni correction. A Manhattan plot and quantile-quantile plot are generated using the qqman R package [37]. Regional association plots are generated using LocusZoom [38, 39]. The Genome Reference Consortium Human Build 38 is used for gene annotation.

Results

The results of the GWAS were analyzed for albuminuria for 6,768 subjects from the Taiwan Biobank. After imputation and quality control, 3,638,350 autosomal SNPs remain. The baseline characteristics for the study population are listed in online supplementary Table 1 (for all online suppl. material, see <https://doi.org/10.1159/000531783>). Suggestive variants for albuminuria are listed in Table 1. Figure 1 and online supplementary Figure 1, respectively, show the Manhattan plot and the quantile-quantile plot. The genomic inflation factor, which is defined as the median of the observed χ^2 test statistics divided by the expected median of the corresponding χ^2 distribution, is 0.985.

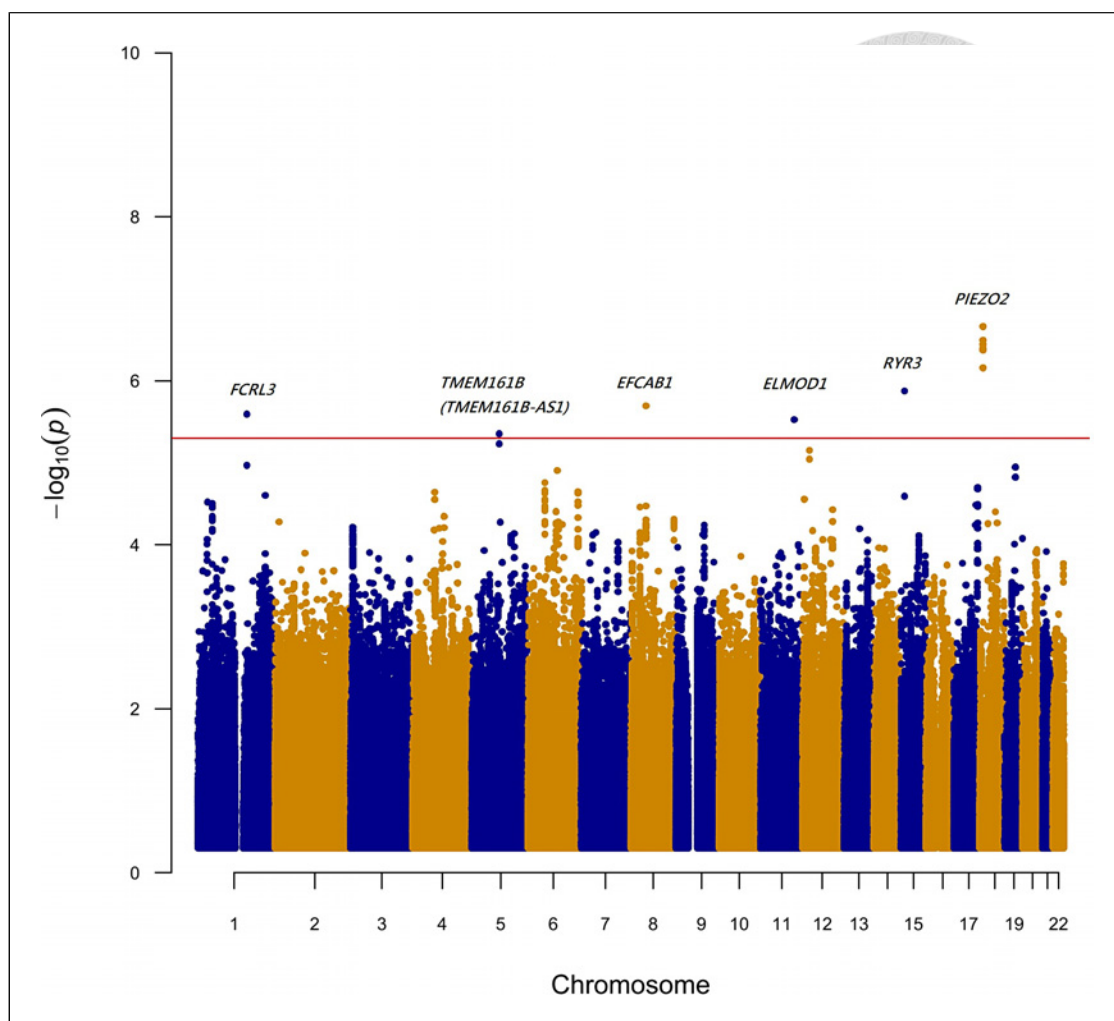


Fig. 1. Manhattan plot for SNPs that are associated with albuminuria.

Regional association plots for the top SNPs are shown in Figure 2. Six candidate loci are identified: rs141851930 near the *FCRL3* gene on chromosome 1 (beta 0.0733, $p = 2.56 \times 10^{-6}$; Fig. 2a), rs140005329 in or near the *TMEM161*(*TMEM161B-AS1*) gene on chromosome 5 (beta 0.0957, $p = 4.43 \times 10^{-6}$; Fig. 2b), rs72639070 near the *EFCAB1* gene on chromosome 8 (beta 0.0344, $p = 2.03 \times 10^{-6}$; Fig. 2c), rs1939894 in the *ELMOD1* gene on chromosome 11 (beta 0.0245, $p = 2.97 \times 10^{-6}$; Fig. 2d), rs12101967 near the *RYR3* gene on chromosome 15 (beta 0.0204, $p = 1.34 \times 10^{-6}$; Fig. 2e), and rs4796904 in the *PIEZO2* gene on chromosome 18 (beta 0.0344, $p = 2.19 \times 10^{-7}$; Fig. 2f).

The respective proportion of variance that is explained by the individual SNPs is 0.33, 0.31, 0.34, 0.33, 0.35, and 0.4 for rs141851930, rs140005329, rs72639070, rs1939894, rs12101967, and rs4796904. These six loci

explain 2.06% of the logUACR level. A list of studies to determine the role of suggestive loci with a p value between 5×10^{-5} and 5×10^{-6} with a common renal trait is shown in online supplementary Table 2.

The sample size is small, but the results for SNPs with association signals of p values between 5×10^{-6} and 5×10^{-5} show 5 SNPs in 5 genes that are plausibly biologically implicated in the development of albuminuria. These include *TAL1*, *GRM7*, *IGF2R*, *SLCO1A2*, and *LTA4H* (online suppl. Table 2). As well as the 6 suggestive genes in Table 1, the *RYR3* and *GRAM7* genes overlap with those that were identified by a previous GWAS for albuminuria in a Japanese population [40].

The MR-Egger slope shows that there is a pleiotropic effect between logUACR and eGFR ($p = 0.035$ for MR-Egger intercept) (Table 2). If SNP rs12101967 (*RYR3*),

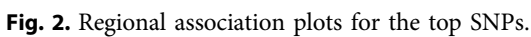


Table 2. Mendelian randomization results and sensitivity analyses for logUACR on eGFR

Number of SNPs	Beta	95% CI	MR-Egger regression intercept (95% CI)	Weight median beta (95% CI)
6 SNPs ($<1 \times 10^{-6}$)	-1.961	-6.013–2.091	0.254 (-18.521 to -0.296)	-2.83 (-7.910–2.249)
5 SNPs*	-3.094	-7.210–1.022	0.269 (-17.658 to -1.740)	-2.385 (-6.840–0.070)

*SNP (rs12101967) removed due to pleiotropic effect.

Table 3. Characteristics of SNPs used in Mendelian randomization sensitivity analysis for logUACR on eGFR

SNP	location	logUACR			eGFR		
		beta	SE	p value	beta	SE	p value
rs141851930	Chr1:157675915	0.0733	0.0156	2.56×10^{-6}	-0.7718	0.3097	0.01272
rs140005329	Chr5:88363221	0.0957	0.0208	4.43×10^{-6}	-0.1569	0.4164	0.7063
rs72639070	Chr8:48593061	0.0344	0.0072	2.03×10^{-6}	-0.09699	0.1503	0.5188
rs1939894	Chr11:107623478	0.0245	0.0052	2.97×10^{-6}	0.06436	0.1049	0.5395
rs12101967*	Chr15:33323086	0.0204	0.0042	1.34×10^{-6}	0.07889	0.0869	0.3639
rs4796904	Chr18:10721279	0.0344	0.0066	2.19×10^{-7}	-0.09976	0.1322	0.4504

UACR, urine albumin creatinine ratio; eGFR, estimated glomerular filtration rate; Chr, chromosome; BP, base pair position in build GRCh38; beta, regression coefficient; SE, standard error of regression coefficient.

which has a dominant association with eGFR and CKD, is excluded, then of the remaining SNPs, only rs1939894 (*ELMOD1*) is positively associated with eGFR (Table 3).

Using an estimated MR, the MR-Egger method shows that logUACR is associated with eGFR (95% CI = -17.658 to -1.740, $p = 0.043$) without a pleiotropic effect ($p = 0.138$ for MR-Egger intercept) (Table 2; Fig. 3). The GRS is not associated with confounding factors, including age (beta = 3.60, 95% CI = -3.862–11.067, $p = 0.344$), sex (odds ratio = -0.67, 95% CI = -2.168–0.847, $p = 0.386$), body mass index (beta = 1.36, 95% CI = -1.296–4.0142, $p = 0.316$), SBP (beta = 8.32, 95% CI = -4.433–21.066, $p = 0.201$) or diastolic blood pressure (beta = 6.10, 95% CI = -1.488–13.680, $p = 0.115$) (Fig. 4).

Discussion

This study identifies five suggestive new variants of albuminuria and replicates one known locus for nondiabetic Taiwanese. Many of the identified genes and alleles are significantly implicated in renal function.

This study shows that rs141851930 in the *FCRL3* gene may be related to logUACR. The *FCRL3* gene encodes a receptor that is expressed in lymphocytes, which binds

secretory IgA [41]. SNPs in the *FCRL3* gene are reported to be associated with IgA nephropathy in the Chinese population [42–44]. A large-scale GWAS showed ethnic heterogeneity of genetic susceptibility to IgA nephropathy. This meta-analysis of 9,507 patients and 20,582 controls with Chinese ancestry and 1,039 patients and 1,289 controls with European ancestry shows that only 17 of the 24 confirmed risk SNPs for IgAN have a consistent association between the two ancestries. The remaining six SNPs are heterogeneous between the ancestries and one (*DEFA* locus) is Chinese-specific [44]. Previous GWAS studies for albuminuria in the European population did not identify loci in the *FCRL3* gene as risk alleles, possibly because of the lower susceptibility to IgAN in this population.

The SNP rs140005329 is an intron variant in the *TMEM161B-AS1* gene that encodes a long noncoding RNA and near a protein-coding gene *TMEM161B*. The SNP rs710375 between the *CCNH-TMEM161B* genes is associated with diabetic nephropathy [45], but the function of *TMEM161B* remains unclear.

The SNP rs72639070 is located near the *EFCAB1* gene. The results of a large GWAS involving 1.5 million individuals [46] show that genetic variants near the *EFCAB1* gene are strongly associated with eGFR. The

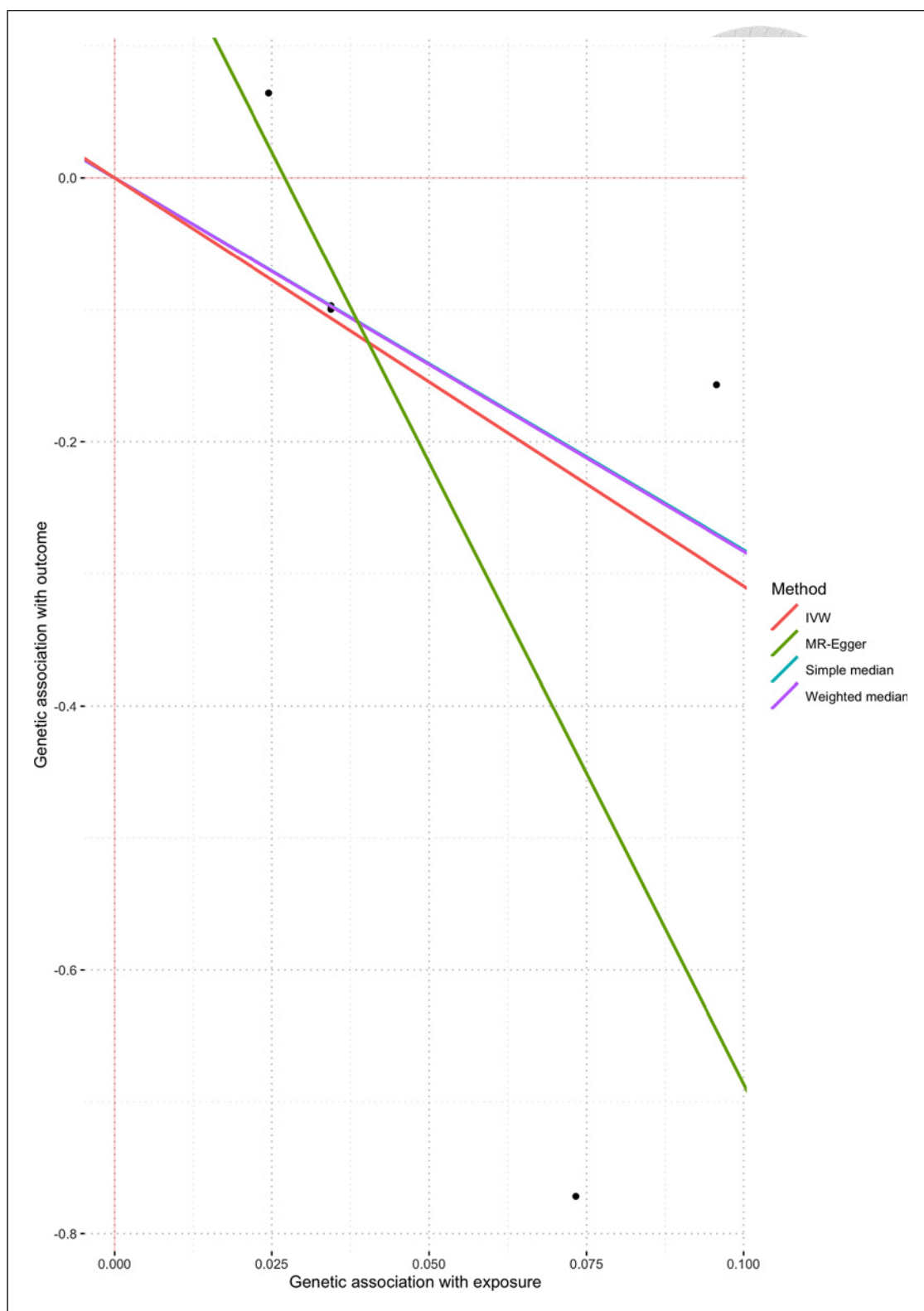


Fig. 3. Scatter plots for Mendelian randomization of 5 SNPs, excluding rs12101967.

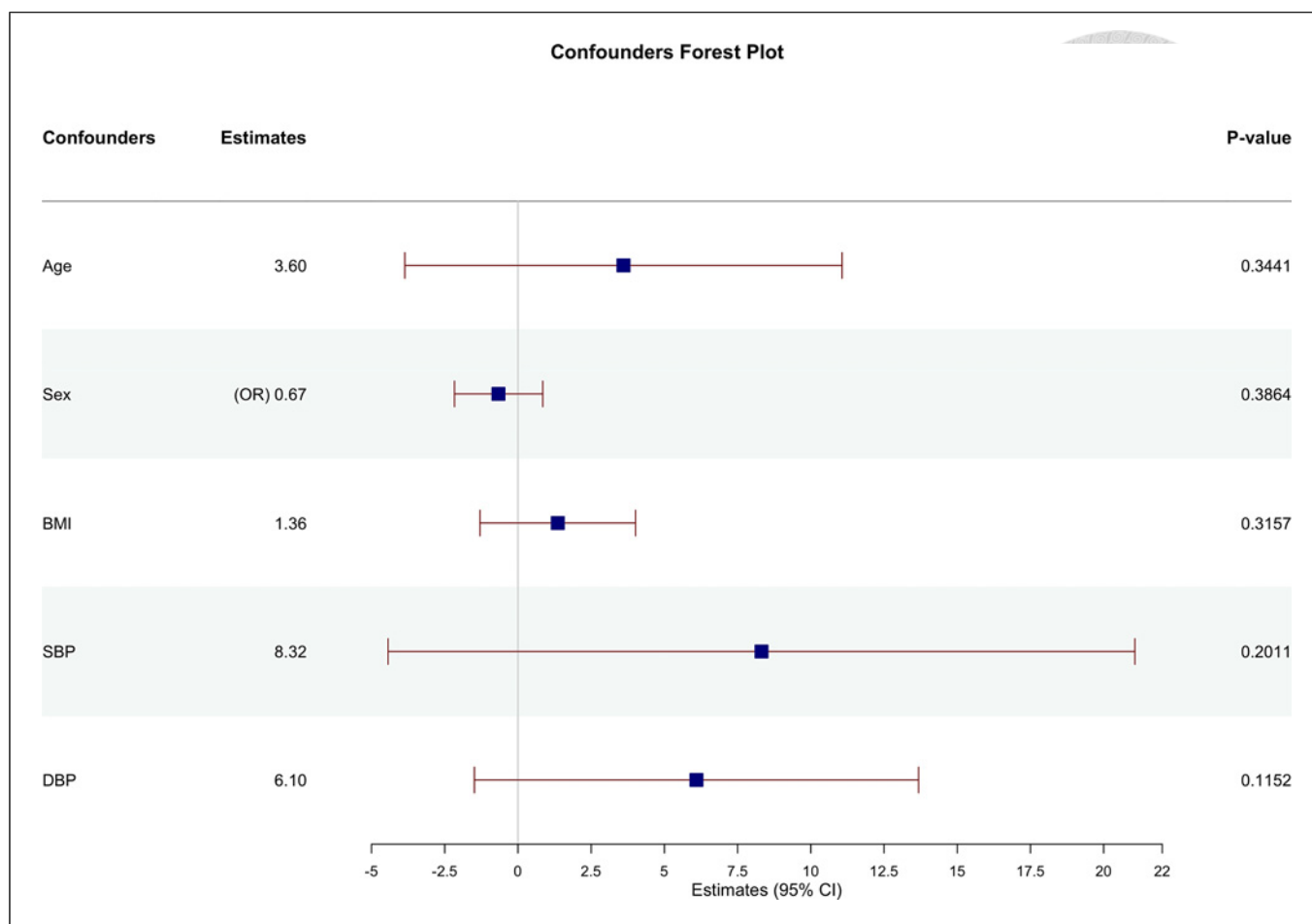


Fig. 4. Grs confounders forest plot.

EFCAB1 gene encodes a Ca^{2+} -binding dynein-associated protein that regulates flagellar and ciliary movement [47]. Variants in the *EFCAB1* gene were also associated with elevated blood pressure in the Japanese population [39]. The *EFCAB1* gene may have a pleiotropic effect on arterial blood pressure and renal function.

The *ELMOD1* gene encodes a protein that belongs to the Arf family GTPase, which affects renal ciliogenesis and cystogenesis by regulating exocytosis [48]. The role of *ELMOD1* in albuminuria is currently unclear.

Ryanodine receptor 3, which is encoded by the *RYR3* gene, releases intracellular calcium and is involved in various cellular processes and human disorders, including heart failure and diabetes [49]. The Family Investigation of Nephropathy and Diabetes (FIND) study reported that the SNP rs2596230 in the *RYR3* gene is associated with diabetic kidney disease in the African-American population [50]. A closely related gene, *RYR2*, is associated

with various nephropathies due to intracellular calcium imbalance [51]. A multiethnic study showed that variants in the *RYR3* gene are associated with a higher eGFR in a population with African ancestry. However, it is not known whether the higher eGFR value indicates glomerular hyperfiltration, which often proceeds proteinuria [52]. These results show that the RYR protein family may have a role in proteinuria and renal disease.

PIEZO2 encodes a sensor for mechanical forces in mesangial cells and renin-producing cells [53]. During dehydration, *PIEZO2* expression is down-regulated in mesangial cells and up-regulated in juxtaglomerular renin-producing cells, and there is overproduction of renin and renin-producing cell proliferation. The expression of renin is also reduced by *PIEZO2* knockdown in juxtaglomerular cells, which demonstrates that *PIEZO2* is an important regulator of glomerular filtration and body fluid status [53]. *PIEZO2* is also a mechanical sensor

for bladder tension [54, 55]. Humans and mice lacking functional *PIEZO2* experience impaired bladder control, and patients who lack *PIEZO2* report an impaired sensation of a full bladder [54]. *PIEZO2* is also involved in enterochromaffin cell response to mechanical forces for regulating the release of serotonin [56].

This study is limited by sample size, so we looked up possible biological role in the development of albuminuria of 5 more SNPs that have association signals with *p* values between 5×10^{-6} and 5×10^{-5} . These include *TAL1*, *GRM7*, *IGF2R*, *SLCO1A2*, and *LTA4H* (online suppl. Table 2). Inserting retrotransposon in the *Tal1* gene causes severe renal disease, which is characterized by progressive glomerulosclerosis, an increase in the glomerular mesangial matrix, deposition of immune complexes, and glomerular enlargement in mice [57, 58]. Genetic variants in the *GRM7* gene are associated with albuminuria in the Japanese population, and a meta-analysis of a GWAS shows that they are also implicated in recovery from acute kidney injury [40, 59]. Genetic variants in the *IGF2R* gene are associated with concentrations of 139 serum metabolites and 41 urine metabolites in patients with CKD, and genetic variants in the *SLCO1A2* gene are associated with urinary concentrations of 1,172 metabolites in another GWAS [60]. *Lta4h* knockout in mice decreases renal fibrosis and the progression of CKD [61].

This study has limitations. Early enrollees in the Taiwan biobank did not have spot urine collected, so the total number of cases for albuminuria analysis is relatively small. Testing numerous SNPs in a small population increases the possibility of overfitting, but candidate loci for albuminuria are identified, and an MR study confirms that logUACR is associated with eGFR. As the number of samples in the Taiwan biobank increases, future replication and validation will be possible. The results of this study are not replicated in databases including CKDgen [62, 63] and the UK Biobank. The former contains data for DM patients that are admixed with non-DM subjects, and study subjects are mostly of European ancestry. The UK Biobank has 3,841 cases with Asian ancestry, 2,429 of whom are nondiabetic, so the results of this study are not replicated (data not shown) for this sample or for trans-ethnic non-DM subjects in the UK Biobank (online suppl. Table 3).

The results of this study are not replicated by either dataset, possibly because there are discrepancies in the minor allele frequency for different populations or the limited number of Asian subjects that these datasets contain. Therefore, the results of this study must be interpreted with caution. However, this study shows that the *RYR3* and *GRM7* genes overlap with the 14 loci identified by a GWAS for albuminuria in a Japanese population of

7,805. The *FCRL3* gene is associated with IgA nephropathy in the Chinese population [42–44]. IgA nephropathy is a frequent cause of albuminuria in nondiabetic populations, which supports the results of this study. A large-scale GWAS or a meta-analysis of the East Asian population is required to determine any association.

Conclusion

This study is the first GWAS on albuminuria for a Taiwanese population with Han ancestry, providing new understanding of possible common pathogeneses for albuminuria and possibly CKD progression also.

Acknowledgments

The authors thank all the participants in the Taiwan Biobank and are grateful for the support of the Taiwan Biobank. The authors also thank the Data Science Statistical Cooperation Center of Academia Sinica (AS-CFII-111-215) for statistical support.

Statement of Ethics

This study is approved by the Institutional Review Board of the National Taiwan University Hospital Research Ethics Committee, approval number 202005001RINC. All subjects provided written informed consent, and all procedures were performed in accordance with the Declaration of Helsinki and relevant regulations.

Conflict of Interest Statement

The authors state that there is no conflict of interest.

Funding Sources

This study is funded by intramural grants from the Yunlin Branch of the National Taiwan University Hospital (NTU-HYL106.X003, NTUHYL107.S004, NTUHYL109.S103, NTU-HYL110.X017, NTUHYL110.X019), the Joint Research Project with Far Eastern Hospital (109 FTN06), an intramural grant from the Institute of Biomedical Sciences of Academia Sinica, Taiwan (AS-TM-111-01-03), intramural grant from the National Taiwan University Hospital (NTUH110-M4813) and a grant from the Ministry of Science and Technology (MOST 106-2314-B-182-043).

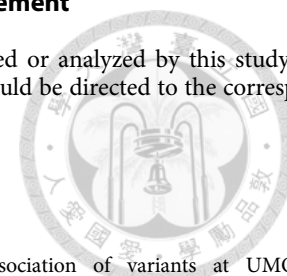
Author Contributions

Wei-Shun Yang and Gwo-Tsann Chuang drafted the article. Gwo-Tsann Chuang, Tony Pan-Hou Che, Li-Yun Chueh, Karen Chia-Wen Liao, and Chen-Yang Shen analyzed the dataset and ran statistics. Wen-Yi Li and Chih-Neng Hsu collected the data and

organized funding. Chia-Ni Hsiung, Hsiao-Chia Ku, Yi-Ching Lin, Yi-Shun Chen, Siow-Wey Hee, Tien-Jyun Chang, Shiau-Mei Chen, Meng-Lun Hsieh, and Hsiao-Lin Lee curated the dataset and revised the draft and graph. Yi-Cheng Chang created the concept for the study and supervised the entire project.

Data Availability Statement

All data that are generated or analyzed by this study are included. Further inquiries should be directed to the corresponding author.



References

- Chen TK, Knicely DH, Grams ME. Chronic kidney disease diagnosis and management: a review. *JAMA*. 2019 Oct 1;322(13):1294–304.
- Berhane AM, Weil EJ, Knowler WC, Nelson RG, Hanson RL. Albuminuria and estimated glomerular filtration rate as predictors of diabetic end-stage renal disease and death. *Clin J Am Soc Nephrol*. 2011 Oct;6(10):2444–51.
- Packham DK, Alves TP, Dwyer JP, Atkins R, de Zeeuw D, Cooper M, et al. Relative incidence of ESRD versus cardiovascular mortality in proteinuric type 2 diabetes and nephropathy: results from the DIAMETRIC (Diabetes Mellitus Treatment for Renal Insufficiency Consortium) database. *Am J Kidney Dis*. 2012 Jan;59(1):75–83.
- Yokoyama H, Araki S, Haneda M, Matsushima M, Kawai K, Hirao K, et al. Chronic kidney disease categories and renal-cardiovascular outcomes in type 2 diabetes without prevalent cardiovascular disease: a prospective cohort study (JDDM25). *Diabetologia*. 2012 Jul;55(7):1911–8.
- Wada T, Haneda M, Furuichi K, Babazono T, Yokoyama H, Iseki K, et al. Clinical impact of albuminuria and glomerular filtration rate on renal and cardiovascular events, and all-cause mortality in Japanese patients with type 2 diabetes. *Clin Exp Nephrol*. 2014 Aug;18(4):613–20.
- Chronic Kidney Disease Prognosis Consortium, Matsushita K, van der Velde M, Astor BC, Woodward M, Levey AS, et al. Association of estimated glomerular filtration rate and albuminuria with all-cause and cardiovascular mortality in general population cohorts: a collaborative meta-analysis. *Lancet*. 2010 Jun 12;375(9731):2073–81.
- Matsushita K, Coresh J, Sang Y, Chalmers J, Fox C, Guallar E, et al. Estimated glomerular filtration rate and albuminuria for prediction of cardiovascular outcomes: a collaborative meta-analysis of individual participant data. *Lancet Diabetes Endocrinol*. 2015 Jul;3(7):514–25.
- Hosoe J, Suzuki K, Miya F, Kato T, Tsunoda T, Okada Y, et al. Structural basis of ethnic-specific variants of PAX4 associated with type 2 diabetes. *Hum Genome*. 2021 Jul 5;8(1):25.
- Takeuchi F, Yokota M, Yamamoto K, Nakashima E, Katsuya T, Asano H, et al. Genome-wide association study of coronary artery disease in the Japanese. *Eur J Hum Genet*. 2012 Mar;20(3):333–40.
- Levy D, Ehret GB, Rice K, Verwoert GC, Launer LJ, Dehghan A, et al. Genome-wide association study of blood pressure and hypertension. *Nat Genet*. 2009 Jun;41(6):677–87.
- Moltke I, Grarup N, Jorgensen ME, Bjerregaard P, Treebak JT, Fumagalli M, et al. A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes. *Nature*. 2014 Aug 14;512(7513):190–3.
- Manousaki D, Kent JW Jr, Haack K, Zhou S, Xie P, Greenwood CM, et al. Toward precision medicine: TBC1D4 disruption is common among the Inuit and leads to underdiagnosis of type 2 diabetes. *Diabetes Care*. 2016 Nov;39(11):1889–95.
- Grant SF, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J, et al. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat Genet*. 2006 Mar;38(3):320–3.
- Chang YC, Chang TJ, Jiang YD, Kuo SS, Lee KC, Chiu KC, et al. Association study of the genetic polymorphisms of the transcription factor 7-like 2 (TCF7L2) gene and type 2 diabetes in the Chinese population. *Diabetes*. 2007 Oct;56(10):2631–7.
- Ng MC, Tam CH, Lam VK, So WY, Ma RC, Chan JC. Replication and identification of novel variants at TCF7L2 associated with type 2 diabetes in Hong Kong Chinese. *J Clin Endocrinol Metab*. 2007 Sep;92(9):3733–7.
- Yasuda K, Miyake K, Horikawa Y, Hara K, Osawa H, Furuta H, et al. Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. *Nat Genet*. 2008 Sep;40(9):1092–7.
- Mahajan A, Go MJ, Zhang W, Below JE, Gaulton KJ, Ferreira T, et al. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat Genet*. 2014 Mar;46(3):234–44.
- Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016 Aug 18;536(7616):285–91.
- Kottgen A, Glazer NL, Dehghan A, Hwang SJ, Katz R, Li M, et al. Multiple loci associated with indices of renal function and chronic kidney disease. *Nat Genet*. 2009 Jun;41(6):712–7.
- Gudbjartsson DF, Holm H, Indridason OS, Thorleifsson G, Edvardsson V, Sulem P, et al. Association of variants at UMOD with chronic kidney disease and kidney stones-role of age and comorbid diseases. *Plos Genet*. 2010 Jul 29;6(7):e1001039.
- Gorski M, Tin A, Garnaas M, McMahon GM, Chu AY, Tayo BO, et al. Genome-wide association study of kidney function decline in individuals of European descent. *Kidney Int*. 2015 May;87(5):1017–29.
- Boger CA, Chen MH, Tin A, Olden M, Kottgen A, de Boer IH, et al. CUBN is a gene locus for albuminuria. *J Am Soc Nephrol*. 2011 Mar;22(3):555–70.
- Trudu M, Janas S, Lanzani C, Debaix H, Schaeffer C, Ikehata M, et al. Common noncoding UMOD gene variants induce salt-sensitive hypertension and kidney damage by increasing uromodulin expression. *Nat Med*. 2013 Dec;19(12):1655–60.
- Franceschini N, Haack K, Almay L, Laston S, Lee ET, Best LG, et al. Generalization of associations of kidney-related genetic loci to American Indians. *Clin J Am Soc Nephrol*. 2014 Jan;9(1):150–8.
- Okada Y, Sim X, Go MJ, Wu JY, Gu D, Takeuchi F, et al. Meta-analysis identifies multiple loci associated with kidney function-related traits in east Asian populations. *Nat Genet*. 2012 Jul 15;44(8):904–9.
- Hishida A, Nakatomi M, Akiyama M, Kamatani Y, Nishiyama T, Ito H, et al. Genome-wide association study of renal function traits: results from the Japan multi-institutional collaborative cohort study. *Am J Nephrol*. 2018;47(5):304–16.
- Lee J, Lee Y, Park B, Won S, Han JS, Heo NJ. Genome-wide association analysis identifies multiple loci associated with kidney disease-related traits in Korean populations. *PLoS One*. 2018;13(3):e0194044.
- Chen CH, Yang JH, Chiang CWK, Hsiung CN, Wu PE, Chang LC, et al. Population structure of Han Chinese in the modern Taiwanese population based on 10,000 participants in the Taiwan biobank project. *Hum Mol Genet*. 2016 Dec 15;25(24):5321–31.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007 Sep;81(3):559–75.
- Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods*. 2013 Jan;10(1):5–6.

- 31 Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet.* 2012 Jul 22;44(8):955–9.
- 32 Marchini J, Howie B. Genotype imputation for genome-wide association studies. *Nat Rev Genet.* 2010 Jul;11(7):499–511.
- 33 Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics.* 2011 Aug 1;27(15):2156–8.
- 34 Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol.* 2008 May;32(4):381–5.
- 35 Duggal P, Gillanders EM, Holmes TN, Bailey-Wilson JE. Establishing an adjusted p-value threshold to control the family-wide type 1 error in genome wide association studies. *BMC Genomics.* 2008 Oct 31;9:516.
- 36 Fallin MD, Kao WH. Is “X”-WAS the future for all of epidemiology? *Epidemiology.* 2011 Jul;22(4):457–9; discussion 67–8.
- 37 Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *bioRxiv.* 2014:005165.
- 38 Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics.* 2010 Sep 15;26(18):2336–7.
- 39 Team Rdc. *R: a language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing; 2010.
- 40 Okuda H, Okamoto K, Abe M, Ishizawa K, Makino S, Tanabe O, et al. Genome-wide association study identifies new loci for albuminuria in the Japanese population. *Clin Exp Nephrol.* 2020 Aug;24(8):1–9.
- 41 Agarwal S, Kraus Z, Dement-Brown J, Alabi O, Starost K, Tolnay M. Human fc receptor-like 3 inhibits regulatory T cell function and binds secretory IgA. *Cell Rep.* 2020 Feb 4;30(5):1292–9.e3.
- 42 Yu XQ, Li M, Zhang H, Low HQ, Wei X, Wang JQ, et al. A genome-wide association study in Han Chinese identifies multiple susceptibility loci for IgA nephropathy. *Nat Genet.* 2011 Dec 25;44(2):178–82.
- 43 Zhong Z, Feng S, Shi D, Xu R, Yin P, Wang M, et al. Association of FCRL3 gene polymorphisms with IgA nephropathy in a Chinese han population. *DNA Cel Biol.* 2019 Oct;38(10):1155–65.
- 44 Li M, Wang L, Shi DC, Foo JN, Zhong Z, Khor CC, et al. Genome-wide meta-analysis identifies three novel susceptibility loci and reveals ethnic heterogeneity of genetic susceptibility for IgA nephropathy. *J Am Soc Nephrol.* 2020 Dec;31(12):2949–63.
- 45 Taira M, Imamura M, Takahashi A, Kamatani Y, Yamauchi T, Araki SI, et al. A variant within the FTO confers susceptibility to diabetic nephropathy in Japanese patients with type 2 diabetes. *PLoS One.* 2018;13(12):e0208654.
- 46 Liu H, Doke T, Guo D, Sheng X, Ma Z, Park J, et al. Epigenomic and transcriptomic analyses define core cell types, genes and targetable mechanisms for kidney disease. *Nat Genet.* 2022 Jul;54(7):950–62.
- 47 Sasaki K, Shiba K, Nakamura A, Kawano N, Satouh Y, Yamaguchi H, et al. Calaxin is required for cilia-driven determination of vertebrate laterality. *Commun Biol.* 2019;2:226.
- 48 Lipschutz JH. The role of the exocyst in renal ciliogenesis, cystogenesis, tubulogenesis, and development. *Kidney Res Clin Pract.* 2019 Sep 30;38(3):260–6.
- 49 Kushnir A, Wajsborg B, Marks AR. Ryanodine receptor dysfunction in human disorders. *Biochim Biophys Acta Mol Cell Res.* 2018 Nov;1865(11 Pt B):1687–97.
- 50 Iyengar SK, Sedor JR, Freedman BI, Kao WH, Kretzler M, Keller BJ, et al. Genome-wide association and trans-ethnic meta-analysis for advanced diabetic kidney disease: family investigation of nephropathy and diabetes (FIND). *Plos Genet.* 2015 Aug;11(8):e1005352.
- 51 Park SJ, Li C, Chen YM. Endoplasmic reticulum calcium homeostasis in kidney disease: pathogenesis and therapeutic targets. *Am J Pathol.* 2021 Feb;191(2):256–65.
- 52 Horimoto A, Xue D, Cai J, Lash JP, Daviglus ML, Franceschini N, et al. Genome-wide admixture mapping of estimated glomerular filtration rate and chronic kidney disease identifies European and african ancestry-of-origin loci in hispanic and latino individuals in the United States. *J Am Soc Nephrol.* 2022 Jan;33(1):77–87.
- 53 Mochida Y, Ochiai K, Nagase T, Nonomura K, Akimoto Y, Fukuhara H, et al. Piezo2 expression and its alteration by mechanical forces in mouse mesangial cells and renin-producing cells. *Sci Rep.* 2022 Mar 10;12(1):4197.
- 54 Marshall KL, Saade D, Ghitani N, Coombs AM, Szczot M, Keller J, et al. PIEZO2 in sensory neurons and urothelial cells coordinates urination. *Nature.* 2020 Dec;588(7837):290–5.
- 55 Dalghi MG, Ruiz WG, Clayton DR, Montalbetti N, Daugherty SL, Beckel JM, et al. Functional roles for PIEZO1 and PIEZO2 in urothelial mechanotransduction and lower urinary tract interoception. *JCI Insight.* 2021 Oct 8;6(19):e152984.
- 56 Wang F, Knutson K, Alcaino C, Linden DR, Gibbons SJ, Kashyap P, et al. Mechanosensitive ion channel Piezo2 is important for enterochromaffin cell response to mechanical forces. *J Physiol.* 2017 Jan 1;595(1):79–91.
- 57 Shultz LD, Lane PW, Coman DR, Taylor S, Hall E, Lyons B, et al. Hairpatches, a single gene mutation characterized by progressive renal disease and alopecia in the mouse. A potential model for a newly described heritable human disorder. *Lab Invest.* 1991 Nov;65(5):588–600.
- 58 Hosur V, Cox ML, Burzenski LM, Riding RL, Alley L, Lyons BL, et al. Retrotransposon insertion in the T-cell acute lymphocytic leukemia 1 (Tal1) gene is associated with severe renal disease and patchy alopecia in Hairpatches (Hpt) mice. *PLoS One.* 2013;8(1):e53426.
- 59 Larach DB, Engoren MC, Schmidt EM, Heung M. Genetic variants and acute kidney injury: a review of the literature. *J Crit Care.* 2018 Apr;44:203–11.
- 60 Schlosser P, Li Y, Sekula P, Raffler J, Grundner-Culemann F, Pietzner M, et al. Genetic studies of urinary metabolites illuminate mechanisms of detoxification and excretion in humans. *Nat Genet.* 2020 Feb;52(2):167–76.
- 61 Montford JR, Bauer C, Dobrinskikh E, Hopp K, Levi M, Weiser-Evans M, et al. Inhibition of 5-lipoxygenase decreases renal fibrosis and progression of chronic kidney disease. *Am J Physiol Ren Physiol.* 2019 Apr 1;316(4):F732–42.
- 62 Teumer A, Tin A, Sorice R, Gorski M, Yeo NC, Chu AY, et al. Genome-wide association studies identify genetic loci associated with albuminuria in diabetes. *Diabetes.* 2016 Mar;65(3):803–17.
- 63 Teumer A, Li Y, Ghasemi S, Prins BP, Wuttke M, Hermle T, et al. Genome-wide association meta-analyses and fine-mapping elucidate pathways influencing albuminuria. *Nat Commun.* 2019 Sep 11;10(1):4130.