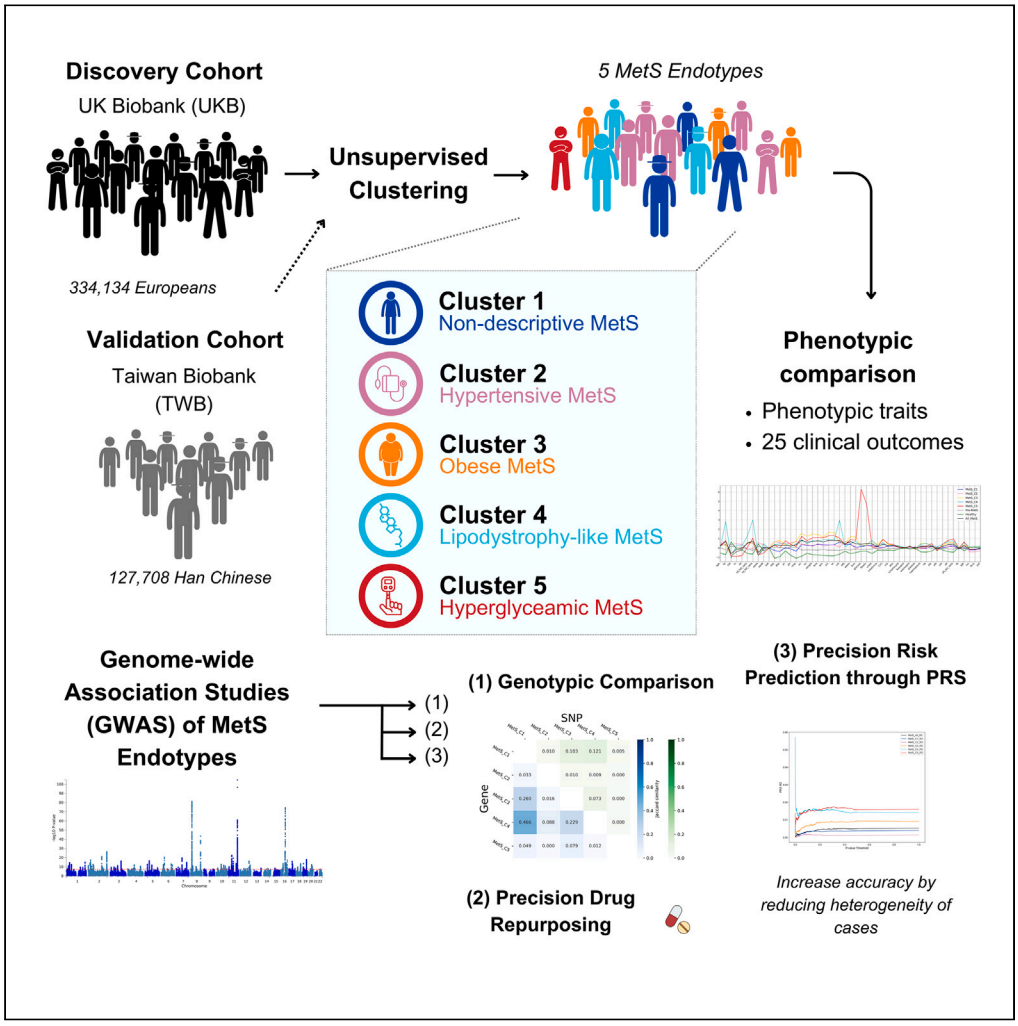
iScience

CellPress
OPEN ACCESS

Article

Unsupervised clustering identified clinically relevant metabolic syndrome endotypes in UK and Taiwan Biobanks



Aylwin Ming Wee Lim, Evan Unit Lim, Pei-Lung Chen, Cathy Shen Jang Fann

csjfann@ibms.sinica.edu.tw

Highlights

Through clustering, we found distinct clinically relevant MetS endotypes

We identified unique genotypic traits for each MetS endotype

We performed first-of-its-kind analysis of polygenic risk scores (PRS) on endotypes

Homogeneity of cases can improve the accuracy of PRS models

Lim et al., iScience 27, 109815
July 19, 2024
https://doi.org/10.1016/j.isci.2024.109815

iScience

CellPress
OPEN ACCESS

Article

Unsupervised clustering identified clinically relevant metabolic syndrome endotypes in UK and Taiwan Biobanks

Aylwin Ming Wee Lim,[1,2,3] Evan Unit Lim,[2,4] Pei-Lung Chen,[5,6] and Cathy Shen Jang Fann[1,2,7,*]

SUMMARY

Metabolic syndrome (MetS) is a collection of cardiovascular risk factors; however, the high prevalence and heterogeneity impede effective clinical management. We conducted unsupervised clustering on individuals from UK Biobank to reveal endotypes. Five MetS subgroups were identified: Cluster 1 (C1): non-descriptive, Cluster 2 (C2): hypertensive, Cluster 3 (C3): obese, Cluster 4 (C4): lipodystrophy-like, and Cluster 5 (C5): hyperglycemic. For all of the endotypes, we identified the corresponding cardiometabolic traits and their associations with clinical outcomes. Genome-wide association studies (GWASs) were conducted to identify associated genotypic traits. We then determined endotype-specific genotypic traits and constructed polygenic risk score (PRS) models specific to each endotype. GWAS of each MetS clusters revealed different genotypic traits. C1 GWAS revealed novel findings of *TRIM63*, *MYBPC3*, *MYLPF*, and *RAPSN*. Intriguingly, C1, C3, and C4 were associated with genes highly expressed in brain tissues. MetS clusters with comparable phenotypic and genotypic traits were identified in Taiwan Biobank.

INTRODUCTION

Non-communicable diseases (metabolic diseases, cardiovascular diseases, neurodegenerative diseases, cancer, and pulmonary diseases) are complex—influenced by both genetics and non-genetics factors.[1,2] To understand the genetic causes and pathophysiology of complex diseases, genome-wide association studies (GWASs) are widely utilized,[3–6] however, the heterogeneous nature of complex disease has impeded effective translation from GWAS to clinical, especially in case-control GWAS of complex diseases.[7–9] Metabolic syndrome (MetS) is the very definition of a highly heterogeneous complex disease because MetS encompasses a spectrum of obvolute conditions such as hypertension, dyslipidemia, type 2 diabetes (T2D), and obesity.[10,11]

MetS represents a collection of known cardiovascular risk factors, which contributes to much morbidity and mortality associated with cardiovascular diseases (CVDs) such as myocardial infarction, stroke, atherosclerosis, and heart failure.[12–14] In addition to that, MetS also associated with a multitude of other non-cardiovascular adverse health outcomes: malignancies, renal diseases, and neurological complications (dementia and Alzheimer disease).[15–17] For these reasons, identifying individuals with MetS with high risk of complications and preventing MetS-associated diseases are crucial in treating MetS. Mulugeta et al. study on the UK Biobank further emphasizes this point, demonstrating the importance of subgrouping in understanding cardiometabolic multimorbidity and its implications for public health and clinical interventions.[18]

The diagnosis and classification of MetS are arbitrary and continually evolving according to experts' consensus and clinical evidence from observational studies.[19] The prevalence of MetS is high and increasing; 34.2% in United States,[20] 22.1% in Australia,[21] 33.9% in Türkiye,[22] 29.3% in South Korea,[23] and 25.5% in Taiwan.[24] With such high MetS prevalence, almost one-third of the general population, managing MetS poses an immense burden on healthcare systems.

However, the binary classification of MetS is insufficient to reflect the heterogeneity, differing risk to disease outcomes and unpredictable pharmacotherapy effectiveness.[25,26] k-means clustering is a well-known unsupervised learning approach that groups objects with similar characteristics into clusters for identifying subtypes of complex diseases.[27,28] Ahlqvist et al. through k-means clustering on T2D revealed clinically relevant subtypes of T2D in Swedish cohorts,[28] which was later replicated in other populations[29–31]; however, there are some variations that indicated population-specific subtypes. Dennis et al. further evaluated the data-driven cluster analysis using clinical trials data but suggested that models based on simple clinical features might be more useful in stratifying patients.[32]

[1]Taiwan International Graduate Program in Molecular Medicine, National Yang Ming Chiao Tung University and Academia Sinica, Taipei 112304, Taiwan
[2]Institute of Biomedical Sciences, Academia Sinica, Taipei 115, Taiwan
[3]ASUS Intelligent Cloud Services (AICS), Taipei 112, Taiwan
[4]College of Computing, Georgia Institute of Technology, Atlanta, GA 30332, USA
[5]Graduate Institute of Medical Genomics and Proteomics, College of Medicine, National Taiwan University, Taipei 10617, Taiwan
[6]Department of Medical Genetics, National Taiwan University Hospital, Taipei 100, Taiwan
[7]Lead contact
*Correspondence: csjfann@ibms.sinica.edu.tw
https://doi.org/10.1016/j.isci.2024.109815

Endotypes refer to disease subtypes with differing pathophysiology; and to be clinically relevant, endotypes should be dependent on biological pathways and pathophysiology on top of the ability to distinguish between the heterogeneity of various clinical outcomes.[33] Furthermore, the endotypes of complex diseases should allow for stratified or precision treatment approach with differing clinical course and treatment responses. Unsupervised clustering can be a valuable technique to reclassify MetS, uncovering distinct endotypes and reducing the overall heterogeneity of MetS. Furthermore, as MetS is more encompassing, which comprises multiple overlapping cardiometabolic conditions, applying unsupervised clustering on MetS, the bigger picture, might reveal intriguing insights.

In our study, we applied unsupervised clustering on MetS in UK Biobank (UKB) cohort to reveal clinically relevant endotypes of MetS. We conducted first-of-its-kind GWAS of endotypes to identify sub-phenotypes-associated genotypic traits and potential drug repurposing targets. MetS endotypes were further predicted and evaluated in Taiwan Biobank (TWB).

RESULTS

Prevalence and characterization of metabolic syndrome in UK Biobank

In our comprehensive analysis of 334,134 white British individuals from the UK Biobank (UKB), 31.1% (n = 103,996) were identified as having metabolic syndrome (MetS), meeting at least three of the five established MetS criteria. Additionally, 55.3% (n = 184,644) were classified as pre-MetS, meeting one or two criteria, underscoring the widespread prevalence of MetS and its precursor states.

Five MetS clusters were identified with distinct phenotypic traits

The optimal number of clusters, k, was determined at five through silhouette coefficient and the elbow method (Figure S2). Individuals with MetS formed five clusters: Cluster 1 (n = 33,707; 32.4%), Cluster 2 (n = 23,215; 22.3%), Cluster 3 (n = 30,089; 28.9%), Cluster 4 (n = 13,116; 12.6%), and Cluster 5 (n = 3,869; 3.7%). From the principal-component analysis (PCA) plot (Figure 1) of the MetS criteria, the five MetS clusters were located away from the healthy individuals with pre-MetS interspersed among healthy and MetS. MetS Clusters 1, 3, and 4 were in close proximity with some overlaps. MetS Cluster 1 appeared to be more consolidated due to smaller variation of principal component 1 and 2, which is reflected in smaller standard deviations of the MetS criteria (Table S3).

Our results revealed marked differences in lipid profiles, blood pressure, anthropometric measurements, glycemic traits, and liver enzymes across the clusters (Figure 2). For instance, Cluster 5, designated as "hyperglycemic," displayed the most severe glycemic abnormalities, including the highest prevalence of diagnosed type 2 diabetes. In contrast, Cluster 4, with "lipodystrophy-like" characteristics, showed an atypical MetS phenotype, mirroring findings from previous studies by Yaghootkar et al.[34] and Udler et al.[35] The "obese" Cluster 3 was characterized by the most pronounced obesity-related traits. Interestingly, despite having the highest blood pressure readings, Cluster 2 exhibited lower lung function, a novel finding warranting further investigation. Cluster 1 was less distinctive but was notable for its lower concentrations of ketone bodies and apolipoproteins (Figure S5). The interrelation of these phenotypic traits (Figure S3) underscores the complex nature of MetS.

Our sex-specific clustering also revealed MetS clusters rather similar to the sex-combined clustering with slight differences shown in Figures S13–S15. As the results from sex-specific clustering appear to be rather similar to that of sex-combined clustering, we did not proceed with further analysis for sex-specific MetS clusters and will explore the sex-specific clusters in future work, which will reveal more sex-specific effects on MetS.

Clinical implications of MetS cluster stratification

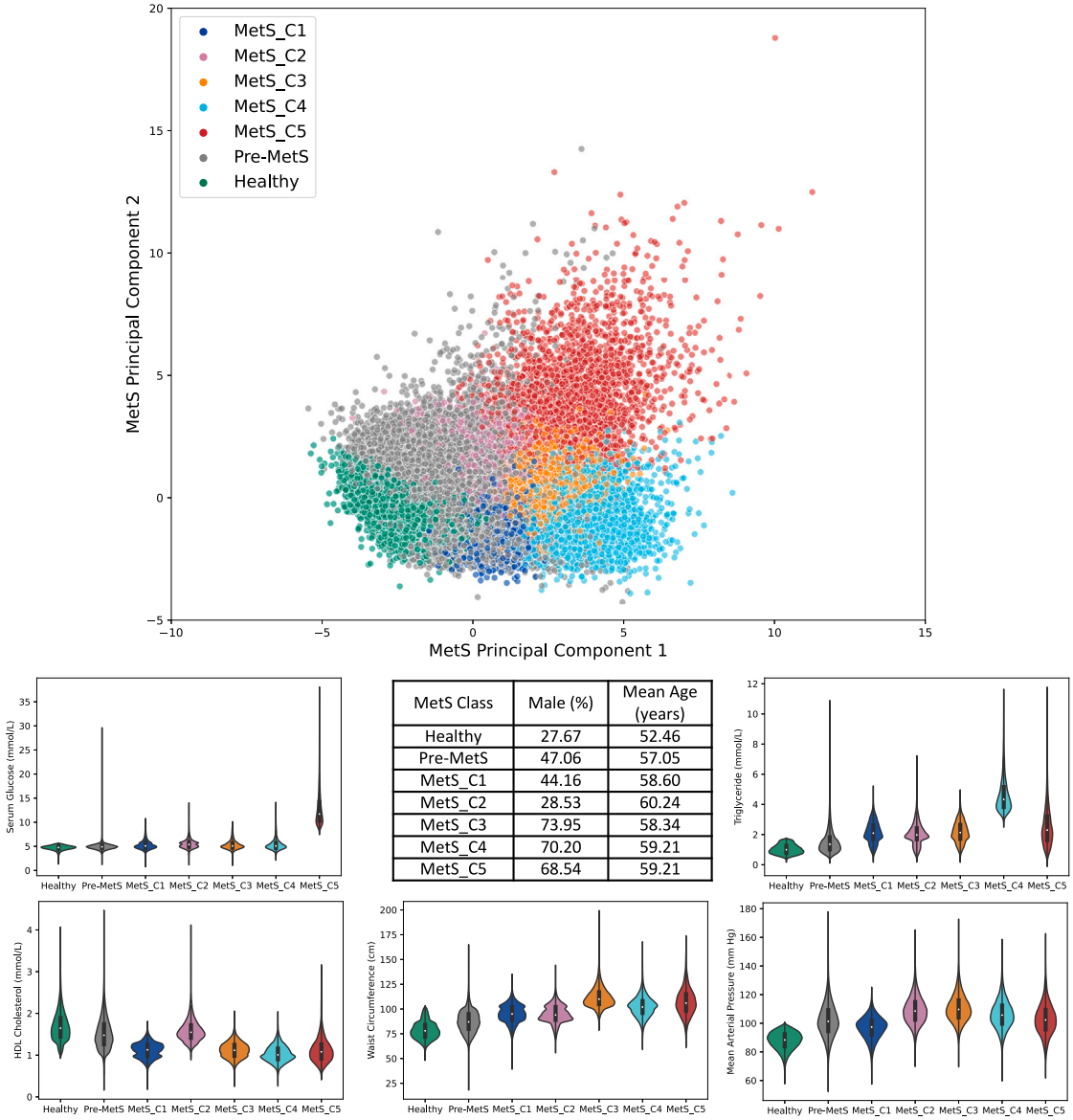
Utilizing multivariable logistic regression models, we assessed the association of each MetS cluster with 25 clinical outcomes, including cardiovascular diseases, chronic kidney disease, various cancer types, dementia, and Alzheimer disease (Figures 3 and S6 and Table S5). Our findings align with the known association of MetS with increased cardiovascular risk, particularly in specific clusters. MetS is a known risk factor for multiple common cancers[17]; however, the mechanism linking MetS and cancer is not well elucidated. Furthermore, the cancer risk varied significantly among the clusters, with Cluster 5 showing the highest odds for liver and pancreas cancer and Cluster 4 being most associated with breast and colorectal cancer. The relative risks of various cancers are similar to that of odds ratio (Table S21). These differential risks underscore the heterogeneity within MetS and the importance of cluster-specific medical management.

Genotypic traits of MetS clusters

Through genome-wide association studies (GWASs) and FUMA SNP2GENE analysis (Figure 4 and Tables S8A–S8F and S9A–S9F), we identified distinct genetic signatures for each MetS cluster. Notably, a comparison of Manhattan plots (Figure 5) revealed both common and unique genomic loci across the clusters. The differential gene expression across various tissues (Figures S7A–S7F) further highlights the complexity of MetS at the genetic level. GWAS of MetS clusters identified rather contrasting mapped genes from that of GWAS of MetS criteria, with only $\leq$10 genes similar across that in all MetS clusters.

Comparison of genotypic traits among clusters

The pairwise genotype comparison among MetS clusters by Jaccard and cosine similarity index on independent significant SNPs and mapped genes are shown in Figure 5 and Table S15. MetS Clusters 1, 3, and 4 were the most similar in term of genotypic traits, whereas MetS Clusters 2 and 5 were the most contrasting from other MetS clusters. Furthermore, the comparison of genotypic traits of MetS clusters



MetS Class	Male (%)	Mean Age (years)
Healthy	27.67	52.46
Pre-MetS	47.06	57.05
MetS_C1	44.16	58.60
MetS_C2	28.53	60.24
MetS_C3	73.95	58.34
MetS_C4	70.20	59.21
MetS_C5	68.54	59.21

Figure 1. Principal-component analysis (PCA) plot and violin plot of five MetS criteria in UK Biobank

Principal-component analysis plot shows principal component 1 versus principal component 2. Principal-component analysis was constructed based on the five MetS criteria namely serum glucose, waist circumference, triglyceride, HDL cholesterol, and mean arterial pressure. Principal component 1 explained 40.2% of the variance, whereas principal component 2 explained 19.6% of the variance. Violin plots show the distribution of the five MetS criteria across healthy, pre-MetS, and MetS clusters.

with all MetS showed that each of the MetS clusters GWAS managed to identify unique SNPs and genes despite them being subgroups of all MetS (Table S16).

Precision drug repurposing and precision disease risk prediction

Leveraging the genetic insights from our GWAS, we performed GREP enrichment analysis to identify potential drug repurposing opportunities. The analysis revealed distinct ATC drug groups associated with each MetS endotype, suggesting tailored therapeutic strategies (Table 1). For instance, genes associated with Cluster 1 were enriched in cardiovascular drugs like anti-obesity and antihypertensive agents. Cluster 4's gene associations were specific to lipid-modifying agents, underscoring the potential for targeted lipid management in this group.

In predictive modeling, the polygenic risk scores (PRS) for MetS clusters demonstrated varying levels of predictive accuracy. Cluster 4's PRS was particularly effective at lower *p*-value thresholds, whereas the PRS for the collective MetS group excelled at higher thresholds (Figure 6A).

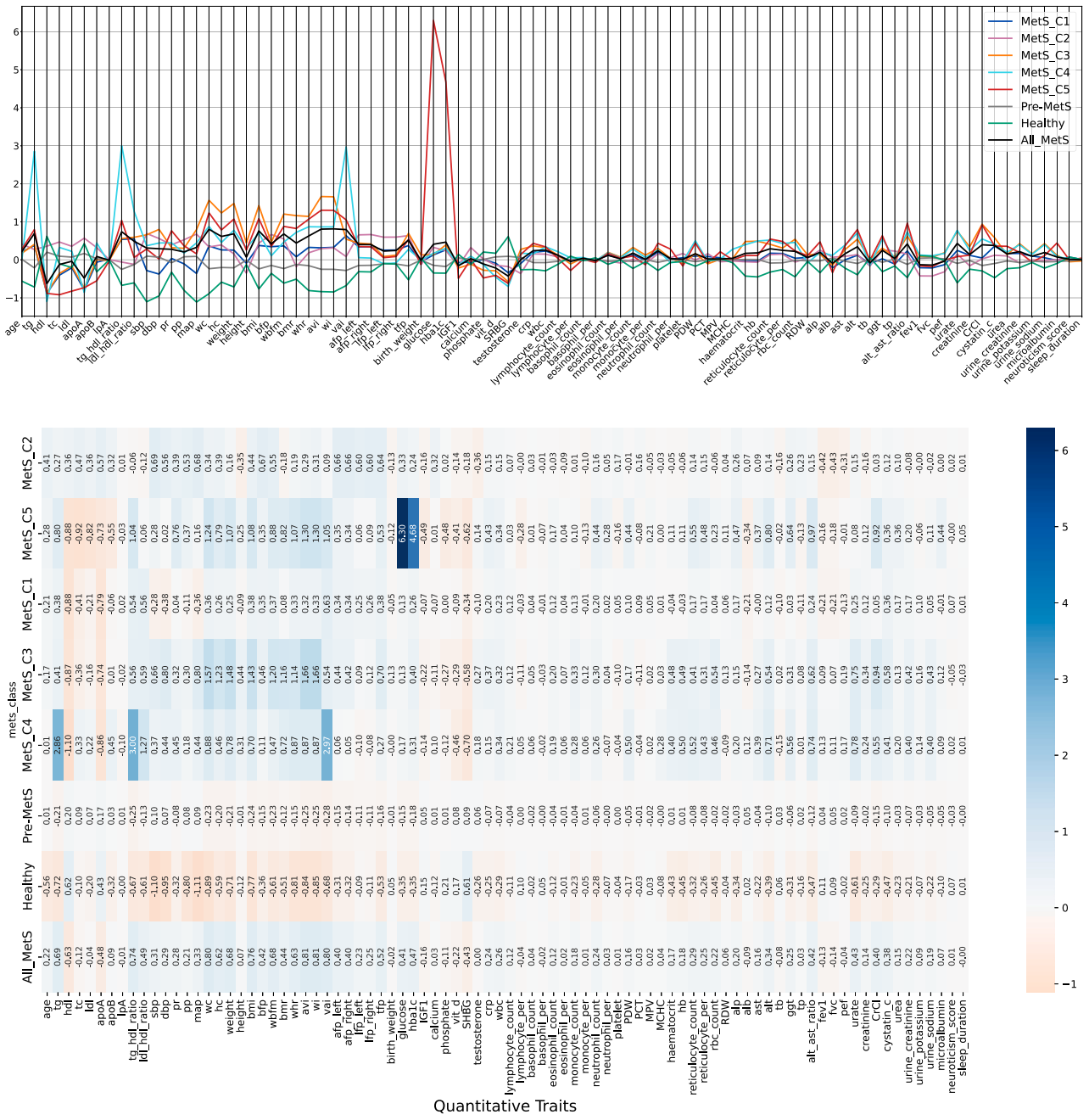


Figure 2. Parallel plot and heatmap displaying standard scores (*Z* scores) for 87 quantitative traits across metabolic syndrome clusters, pre-metabolic syndrome, and healthy individuals in the UK Biobank

Figure 2 provides a detailed parallel plot and heatmap analysis, showcasing the standard scores (*Z* scores) for 87 distinct quantitative traits across metabolic syndrome (MetS) clusters, pre-MetS, and healthy individuals within the UK Biobank dataset. This figure is designed to illuminate the complex and multifaceted nature of MetS, revealing the unique phenotypic patterns that distinguish each MetS cluster from one another and from pre-MetS and healthy baselines. Key observations include the pronounced obesity-related markers in Cluster 3 and the markedly elevated glucose levels and HbA1c in Cluster 5, illustrating the heterogeneity within MetS diagnoses. By incorporating a wide range of cardiometabolic parameters, Figure 2 underscores the variability within MetS clusters and highlights the critical need for a nuanced understanding of MetS subtypes in enhancing precision medicine approaches. tg, triglyceride; hdl, high-density lipoprotein; tc, total cholesterol; ldl, low-density lipoprotein; apoA, apolipoprotein A; apoB, apolipoprotein B; sbp, systolic blood pressure; dbp, diastolic blood pressure; pr, pulse rate; pp, pulse pressure; map, mean arterial pressure; wc, waist circumference; hc, hip circumference; bmi: body mass index; bfp: body fat percentage; wbfm: whole-body fat mass; bmr: basal metabolic rate; whr: waist-to-hip ratio; avi: abdominal visceral index; wi: waist index; vai: visceral adiposity index; afp_left: arm fat percentage (left); afp_right: arm fat percentage (right); lfp_left: leg fat percentage (left); lfp_right: leg fat percentage (right); tfp: trunk fat percentage; IGF1: insulin growth factor 1; vit_d: vitamin d; SHBG: sex hormone binding

Figure 2. *Continued*

globulin; crp: c-reactive protein; wbc: white blood cell; PDW: platelet distribution width; PCT: plateletcrit; MPV: mean platelet volume; MCHC: mean corpuscular hemoglobin concentration; hb: hemoglobin; rbc_count: red blood cell count; RDW: red cell distribution width; alp: alkaline phosphatase; alb: albumin; ast: aspartate aminotransferase; alt: alanine transaminase; tb: total bilirubin; ggt: gamma-glutamyl transferase; tp: total protein; fvc: forced vital capacity; fev1: forced expiratory volume in 1s; pef: peak expiratory flow; CrCl: creatine clearance.

To mitigate sample size disparities, we conducted a random sampling of 3,800 cases per group, revealing that PRS models for three out of the five MetS endotypes outperformed the all-MetS model across various *p*-value thresholds (Figure 6B). This finding highlights the importance of considering sample size and genetic heritability in PRS model performance.

Similar MetS endotypes identified in TWB

Extending our analysis to the Taiwan Biobank (TWB), we identified similar MetS clusters, indicating a cross-population consistency in MetS phenotypes (Figures 7 and S9 and Table S12). A notable variance was observed in Cluster 4 (lipodystrophy-like MetS) of TWB, which exhibited distinct lipid profiles and a lower prevalence of T2D and hypertension compared to its UKB counterpart. This highlights potential population-specific phenotypic expressions of MetS.

Despite the smaller MetS sample size in TWB, we observed significant overlaps in mapped genes between the two biobanks, although independent significant SNPs were largely population-specific. For example, Cluster 1 in TWB shared 43 mapped genes with its UKB counterpart, indicating potential common genetic pathways across populations. However, one GWAS in TWB (Cluster 2) did not identify any significant genetic loci due to the smaller sample size (Figure S11), underscoring the challenges in cross-population genetic studies.

DISCUSSION

Our study utilized unsupervised clustering on MetS criteria in the UK Biobank (UKB) to delineate five clinically relevant MetS endotypes. These endotypes are semi-distinct in both phenotypic and genotypic traits, as evidenced by our analyses across 25 clinical outcomes. This novel approach, combining unsupervised clustering with genome-wide association studies (GWASs), has revealed critical insights into the underlying genotypes that drive the pathophysiology of MetS. Remarkably, we validated these MetS clusters in the Taiwan Biobank (TWB), demonstrating their applicability across populations with diverse ancestries.

Intriguingly, all MetS clusters except Cluster 2 shared only three genotypes of *LPCAT2* (lysophosphatidylcholine acyltransferase 2), *NUDT21* (nudix hydrolase 21), and *OGFOD1* (2-oxoglutarate and iron-dependent oxygenase domain containing 1). All three genes were also identified in Cluster 1 and Cluster 3 of TWB. *NUDT21* and *OGFOD1* have both been reported to be associated with BMI,[36,37] highlighting the common shared obesity trait that might predispose to MetS in individuals within these clusters. *LPCAT2* has never been reported to be associated with any cardiometabolic traits. Single-cell RNA-sequencing data showed high *LPCAT2* expression in immune cells, specifically basophil, eosinophil, neutrophil, and monocytes.[38,39] Furthermore, *LPCAT2* has reported to be responsible for increased expression of inflammatory genes in response to bacterial stimuli.[40] All these previous findings of *LPCAT2* could indicate chronic inflammation being a key player in MetS pathophysiology.

One of our study's most intriguing findings is the unique profile of MetS Cluster 1. Despite lacking specific cardiometabolic traits, this cluster is linked to a broad range of clinical outcomes, including cardiovascular disease (CVD), chronic kidney disease, dementia, and bladder cancer. Notably, Cluster 1 exhibits the highest CVD odds, even after adjusting for T2D status, a trend that is also observed in its TWB counterpart. This poses a significant clinical challenge, as patients in this cluster may not be readily identified as high-risk due to their inconspicuous clinical presentation. Beyond just the aggregation of MetS risk factors, Cluster 1's elevated risk suggests a deeper, potentially unrecognized pathophysiological mechanism. This is further hinted at by its low levels of ketone bodies and medium high-density lipoprotein (HDL) lipid composition, factors inversely associated with insulin resistance and CVD risk.[41,42] However, determining whether these metabolomic traits are consequences or indicators of the disease remains challenging. The GWAS of Cluster 1 has highlighted several genes, including *TRIM63* (tripartite motif containing 63), *MYBPC3* (myosin-binding protein C3), *MYLPF* (myosin light chain, phosphorylatable, fast skeletal muscle), RAPSN (receptor-associated protein of the synapse), and *LPL* (lipoprotein lipase), which may provide insights into underlying mechanisms. Particularly, the expression of *TRIM63* and *MYBPC3* in cardiomyocytes[43,44] and *MYLPF* and *RAPSN* in skeletal muscles[45,46] could be indicative of a link between muscle function, skeletal insulin resistance,[47] and increased CVD risk in this cluster.

The phenotypic traits of MetS Cluster 4 are similar to that of lipodystrophy-like and liver/lipid T2D clusters identified by Udler et al.[35] with lipodystrophy-like features such as low obesity traits and dyslipidemia. Furthermore, Cluster 4 GWAS highlighted some of the genes previously reported in "lipodystrophy-like" insulin resistance cluster by both Yaghootkar et al.[34] and Udler et al.[35] such as *GRB14* (growth factor receptor bound protein 14), *GCKR* (glucokinase regulator), and *IRS1* (insulin receptor substrate 1). We also identified several genes from similar gene families such as *FAM* (family with sequence similarity member): *FAM76A*, *FAM171A2*, *FAM192A*, *FAM89B*, and *FAM180B*, similar to *FAM13A* in Yaghootkar et al.[34] Genetic variants mapped to *GCKR* were discovered in Cluster 4 of both UKB and TWB. Cluster 4 supports the importance of identifying individuals of normal weight obesity due to the high CVD risks.[48]

MetS and individual MetS components such as hypertension, hyperglycemia, and obesity are known risk factors for AF.[49–53] Cluster 3 had the highest odds for AF after adjusting for type 2 diabetes status and highlights the alarming risk of developing AF in this MetS endotype, probably due to the role of obesity in MetS toward AF. Cluster 3 GWAS highlighted genes that are uniquely expressed in adipose tissue, brain

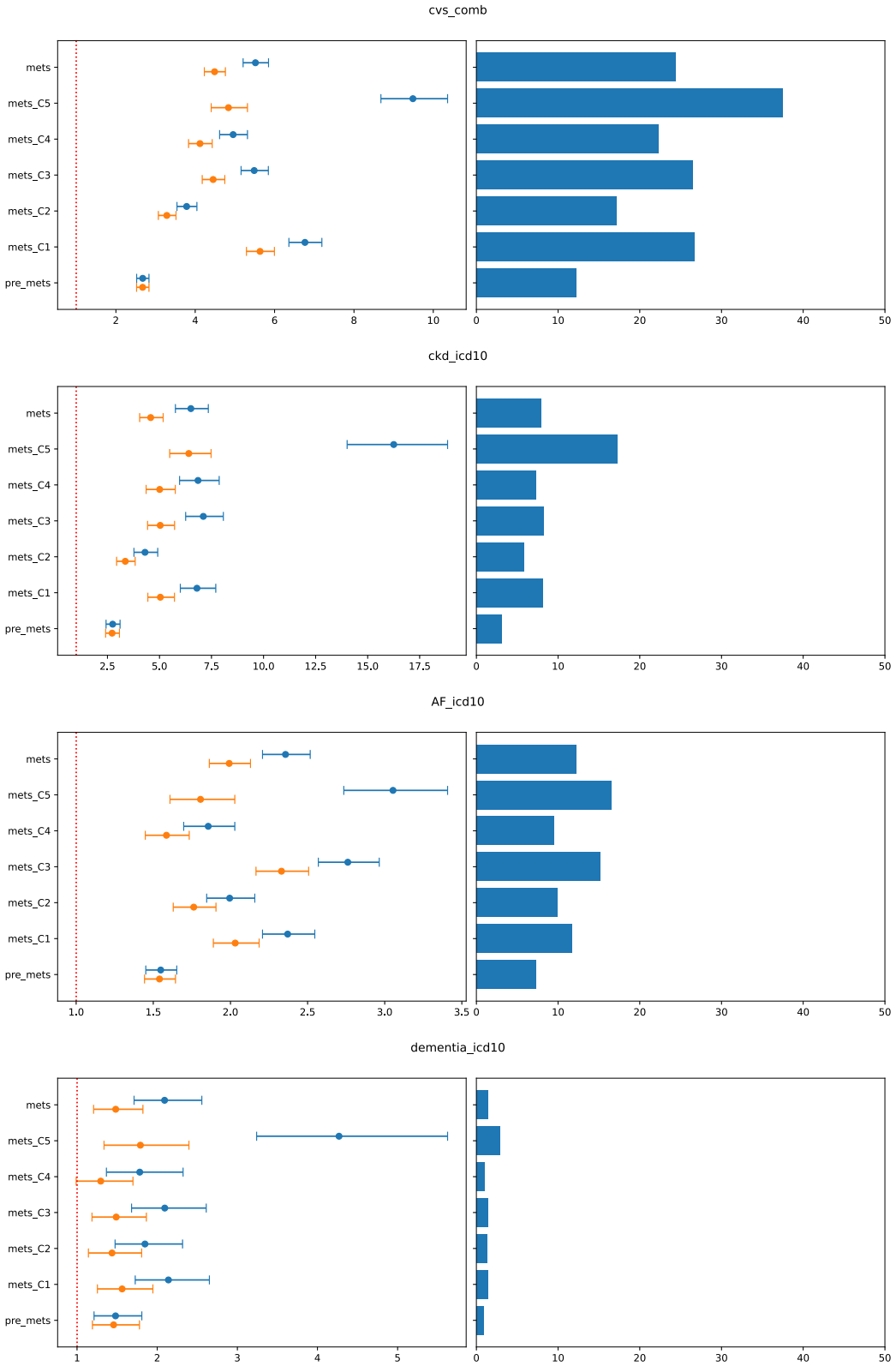

cvs_comb
mets
mets_C5
mets_C4
mets_C3
mets_C2
mets_C1
pre_mets
2
4
6
8
10
0
10
20
30
40
50
ckd_icd10
mets
mets_C5
mets_C4
mets_C3
mets_C2
mets_C1
pre_mets
2.5
5.0
7.5
10.0
12.5
15.0
17.5
0
10
20
30
40
50
AF_icd10
mets
mets_C5
mets_C4
mets_C3
mets_C2
mets_C1
pre_mets
1.0
1.5
2.0
2.5
3.0
3.5
0
10
20
30
40
50
dementia_icd10
mets
mets_C5
mets_C4
mets_C3
mets_C2
mets_C1
pre_mets
1
2
3
4
5
0
10
20
30
40
50

Figure 3. Odds ratios (and 95% confidence intervals) for health outcomes (composite CVD outcomes, atrial fibrillation, depression, and all cancers) in metabolic syndrome, its clusters, and pre-metabolic syndrome compared to healthy controls (left); percentage of cases in each category (right)

Left panel: x axis represents the odds ratios for health outcomes, adjusted for age and sex (blue), and further adjusted for T2D status (orange). A red dotted line indicates an odds ratio of 1, signifying equal event odds in metabolic syndrome groups versus healthy controls. Right panel: x axis shows the percentage of each health outcome; y axis categorizes overall metabolic syndrome, its clusters, and pre-metabolic syndrome. Table S5 provides further details.

amygdala, or associated with obesity traits: for example, *FTO* (fat-mass- and obesity-associated protein), *MC4R* (melanocortin 4 receptor), *CALCRL* (calcitonin-receptor-like receptor), and *IL34* (interleukin-34). *FTO* and *MC4R* are well-known obesity genes, which were associated with obesity traits[36,54] and type 2 diabetes.[55,56] *CALCRL* have been reported to be associated with obesity traits,[57] highly expressed in adipose tissue[45] and negatively associated with leptin, a hormone that helps maintain normal body weight.[58] Genetic variants of *IL34* were associated with Alzheimer disease,[59] BMI,[60] and type 2 diabetes in multi-ancestry cohort.[61] However, it is unsure which specific genetic traits underlie the pathophysiology of obesity and AF in this MetS cluster.

We noticed that the intersection of genotypic traits among MetS Clusters 1, 3, and 4 was highly expressed in the brain, specifically hypothalamus and pituitary, such as *CNIH2* (cornichon family AMPA receptor auxiliary protein 2), *TMEM151A* (transmembrane protein 151A), *C1QTNF4* (C1q and TNF related 4), and *MT3* (metallothionein 3). The neuroendocrine systems, especially the hypothalamus and pituitary, are involved in energy thermoregulation and satiety control.[62] Our results highlighted the importance of the neuroendocrine system shared by these three MetS endotypes. For instance, *C1QTNF4* modulated food intake patterns and systemic energy metabolism in obese mice.[63] *MT3* expression in hypothalamus of mice may be involved in leptin signaling and peripheral energy expenditure.[64]

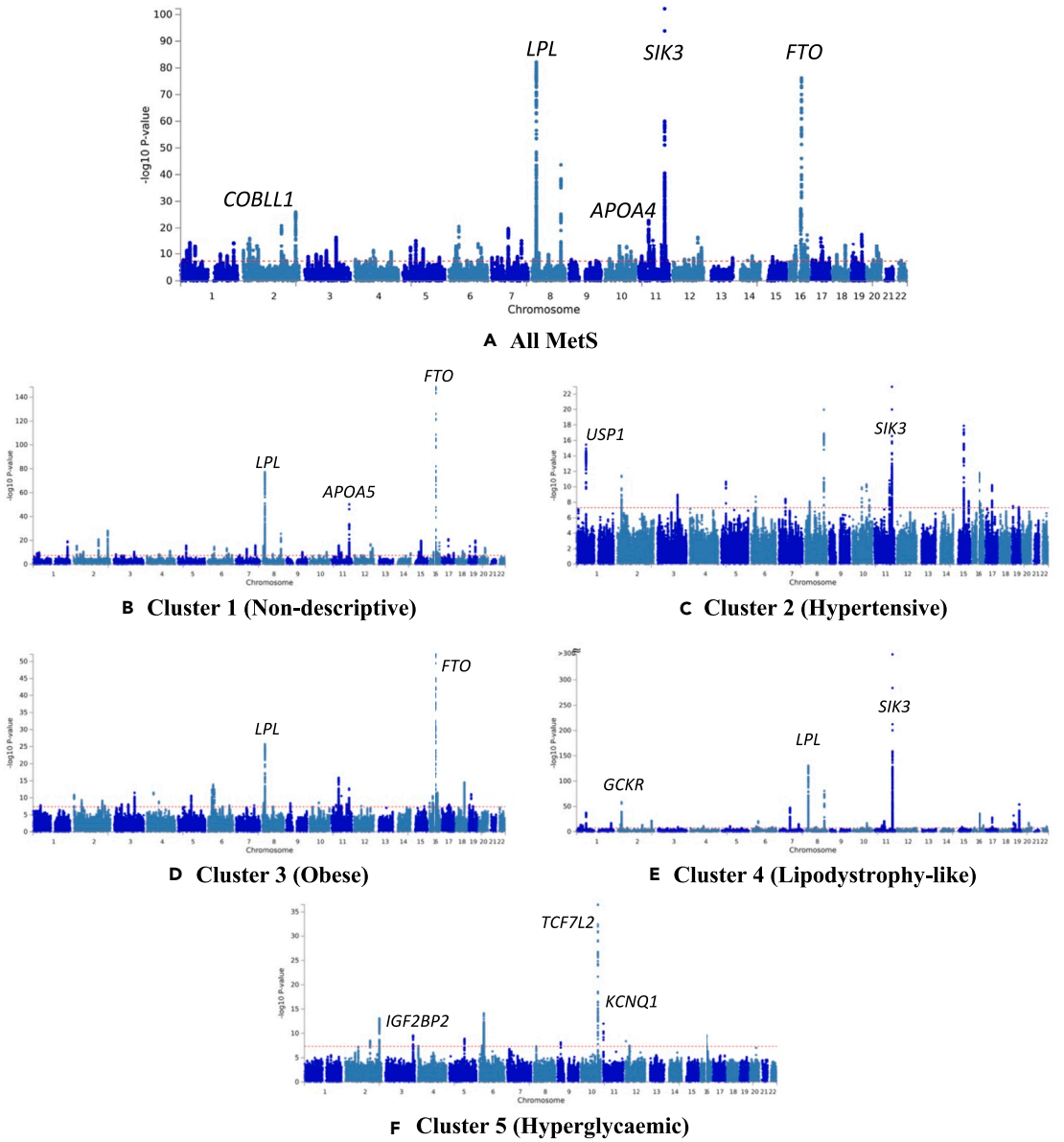
Cluster 2, despite its higher blood pressure traits, exhibited lower odds for CVD, similar to that observed in the pre-MetS group. This phenomenon may be attributed to the higher proportion of females in the cluster, which is generally considered a cardioprotective factor[65,66] and higher proportion of subcutaneous fat over visceral fat.[67] Additionally, the influence of menopausal status, particularly early menopause, as a risk factor for CVD, was notable. Adjustments for menopausal status reduced the risks for clinical outcomes, but these risks still exceeded those of the healthy and pre-MetS groups (Tables S6, and S7), suggesting menopausal status as a significant, yet not sole, factor in CVD risk. The absence of data on postmenopausal hormone therapy, which can impact cardiovascular risk,[68] is a limitation. Cluster 2 GWAS highlighted some brain-specific genes such as *C1QL1* (complement C1q like 1), *MAPT* (microtubule-associated protein tau), and *GFAP* (glial fibrillary acidic protein). *C1QL1*, *MAPT*, and *GFAP* are known genotypes associated with blood pressure traits.[69–71]

MetS Cluster 5 featured the role of type 2 diabetes and hyperglycemia as a critical component of MetS and also the strong association with CVD, chronic kidney disease, and neurovascular diseases. It should be noted that there were around 20% prevalence of T2D in other clusters, highlighting the heterogeneity of T2D. Despite being the smallest MetS cluster, GWAS of Cluster 5 managed to identify important well-known type 2 diabetes genes such as *TCF7L2* (transcription factor 7-like 2), *BBIP1* (BBSome interacting protein 1), GIN1 (gypsy retrotransposon integrase 1), and *IRS1* (insulin receptor substrate 1).[55,72,73] The discovery of multiple important type 2 diabetes genes, despite the small sample size of cases and lack of ancestral diversity compared to previously reported GWASs,[55,72,73] implied that size does not always matter but the homogeneity of the cases does.

GWAS signals had successfully identified drug targets for various complex diseases such as statins targeting *HMGCR* for lowering LDL,[74] ustekinumab and risankizumab repurposing for Crohn disease by targeting *IL23R*,[75] and antiarrhythmics in AF targeting *SCN5A*.[76] In our study, *SLC12A3* (solute carrier family 12 member 3) was unique to Cluster 1 and is a known drug target for thiazide diuretics,[77] which could imply that thiazide diuretics might be a useful drug class for Cluster 1 to reduce CVD risk.[78] *SLC5A11* (solute carrier family 5 member 11), also known as SGLT6, is one of the dapagliflozin target receptors.[79] Inhibition of SGLT6 by dapagliflozin had been reported to reduce oxidative stress, which could be beneficial in reversing diabetic cardiomyopathy.[80] *SLC5A11* was exclusively associated with Cluster 3. *NCAN* (neurocan), which was unique to Cluster 5, had been reported to be associated to T2D,[61] diastolic blood pressure,[81] and cholesterol[82] in previous GWAS. *NCAN* is a drug target of hyaluronic acid, a glycosaminoglycan commonly used in cosmetic treatment, wound healing, and joint pain.[83] Dimethyl fumarate (anti-inflammatory for multiple sclerosis)[84] and edasalonexent (still in clinical trials for type 2 diabetes and Duchenne muscular dystrophy)[85] target *RELA* (proto-oncogene, NF-kB subunit), which was shared by Clusters 1, 3, and 4. Phentermine, widely prescribed to promote weight loss, is an inhibitor of the sodium-dependent noradrenaline transporter encoded by *SLC6A2* (solute carrier family 6 member 2).[86] *SLC6A2* was specifically associated with Clusters 1, 3, and 5 in UKB on top of Cluster 3 of TWB. Lipid-lowering agents such as clofibrate and gemfibrozil target *LPL*, which was identified in both Cluster 1 and 4 of UKB and TWB. All these findings might indicate effective drug repurposing for specific MetS endotypes across two populations.

In our first-of-a-kind PRS comparison between the more homogeneous endotypes of MetS against the heterogeneous all MetS, we managed to show that accuracy of some PRS models can be improved by reducing the heterogeneity of cases especially for endotypes that are highly influenced by genetics. PRS for endotypes that are affected more by non-genetic factors or environmental influences will be less useful such as that for Cluster 2. We expect that our approach will assist in improving the accuracy of precision risk prediction with polygenic scores for complex diseases. However, this finding should be evaluated and further validated in other complex diseases with varying heritability.

MetS is a highly heterogeneous condition with multiple clinically relevant endotypes that are semi-distinctive in terms of phenotypic and genotypic traits across two populations. The identification of the endotypes is a key step toward precision medicine that could allow treatment stratification according to the phenotypic and genotypic traits of endotypes.



MetS Cluster	C1	C2	C3	C4	C5	All
# Genomic risk loci	38	11	19	32	5	64
# Lead SNPs	65	13	26	82	5	94
# Ind. Sig. SNPs	184	25	72	296	10	247
# Cluster-specific Ind. Sig. SNPs	122	22	38	233	9	-
# Candidate SNPs	4465	788	2767	6109	418	9066
# Mapped genes	481	58	256	461	32	800
# Cluster-specific mapped genes	156	16	98	133	8	-
# MAGMA gene-sets	20	4	2	9	3	4

Figure 4. GWAS Manhattan plots and summary statistics

(A–F) Manhattan plots for GWAS of overall metabolic syndrome and clusters 1–5 in the UK Biobank; summary of GWAS and FUMA SNP2GENE with cluster-specific independent significant SNPs and prioritized genes for MetS.

Figure 5. Comparative analysis of metabolic syndrome clusters using Jaccard similarity index: independent SNPs and associated genes in the UK Biobank

Limitations of the study

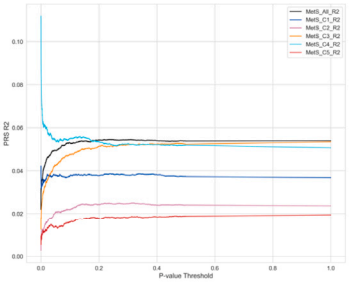
The major limitation of our study is that the hard clustering approach by k-means clustering might not be the most ideal method for identification of diseases subtypes as pathophysiology for diseases often overlap. Nonetheless, overlaps between subgroups in soft clustering might interfere with the assumptions of various statistical tests in subsequent analyses. The clustering based on phenotypic data also possessed an inherent limitation where phenotypic data tend to fluctuate with disease progression and are affected by pharmacotherapy; nevertheless, our MetS clusters still show semi-distinct differences in terms of genotypic data. One thing to note is that clustering is an NP-hard problem; implementation of other effective metaheuristics needs to be further explored when clustering based on genomic data.[87] The drug repurposing targets merely serve as indicators that certain group of drugs could be more effective for specific MetS endotypes. We also advise caution when interpreting the precision risk prediction results in clinical settings due to the minimal differences in accuracy across the MetS endotypes. More studies are needed to validate our findings for clinical implementation especially with further longitudinal follow-up data and prescription information. In our UKB application, participants' dates of birth are not accessible; consequently, determining the age at which diseases occur is not feasible, which restricts our ability to conduct Cox regression analyses. Furthermore, caution should be taken with regard to the interpretation on comparison between two distinct biobanks of different ancestry with contrasting environmental influences, diverse lifestyle factors, dissimilar genotyping techniques, and different laboratory investigation standards. Lastly, the addition of other omics[88] such as proteomics, metabolomics,[89] and metagenomics[90] will provide further insights to the pathophysiology of the different MetS endotypes.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Ethics, consent and permissions
- METHOD DETAILS
  - Unsupervised clustering for mets clusters
  - Genetic QC and GWAS
  - Genotypic traits comparison among MetS clusters
  - Precision drug repurposing

A (Entire sample size of cases)



Phenotype	Heritability (h^2)	Sample size (cases)
MetS_All	0.2319	103,996
MetS_C1	0.2462	33,707
MetS_C2	0.2198	23,215
MetS_C3	0.3138	30,089
MetS_C4	0.3243	13,116
MetS_C5	0.1702	3,869

B (Random sampling of 3,800 cases)

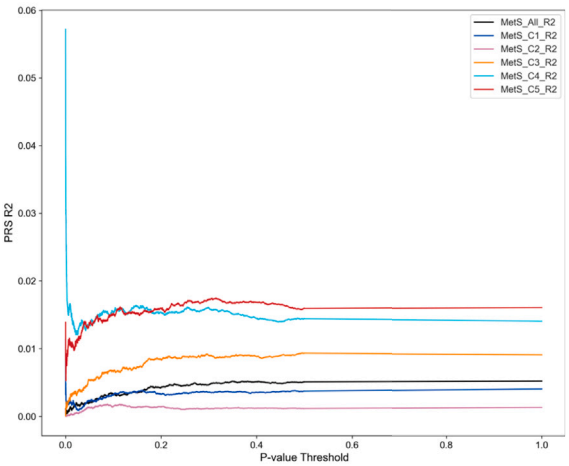


Figure 6. Predictive accuracy (R2) of PRS models for overall metabolic syndrome and clusters 1–5 in the UK Biobank

(A) PRS models constructed using the full sample size of cases.

(B) PRS models developed with random sampling of 3,800 cases.

- Precision disease risk prediction
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.109815.

Table 1. Selected precision drug repurposing targets for MetS endotypes identified through GREP by Anatomical Therapeutic Chemical (ATC) classification system

MetS Cluster	ATC Group Name	Odds Ratio	*p*-value	Target Gene: Drug Names
C1	Cardiovascular system	2.492	0.037	*ABCA1*: probucol *APOB*: mipomersen [a]*LPL*: clofibrate, gemfibrozil *SCN9A*: lidocaine *SLC12A3*: bendroflumethiazide, hydrochlorothiazide, chlorothiazide, polythiazide, quinethazone, metolazone *SLC12A4*: bumetanide *SLC6A2*: phentermine, amfepramone, mazindol, sibutramine
C3	Blood and blood forming system	8.004	0.01	*F2*: lepirudin, argatroban, melagatran, ximelagatran, bivalirudin, dabigatran etexilate, conestat alfa
	Anti-obesity agents	14.314	0.079	[a]*SLC6A2*: phentermine, amfepramone, mazindol, sibutramine *SLC5A11:* dapagliflozin
C4	Lipid-modifying agents	9.464	0.026	*APOB*: mipomersen [a]*LPL*: clofibrate, gemfibrozil
C5	Alimentary tract and metabolism	∞	0.004	*CES1*: cholic acid *SLC6A2*: phentermine, amfepramone, mazindol, sibutramine

[a]Similar drug repurposing targets in TWB.

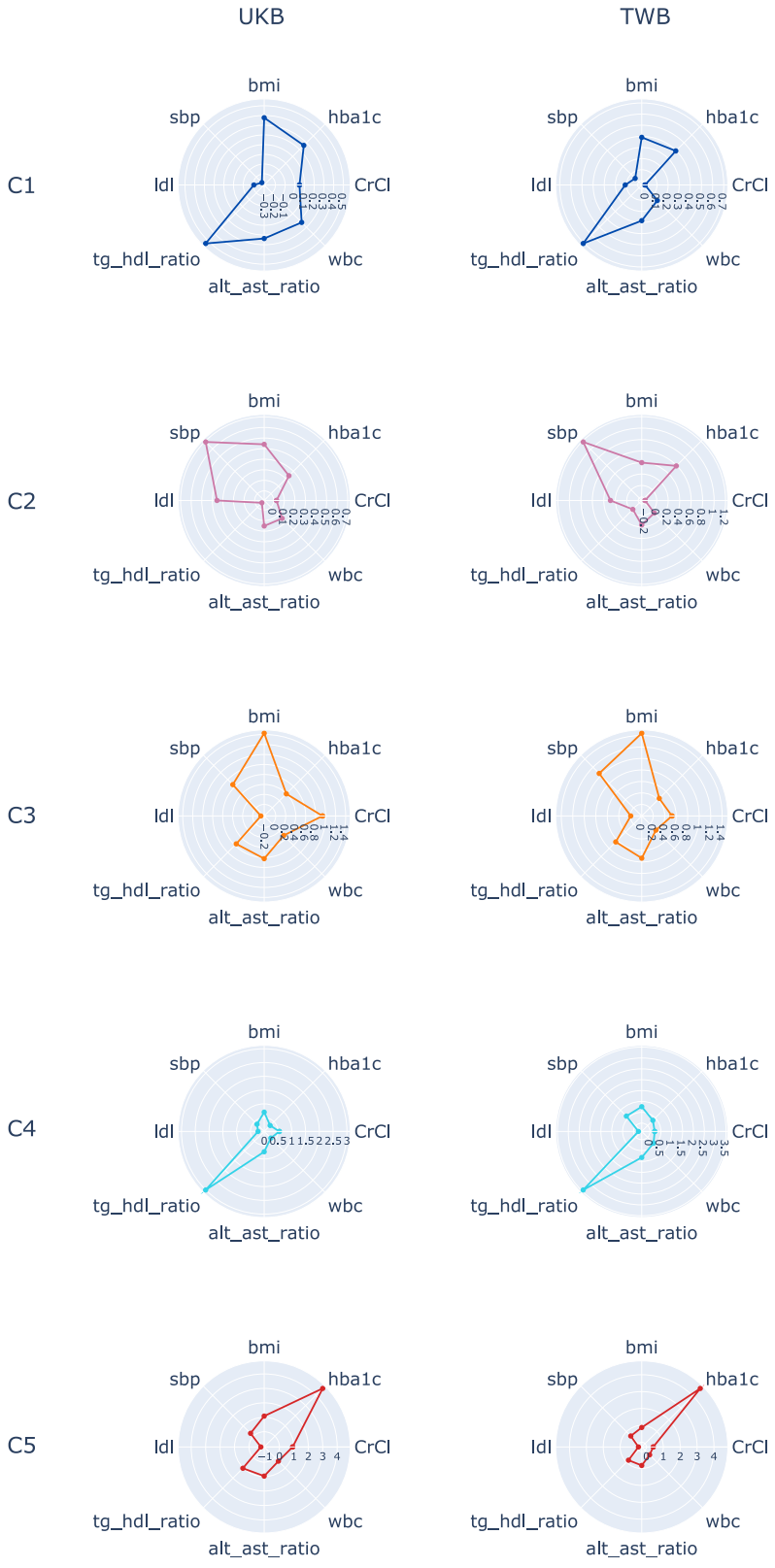


Figure 7. Radar plot of *Z* scores for MetS-related features (BMI, HbA1C, CrCl, WBC, ALT/AST ratio, TG/HDL ratio, LDL, and SBP) in metabolic syndrome clusters of the UK Biobank and Taiwan Biobank

ACKNOWLEDGMENTS

This project is supported by MOST 111-2314-B-001-008 grant from Ministry of Science and Technology of Taiwan. The authors appreciate UKB and TWB participants and all staff for their contribution. We thank the Institute of Biomedical Sciences, Academia Sinica for support.

AUTHOR CONTRIBUTIONS

A.L., E.L., P.C., and C.F. had full access to the study data and take responsibility for the integrity of the data and the accuracy of the data analysis. A.L. designed the study with contributions from E.L. A.L., E.L., and C.F. acquired the funding for the study. A.L. and E.L. contributed to data analysis and results interpretation. E.L. and A.L. were responsible for data visualization. A.L. drafted the manuscript. E.L. and A.L. revised the manuscript with contributions from C.F. and P.C. C.F. and P.C. supervised the project. All authors gave final approval of the version to be published.

DECLARATION OF INTERESTS

All authors declare no competing interests.

Received: September 29, 2023
Revised: February 2, 2024
Accepted: April 23, 2024
Published: April 25, 2024

REFERENCES

1. Hunter, D.J. (2005). Gene-environment interactions in human diseases. Nat. Rev. Genet. *6*, 287–298. https://doi.org/10.1038/nrg1578.
2. Duarte, C.W., Vaughan, L.K., Beasley, T.M., and Tiwari, H.K. (2013). Chapter 12 - Multifactorial Inheritance and Complex Diseases. In Emery and Rimoin's Principles and Practice of Medical Genetics, Sixth Edition, D. Rimoin, R. Pyeritz, and B. Korf, eds. (Academic Press), pp. 1–15. https://doi.org/10.1016/B978-0-12-383834-6.00014-8.
3. Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., and Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. Am. J. Hum. Genet. *101*, 5–22. https://doi.org/10.1016/j.ajhg.2017.06.005.
4. Chen, J., Spracklen, C.N., Marenne, G., Varshney, A., Corbin, L.J., Luan, J., Willems, S.M., Wu, Y., Zhang, X., Horikoshi, M., et al. (2021). The trans-ancestral genomic architecture of glycemic traits. Nat. Genet. *53*, 840–860. https://doi.org/10.1038/s41588-021-00852-9.
5. Malik, R., Chauhan, G., Traylor, M., Sargurupremraj, M., Okada, Y., Mishra, A., Rutten-Jacobs, L., Giese, A.K., van der Laan, S.W., Gretarsdottir, S., et al. (2018). Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. Nat. Genet. *50*, 524–537. https://doi.org/10.1038/s41588-018-0058-3.
6. Jansen, I.E., Savage, J.E., Watanabe, K., Bryois, J., Williams, D.M., Steinberg, S., Sealock, J., Karlsson, I.K., Hägg, S., Athanasiu, L., et al. (2019). Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. Nat. Genet. *51*, 404–413. https://doi.org/10.1038/s41588-018-0311-9.
7. Manchia, M., Cullis, J., Turecki, G., Rouleau, G.A., Uher, R., and Alda, M. (2013). The impact of phenotypic and genetic heterogeneity on results of genome wide association studies of complex diseases. PLoS One *8*, e76295. https://doi.org/10.1371/journal.pone.0076295.
8. Edwards, B.J., Haynes, C., Levenstien, M.A., Finch, S.J., and Gordon, D. (2005). Power and sample size calculations in the presence of phenotype errors for case/control genetic association studies. BMC Genet. *6*, 18. https://doi.org/10.1186/1471-2156-6-18.
9. Gordon, D., Haynes, C., Yang, Y., Kramer, P.L., and Finch, S.J. (2007). Linear trend tests for case-control genetic association that incorporate random phenotype and genotype misclassification error. Genet. Epidemiol. *31*, 853–870. https://doi.org/10.1002/gepi.20246.
10. Alberti, K.G.M.M., Eckel, R.H., Grundy, S.M., Zimmet, P.Z., Cleeman, J.I., Donato, K.A., Fruchart, J.C., James, W.P.T., Loria, C.M., Smith, S.C., et al. (2009). Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. Circulation *120*, 1640–1645. https://doi.org/10.1161/CIRCULATIONAHA.109.192644.
11. Kuk, J.L., and Ardern, C.I. (2010). Age and sex differences in the clustering of metabolic syndrome factors: association with mortality risk. Diabetes Care *33*, 2457–2461. https://doi.org/10.2337/dc10-0942.
12. Mottillo, S., Filion, K.B., Genest, J., Joseph, L., Pilote, L., Poirier, P., Rinfret, S., Schiffrin, E.L., and Eisenberg, M.J. (2010). The metabolic syndrome and cardiovascular risk a systematic review and meta-analysis. J. Am. Coll. Cardiol. *56*, 1113–1132. https://doi.org/10.1016/j.jacc.2010.05.034.
13. Daly, C.A., Hildebrandt, P., Bertrand, M., Ferrari, R., Remme, W., Simoons, M., and Fox, K.M.; EUROPA investigators (2007). Adverse prognosis associated with the metabolic syndrome in established coronary artery disease: data from the EUROPA trial. Heart *93*, 1406–1411. https://doi.org/10.1136/hrt.2006.113084.
14. Bertoluci, M.C., and Rocha, V.Z. (2017). Cardiovascular risk assessment in patients with diabetes. Diabetol. Metab. Syndr. *9*, 25. https://doi.org/10.1186/s13098-017-0225-1.
15. Abou Ziki, M.D., and Mani, A. (2016). Metabolic syndrome: genetic insights into disease pathogenesis. Curr. Opin. Lipidol. *27*, 162–171. https://doi.org/10.1097/MOL.0000000000000276.
16. Borshchev, Y.Y., Uspensky, Y.P., and Galagudza, M.M. (2019). Pathogenetic pathways of cognitive dysfunction and dementia in metabolic syndrome. Life Sci. *237*, 116932. https://doi.org/10.1016/j.lfs.2019.116932.
17. Esposito, K., Chiodini, P., Colao, A., Lenzi, A., and Giugliano, D. (2012). Metabolic syndrome and risk of cancer: a systematic review and meta-analysis. Diabetes Care *35*, 2402–2411. https://doi.org/10.2337/dc12-0336.
18. Mulugeta, A., Hyppönen, E., Ala-Korpela, M., and Mäkinen, V.P. (2022). Cross-sectional metabolic subgroups and 10-year follow-up of cardiometabolic multimorbidity in the UK Biobank. Sci. Rep. *12*, 8590. https://doi.org/10.1038/s41598-022-12198-1.
19. Kassi, E., Pervanidou, P., Kaltsas, G., and Chrousos, G. (2011). Metabolic syndrome: definitions and controversies. BMC Med. *9*, 48. https://doi.org/10.1186/1741-7015-9-48.
20. Moore, J.X., Chaudhary, N., and Akinyemiju, T. (2017). Metabolic Syndrome Prevalence by Race/Ethnicity and Sex in the United States, National Health and Nutrition Examination Survey, 1988-2012. Prev. Chronic Dis. *14*, E24. https://doi.org/10.5888/pcd14.160287.
21. Cameron, A.J., Magliano, D.J., Zimmet, P.Z., Welborn, T., and Shaw, J.E. (2007). The metabolic syndrome in Australia: prevalence using four definitions. Diabetes Res. Clin. Pract. *77*, 471–478. https://doi.org/10.1016/j.diabres.2007.02.002.

22. Sliem, H.A., Ahmed, S., Nemr, N., and El-Sherif, I. (2012). Metabolic syndrome in the Middle East. Indian J. Endocrinol. Metab. *16*, 67–71. https://doi.org/10.4103/2230-8210.91193.
23. Lim, S., Jang, H.C., Lee, H.K., Kimm, K.C., Park, C., and Cho, N.H. (2006). A rural-urban comparison of the characteristics of the metabolic syndrome by gender in Korea: the Korean Health and Genome Study (KHGS). J. Endocrinol. Invest. *29*, 313–319. https://doi.org/10.1007/BF03344102.
24. Yeh, C.J., Chang, H.Y., and Pan, W.H. (2011). Time trend of obesity, the metabolic syndrome and related dietary pattern in Taiwan: from NAHSIT 1993-1996 to NAHSIT 2005-2008. Asia Pac. J. Clin. Nutr. *20*, 292–300.
25. Gale, E.A.M. (2005). The myth of the metabolic syndrome. Diabetologia *48*, 1679–1683. https://doi.org/10.1007/s00125-005-1873-5.
26. Kahn, R., Buse, J., Ferrannini, E., and Stern, M. (2005). The metabolic syndrome: time for a critical appraisal. Joint statement from the American Diabetes Association and the European Association for the Study of Diabetes. Diabetologia *48*, 1684–1699. https://doi.org/10.1007/s00125-005-1876-2.
27. Fereshtehnejad, S.M., Zeighami, Y., Dagher, A., and Postuma, R.B. (2017). Clinical criteria for subtyping Parkinson's disease: biomarkers and longitudinal progression. Brain *140*, 1959–1976. https://doi.org/10.1093/brain/awx118.
28. Ahlqvist, E., Storm, P., Käräjämäki, A., Martinell, M., Dorkhan, M., Carlsson, A., Vikman, P., Prasad, R.B., Aly, D.M., Almgren, P., et al. (2018). Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. Lancet Diabetes Endocrinol. *6*, 361–369. https://doi.org/10.1016/s2213-8587(18)30051-2.
29. Anjana, R.M., Baskar, V., Nair, A.T.N., Jebarani, S., Siddiqui, M.K., Pradeepa, R., Unnikrishnan, R., Palmer, C., Pearson, E., and Mohan, V. (2020). Novel subgroups of type 2 diabetes and their association with microvascular outcomes in an Asian Indian population: a data-driven cluster analysis: the INSPIRED study. BMJ Open Diabetes Res. Care *8*, e001506. https://doi.org/10.1136/bmjdrc-2020-001506.
30. Zaharia, O.P., Strassburger, K., Strom, A., Bönhof, G.J., Karusheva, Y., Antoniou, S., Bódis, K., Markgraf, D.F., Burkart, V., Müssig, K., et al. (2019). Risk of diabetes-associated diseases in subgroups of patients with recent-onset diabetes: a 5-year follow-up study. Lancet Diabetes Endocrinol. *7*, 684–694. https://doi.org/10.1016/s2213-8587(19)30187-1.
31. Bello-Chavolla, O.Y., Bahena-Lopez, J.P., Vargas-Vazquez, A., Antonio-Villa, N.E., Marquez-Salinas, A., Fermin-Martinez, C.A., Rojas, R., Mehta, R., Cruz-Bautista, I., Hernandez-Jimenez, S., et al. (2020). Clinical characterization of data-driven diabetes subgroups in Mexicans using a reproducible machine learning approach. BMJ Open Diabetes Res Care *8*, e001550. https://doi.org/10.1136/bmjdrc-2020-001550.
32. Dennis, J.M., Shields, B.M., Henley, W.E., Jones, A.G., and Hattersley, A.T. (2019). Disease progression and treatment response in data-driven subgroups of type 2 diabetes compared with models based on simple clinical features: an analysis using clinical trial data. Lancet Diabetes Endocrinol. *7*, 442–451. https://doi.org/10.1016/s2213-8587(19)30087-7.
33. Anderson, G.P. (2008). Endotyping asthma: new insights into key pathogenic mechanisms in a complex, heterogeneous disease. Lancet *372*, 1107–1119. https://doi.org/10.1016/S0140-6736(08)61452-X.
34. Yaghootkar, H., Scott, R.A., White, C.C., Zhang, W., Speliotes, E., Munroe, P.B., Ehret, G.B., Bis, J.C., Fox, C.S., Walker, M., et al. (2014). Genetic evidence for a normal-weight "metabolically obese" phenotype linking insulin resistance, hypertension, coronary artery disease, and type 2 diabetes. Diabetes *63*, 4369–4377. https://doi.org/10.2337/db14-0318.
35. Udler, M.S., Kim, J., von Grotthuss, M., Bonàs-Guarch, S., Cole, J.B., Chiou, J., Christopher D Anderson on behalf of METASTROKE and the ISGC, Boehnke, M., Laakso, M., Atzmon, G., et al. (2018). Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: A soft clustering analysis. PLoS Med. *15*, e1002654. https://doi.org/10.1371/journal.pmed.1002654.
36. Pulit, S.L., Stoneman, C., Morris, A.P., Wood, A.R., Glastonbury, C.A., Tyrrell, J., Yengo, L., Ferreira, T., Marouli, E., Ji, Y., et al. (2019). Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. Hum. Mol. Genet. *28*, 166–174. https://doi.org/10.1093/hmg/ddy327.
37. Sakaue, S., Kanai, M., Tanigawa, Y., Karjalainen, J., Kurki, M., Koshiba, S., Narita, A., Konuma, T., Yamamoto, K., Akiyama, M., et al. (2021). A cross-population atlas of genetic associations for 220 human phenotypes. Nat. Genet. *53*, 1415–1424. https://doi.org/10.1038/s41588-021-00931-x.
38. Schmiedel, B.J., Singh, D., Madrigal, A., Valdovino-Gonzalez, A.G., White, B.M., Zapardiel-Gonzalo, J., Ha, B., Altay, G., Greenbaum, J.A., McVicker, G., et al. (2018). Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. Cell *175*, 1701–1715.e16. https://doi.org/10.1016/j.cell.2018.10.022.
39. Monaco, G., Lee, B., Xu, W., Mustafah, S., Hwang, Y.Y., Carré, C., Burdin, N., Visan, L., Ceccarelli, M., Poidinger, M., et al. (2019). RNA-Seq Signatures Normalized by mRNA Abundance Allow Absolute Deconvolution of Human Immune Cell Types. Cell Rep. *26*, 1627–1640.e7. https://doi.org/10.1016/j.celrep.2019.01.041.
40. Abate, W., Alrammah, H., Kiernan, M., Tonks, A.J., and Jackson, S.K. (2020). Lysophosphatidylcholine acyltransferase 2 (LPCAT2) co-localises with TLR4 and regulates macrophage inflammatory gene expression in response to LPS. Sci. Rep. *10*, 10355. https://doi.org/10.1038/s41598-020-67000-x.
41. Würtz, P., Havulinna, A.S., Soininen, P., Tynkkynen, T., Prieto-Merino, D., Tillin, T., Ghorbani, A., Artati, A., Wang, Q., Tiainen, M., et al. (2015). Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. Circulation *131*, 774–785. https://doi.org/10.1161/CIRCULATIONAHA.114.013116.
42. Würtz, P., Mäkinen, V.P., Soininen, P., Kangas, A.J., Tukiainen, T., Kettunen, J., Savolainen, M.J., Tammelin, T., Viikari, J.S., Rönnemaa, T., et al. (2012). Metabolic signatures of insulin resistance in 7,098 young adults. Diabetes *61*, 1372–1380. https://doi.org/10.2337/db11-1355.
43. van der Harst, P., van Setten, J., Verweij, N., Vogler, G., Franke, L., Maurano, M.T., Wang, X., Mateo Leach, I., Eijgelsheim, M., Sotoodehnia, N., et al. (2016). 52 Genetic Loci Influencing Myocardial Mass. J. Am. Coll. Cardiol. *68*, 1435–1448. https://doi.org/10.1016/j.jacc.2016.07.729.
44. Morita, H., Rehm, H.L., Menesses, A., McDonough, B., Roberts, A.E., Kucherlapati, R., Towbin, J.A., Seidman, J.G., and Seidman, C.E. (2008). Shared genetic causes of cardiac hypertrophy in children and adults. N. Engl. J. Med. *358*, 1899–1908. https://doi.org/10.1056/NEJMoa075463.
45. Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. Science *347*, 1260419. https://doi.org/10.1126/science.1260419.
46. Karlsson, M., Zhang, C., Méar, L., Zhong, W., Digre, A., Katona, B., Sjöstedt, E., Butler, L., Odeberg, J., Dusart, P., et al. (2021). A single-cell type transcriptomics map of human tissues. Sci. Adv. *7*, 1275. https://doi.org/10.1126/sciadv.abh2169.
47. Wei, Y., Chen, K., Whaley-Connell, A.T., Stump, C.S., Ibdah, J.A., and Sowers, J.R. (2008). Skeletal muscle insulin resistance: role of inflammatory cytokines and reactive oxygen species. Am. J. Physiol. Regul. Integr. Comp. Physiol. *294*, R673–R680. https://doi.org/10.1152/ajpregu.00561.2007.
48. Jean, N., Somers, V.K., Sochor, O., Medina-Inojosa, J., Llano, E.M., and Lopez-Jimenez, F. (2014). Normal-weight obesity: implications for cardiovascular health. Curr. Atheroscler. Rep. *16*, 464. https://doi.org/10.1007/s11883-014-0464-7.
49. Chamberlain, A.M., Agarwal, S.K., Ambrose, M., Folsom, A.R., Soliman, E.Z., and Alonso, A. (2010). Metabolic syndrome and incidence of atrial fibrillation among blacks and whites in the Atherosclerosis Risk in Communities (ARIC) Study. Am. Heart J. *159*, 850–856. https://doi.org/10.1016/j.ahj.2010.02.005.
50. Mitchell, G.F., Vasan, R.S., Keyes, M.J., Parise, H., Wang, T.J., Larson, M.G., D'Agostino, R.B., Kannel, W.B., Levy, D., and Benjamin, E.J. (2007). Pulse pressure and risk of new-onset atrial fibrillation. JAMA *297*, 709–715. https://doi.org/10.1001/jama.297.7.709.
51. Verdecchia, P., Dagenais, G., Healey, J., Gao, P., Dans, A.L., Chazova, I., Binbrek, A.S., Iacobellis, G., Ferreira, R., Holwerda, N., et al. (2012). Blood pressure and other determinants of new-onset atrial fibrillation in patients at high cardiovascular risk in the Ongoing Telmisartan Alone and in Combination With Ramipril Global Endpoint Trial/Telmisartan Randomized AssessmeNt Study in ACE iNtolerant subjects with cardiovascular Disease studies. J. Hypertens. *30*, 1004–1014. https://doi.org/10.1097/HJH.0b013e3283522a51.
52. Movahed, M.R., Hashemzadeh, M., and Jamal, M.M. (2005). Diabetes mellitus is a strong, independent risk for atrial fibrillation and flutter in addition to other cardiovascular disease. Int. J. Cardiol. *105*,

315–318. https://doi.org/10.1016/j.ijcard.2005.02.050.
53. Wanahita, N., Messerli, F.H., Bangalore, S., Gami, A.S., Somers, V.K., and Steinberg, J.S. (2008). Atrial fibrillation and obesity–results of a meta-analysis. Am. Heart J. *155*, 310–315. https://doi.org/10.1016/j.ahj.2007.10.004.
54. Frayling, T.M., Timpson, N.J., Weedon, M.N., Zeggini, E., Freathy, R.M., Lindgren, C.M., Perry, J.R.B., Elliott, K.S., Lango, H., Rayner, N.W., et al. (2007). A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science *316*, 889–894. https://doi.org/10.1126/science.1141634.
55. Mahajan, A., Taliun, D., Thurner, M., Robertson, N.R., Torres, J.M., Rayner, N.W., Payne, A.J., Steinthorsdottir, V., Scott, R.A., Grarup, N., et al. (2018). Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nat. Genet. *50*, 1505–1513. https://doi.org/10.1038/s41588-018-0241-6.
56. Wood, A.R., Tyrrell, J., Beaumont, R., Jones, S.E., Tuke, M.A., Ruth, K.S., GIANT consortium, Yaghootkar, H., Freathy, R.M., Murray, A., et al. (2016). Variants in the FTO and CDKAL1 loci have recessive effects on risk of obesity and type 2 diabetes, respectively. Diabetologia *59*, 1214–1221. https://doi.org/10.1007/s00125-016-3908-5.
57. Graff, M., Scott, R.A., Justice, A.E., Young, K.L., Feitosa, M.F., Barata, L., Winkler, T.W., Chu, A.Y., Mahajan, A., Hadley, D., et al. (2017). Genome-wide physical activity interactions in adiposity - A meta-analysis of 200,452 adults. PLoS Genet. *13*, e1006528. https://doi.org/10.1371/journal.pgen.1006528.
58. Aguilera, C.M., Gomez-Llorente, C., Tofe, I., Gil-Campos, M., Cañete, R., and Gil, Á. (2015). Genome-wide expression in visceral adipose tissue from obese prepubertal children. Int. J. Mol. Sci. *16*, 7723–7737. https://doi.org/10.3390/ijms16047723.
59. Bellenguez, C., Küçükali, F., Jansen, I.E., Kleineidam, L., Moreno-Grau, S., Amin, N., Naj, A.C., Campos-Martin, R., Grenier-Boley, B., Andrade, V., et al. (2022). New insights into the genetic etiology of Alzheimer's disease and related dementias. Nat. Genet. *54*, 412–436. https://doi.org/10.1038/s41588-022-01024-z.
60. Speliotes, E.K., Willer, C.J., Berndt, S.I., Monda, K.L., Thorleifsson, G., Jackson, A.U., Lango Allen, H., Lindgren, C.M., Luan, J., Mägi, R., et al. (2010). Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat. Genet. *42*, 937–948. https://doi.org/10.1038/ng.686.
61. Vujkovic, M., Keaton, J.M., Lynch, J.A., Miller, D.R., Zhou, J., Tcheandjieu, C., Huffman, J.E., Assimes, T.L., Lorenz, K., Zhu, X., et al. (2020). Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. Nat. Genet. *52*, 680–691. https://doi.org/10.1038/s41588-020-0637-y.
62. Benarroch, E.E. (2007). Thermoregulation: recent concepts and remaining questions. Neurology *69*, 1293–1297. https://doi.org/10.1212/01.wnl.0000275537.71623.8e.
63. Sarver, D.C., Stewart, A.N., Rodriguez, S., Little, H.C., Aja, S., and Wong, G.W. (2020). Loss of CTRP4 alters adiposity and food intake behaviors in obese mice. Am. J. Physiol. Endocrinol. Metab. *319*, E1084–E1100. https://doi.org/10.1152/ajpendo.00448.2020.
64. Byun, H.R., Kim, D.K., and Koh, J.Y. (2011). Obesity and downregulated hypothalamic leptin receptors in male metallothionein-3-null mice. Neurobiol. Dis. *44*, 125–132. https://doi.org/10.1016/j.nbd.2011.06.012.
65. Walli-Attaei, M., Joseph, P., Rosengren, A., Chow, C.K., Rangarajan, S., Lear, S.A., AlHabib, K.F., Davletov, K., Dans, A., Lanas, F., et al. (2020). Variations between women and men in risk factors, treatments, cardiovascular disease incidence, and death in 27 high-income, middle-income, and low-income countries (PURE): a prospective cohort study. Lancet *396*, 97–109. https://doi.org/10.1016/S0140-6736(20)30543-2.
66. Bots, S.H., Peters, S.A.E., and Woodward, M. (2017). Sex differences in coronary heart disease and stroke mortality: a global assessment of the effect of ageing between 1980 and 2010. BMJ Glob. Health *2*, e000298. https://doi.org/10.1136/bmjgh-2017-000298.
67. Després, J.P., Carpentier, A.C., Tchernof, A., Neeland, I.J., and Poirier, P. (2021). Management of Obesity in Cardiovascular Practice: JACC Focus Seminar. J. Am. Coll. Cardiol. *78*, 513–531. https://doi.org/10.1016/j.jacc.2021.05.035.
68. Rozenberg, S., Vandromme, J., and Antoine, C. (2013). Postmenopausal hormone therapy: risks and benefits. Nat. Rev. Endocrinol. *9*, 216–227. https://doi.org/10.1038/nrendo.2013.17.
69. Plotnikov, D., Huang, Y., Khawaja, A.P., Foster, P.J., Zhu, Z., Guggenheim, J.A., and He, M. (2022). High Blood Pressure and Intraocular Pressure: A Mendelian Randomization Study. Invest. Ophthalmol. Vis. Sci. *63*, 29. https://doi.org/10.1167/iovs.63.6.29.
70. Kichaev, G., Bhatia, G., Loh, P.R., Gazal, S., Burch, K., Freund, M.K., Schoech, A., Pasaniuc, B., and Price, A.L. (2019). Leveraging Polygenic Functional Enrichment to Improve GWAS Power. Am. J. Hum. Genet. *104*, 65–75. https://doi.org/10.1016/j.ajhg.2018.11.008.
71. Giri, A., Hellwege, J.N., Keaton, J.M., Park, J., Qiu, C., Warren, H.R., Torstenson, E.S., Kovesdy, C.P., Sun, Y.V., Wilson, O.D., et al. (2019). Trans-ethnic association study of blood pressure determinants in over 750,000 individuals. Nat. Genet. *51*, 51–62. https://doi.org/10.1038/s41588-018-0303-9.
72. Mahajan, A., Spracklen, C.N., Zhang, W., Ng, M.C.Y., Petty, L.E., Kitajima, H., Yu, G.Z., Rüeger, S., Speidel, L., Kim, Y.J., et al. (2022). Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. Nat. Genet. *54*, 560–572. https://doi.org/10.1038/s41588-022-01058-3.
73. Spracklen, C.N., Horikoshi, M., Kim, Y.J., Lin, K., Bragg, F., Moon, S., Suzuki, K., Tam, C.H.T., Tabara, Y., Kwak, S.H., et al. (2020). Identification of type 2 diabetes loci in 433,540 East Asian individuals. Nature *582*, 240–245. https://doi.org/10.1038/s41586-020-2263-3.
74. Teslovich, T.M., Musunuru, K., Smith, A.V., Edmondson, A.C., Stylianou, I.M., Koseki, M., Pirruccello, J.P., Ripatti, S., Chasman, D.I., Willer, C.J., et al. (2010). Biological, clinical and population relevance of 95 loci for blood lipids. Nature *466*, 707–713. https://doi.org/10.1038/nature09270.
75. Duerr, R.H., Taylor, K.D., Brant, S.R., Rioux, J.D., Silverberg, M.S., Daly, M.J., Steinhart, A.H., Abraham, C., Regueiro, M., Griffiths, A., et al. (2006). A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. Science *314*, 1461–1463. https://doi.org/10.1126/science.1135245.
76. Nielsen, J.B., Thorolfsdottir, R.B., Fritsche, L.G., Zhou, W., Skov, M.W., Graham, S.E., Herron, T.J., McCarthy, S., Schmidt, E.M., Sveinbjornsson, G., et al. (2018). Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. Nat. Genet. *50*, 1234–1239. https://doi.org/10.1038/s41588-018-0171-3.
77. Preston, R.A., Afshartous, D., Caizapanta, E.V., Materson, B.J., Rodco, R., Alonso, E., and Alonso, A.B. (2021). Thiazide-Sensitive NCC (Sodium-Chloride Cotransporter) in Human Metabolic Syndrome: Sodium Sensitivity and Potassium-Induced Natriuresis. Hypertension *77*, 447–460. https://doi.org/10.1161/HYPERTENSIONAHA.120.15933.
78. Tsujimoto, T., and Kajio, H. (2020). Thiazide Use and Decreased Risk of Heart Failure in Nondiabetic Patients Receiving Intensive Blood Pressure Treatment. Hypertension *76*, 432–441. https://doi.org/10.1161/HYPERTENSIONAHA.120.15154.
79. Wicik, Z., Nowak, A., Jarosz-Popek, J., Wolska, M., Eyileten, C., Siller-Matula, J.M., von Lewinski, D., Sourij, H., Filipiak, K.J., and Postuła, M. (2022). Characterization of the SGLT2 Interaction Network and Its Regulation by SGLT2 Inhibitors: A Bioinformatic Analysis. Front. Pharmacol. *13*, 901340. https://doi.org/10.3389/fphar.2022.901340.
80. Van Steenbergen, A., Balteau, M., Ginion, A., Ferté, L., Battault, S., Ravenstein, C.d.M.d., Balligand, J.L., Daskalopoulos, E.P., Gilon, P., Despa, F., et al. (2017). Sodium-myoinositol cotransporter-1, SMIT1, mediates the production of reactive oxygen species induced by hyperglycemia in the heart. Sci. Rep. *7*, 41166. https://doi.org/10.1038/srep41166.
81. Surendran, P., Feofanova, E.V., Lahrouchi, N., Ntalla, I., Karthikeyan, S., Cook, J., Chen, L., Mifsud, B., Yao, C., Kraja, A.T., et al. (2020). Discovery of rare variants associated with blood pressure regulation through meta-analysis of 1.3 million individuals. Nat. Genet. *52*, 1314–1332. https://doi.org/10.1038/s41588-020-00713-x.
82. Dönertaş, H.M., Fabian, D.K., Valenzuela, M.F., Partridge, L., and Thornton, J.M. (2021). Common genetic associations between age-related diseases. Nat. Aging *1*, 400–412. https://doi.org/10.1038/s43587-021-00051-5.
83. Zhou, X.H., Brakebusch, C., Matthies, H., Oohashi, T., Hirsch, E., Moser, M., Krug, M., Seidenbecher, C.I., Boeckers, T.M., Rauch, U., et al. (2001). Neurocan is dispensable for brain development. Mol. Cell Biol. *21*, 5970–5978. https://doi.org/10.1128/MCB.21.17.5970-5978.2001.
84. Naismith, R.T., Wundes, A., Ziemssen, T., Jasinska, E., Freedman, M.S., Lembo, A.J., Selmaj, K., Bidollari, I., Chen, H., Hanna, J., et al. (2020). Diroximel Fumarate Demonstrates an Improved Gastrointestinal Tolerability Profile Compared with Dimethyl Fumarate in Patients

with Relapsing-Remitting Multiple Sclerosis: Results from the Randomized, Double-Blind, Phase III EVOLVE-MS-2 Study. CNS Drugs *34*, 185–196. https://doi.org/10.1007/s40263-020-00700-0.

85. Donovan, J.M., Zimmer, M., Offman, E., Grant, T., and Jirousek, M. (2017). A Novel NF-κB Inhibitor, Edasalonexent (CAT-1004), in Development as a Disease-Modifying Treatment for Patients With Duchenne Muscular Dystrophy: Phase 1 Safety, Pharmacokinetics, and Pharmacodynamics in Adult Subjects. J. Clin. Pharmacol. *57*, 627–639. https://doi.org/10.1002/jcph.842.
86. Proietto, J., Fam, B.C., Ainslie, D.A., and Thorburn, A.W. (2000). Novel anti-obesity drugs. Expert Opin. Investig. Drugs *9*, 1317–1326. https://doi.org/10.1517/13543784.9.6.1317.
87. Mahajan, M., Nimbhorkar, P., and Varadarajan, K. (2012). The planar k-means problem is NP-hard. Theor. Comput. Sci. *442*, 13–21.
88. Hasin, Y., Seldin, M., and Lusis, A. (2017). Multi-omics approaches to disease. Genome Biol. *18*, 83. https://doi.org/10.1186/s13059-017-1215-1.
89. Adamski, J. (2012). Genome-wide association studies with metabolomics. Genome Med. *4*, 34. https://doi.org/10.1186/gm333.
90. Cox, T.O., Lundgren, P., Nath, K., and Thaiss, C.A. (2022). Metabolic control by the microbiome. Genome Med. *14*, 80. https://doi.org/10.1186/s13073-022-01092-0.
91. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. Nature *562*, 203–209. https://doi.org/10.1038/s41586-018-0579-z.
92. Wei, C.Y., Yang, J.H., Yeh, E.C., Tsai, M.F., Kao, H.J., Lo, C.Z., Chang, L.P., Lin, W.J., Hsieh, F.J., Belsare, S., et al. (2021). Genetic profiles of 103,106 individuals in the Taiwan Biobank provide insights into the health and history of Han Chinese. NPJ Genom. Med. *6*, 10. https://doi.org/10.1038/s41525-021-00178-9.
93. Grundy, S.M., Cleeman, J.I., Daniels, S.R., Donato, K.A., Eckel, R.H., Franklin, B.A., Gordon, D.J., Krauss, R.M., Savage, P.J., Smith, S.C., et al. (2005). Diagnosis and management of the metabolic syndrome: an American Heart Association/National Heart, Lung, and Blood Institute Scientific Statement. Circulation *112*, 2735–2752. https://doi.org/10.1161/CIRCULATIONAHA.105.169404.
94. Grundy, S.M. (2016). Metabolic syndrome update. Trends Cardiovasc. Med. *26*, 364–373. https://doi.org/10.1016/j.tcm.2015.10.004.
95. Julkunen, H., Cichońska, A., Tiainen, M., Koskela, H., Nybo, K., Mäkelä, V., Nokso-Koivisto, J., Kristiansson, K., Perola, M., Salomaa, V., et al. (2023). Atlas of plasma NMR biomarkers for health and disease in 118,461 individuals from the UK Biobank. Nat. Commun. *14*, 604. https://doi.org/10.1038/s41467-023-36231-7.
96. MacQueen, J. (1967). Classification and Analysis of Multivariate Observations, pp. 281–297.
97. Peter, J.R. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. *20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7.
98. Vassilvitskii, S., and Arthur, D. (2006). k-means++: The Advantages of Careful Seeding, pp. 1027–1035.
99. Pedregosa, F. (2011). Scikit-Learn: Machine Learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.
100. Mbatchou, J., Barnard, L., Backman, J., Marcketta, A., Kosmicki, J.A., Ziyatdinov, A., Benner, C., O'Dushlaine, C., Barber, M., Boutkov, B., et al. (2021). Computationally efficient whole-genome regression for quantitative and binary traits. Nat. Genet. *53*, 1097–1103. https://doi.org/10.1038/s41588-021-00870-7.
101. Watanabe, K., Taskesen, E., van Bochoven, A., and Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. Nat. Commun. *8*, 1826. https://doi.org/10.1038/s41467-017-01261-5.
102. Lind, L. (2020). Genetic Determinants of Clustering of Cardiometabolic Risk Factors in U.K. Biobank. Metab. Syndr. Relat. Disord. *18*, 121–127. https://doi.org/10.1089/met.2019.0096.
103. Sakaue, S., and Okada, Y. (2019). GREP: genome for REPositioning drugs. Bioinformatics *35*, 3821–3823. https://doi.org/10.1093/bioinformatics/btz166.
104. Wishart, D.S., Feunang, Y.D., Guo, A.C., Lo, E.J., Marcu, A., Grant, J.R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., et al. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res. *46*, D1074–D1082. https://doi.org/10.1093/nar/gkx1037.
105. Li, Y.H., Yu, C.Y., Li, X.X., Zhang, P., Tang, J., Yang, Q., Fu, T., Zhang, X., Cui, X., Tu, G., et al. (2018). Therapeutic target database update 2018: enriched resource for facilitating bench-to-clinic research of targeted therapeutics. Nucleic Acids Res. *46*, D1121–D1127. https://doi.org/10.1093/nar/gkx1076.
106. Bulik-Sullivan, B.K., Loh, P.R., Finucane, H.K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics Consortium, Patterson, N., Daly, M.J., Price, A.L., and Neale, B.M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. *47*, 291–295. https://doi.org/10.1038/ng.3211.
107. 1000 Genomes Project Consortium, Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56–65. https://doi.org/10.1038/nature11632.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
UK Biobank dataset	https://www.ukbiobank.ac.uk/	The application number is 46789.
Taiwan Biobank dataset	https://www.biobank.org.tw/english.php	The application number is TWBR10503-02.
Software and algorithms		
Python; RRID: SCR_008394	https://www.python.org/	Version 3.10.6
scikit-learn; RRID: SCR_002577	https://scikit-learn.org/	Version 1.2.2
SciPy; RRID: SCR_008058	https://scipy.org/	Version 1.10.1
MatPlotLib; RRID: SCR_008624	https://matplotlib.org/	Version 3.7.1
Plotly; RRID: SCR_013991	https://plotly.com/python/	Version 5.3.1
seaborn; RRID: SCR_018132	https://seaborn.pydata.org	Version 0.11.2

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Cathy SJ Fann (csjfann@ibms.sinica.edu.tw).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The genetic and phenotype datasets are not publicly available but can be accessed via the UK Biobank data access process (http://www.ukbiobank.ac.uk/register-apply/) and through Taiwan Biobank (https://taiwanview.twbiobank.org.tw/data_appl). GWAS summary statistics are available on GWAS Catalog (https://www.ebi.ac.uk/gwas/).
- This paper does not report original code. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

UKB is a population-based prospective cohort that recruited 502,637 individuals aged 37–73 from year 2006–2010 across the UK. Similarly, TWB is a population-based prospective cohort that recruited 127,708 adult individuals from 2012 to 2019 across Taiwan. Full details of the UKB and TWB have been reported in Bycroft et al.[91] and Wei et al..[92] Among the half a million participants in UKB, 94·7% individuals are of European ancestry. In contrast, TWB mainly consists of East Asian ancestry, specifically Han Chinese with over 99% of the entire cohort.

Ethics, consent and permissions

This study has been approved by the institutional review board of Academia Sinica (AS-IRB01-21065) and conducted according to the principles of Declaration of Helsinki. All participants gave informed consent when joining biobanks, which allow for sharing of all anonymized data to authorized researchers. Participants can withdraw consent to sharing of their data at any stages of their participation.

METHOD DETAILS

Based on the ATP III criteria, we classified individuals as having MetS when fulfilled three or more of the criteria.[93,94] Individuals were defined as pre-MetS if they met one or two of the MetS criteria.

Based on the Third Report of the National Cholesterol Education Program Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (ATP III) criteria[93,94].

(1) Systolic blood pressure ≥130 mmHg or diastolic blood pressure ≥85 mmHg or on antihypertensive treatment or diagnosed or self-reported to have hypertension
(2) Serum glucose ≥100 mg/dL (5·6 mmol/L) or antidiabetic treatment or diagnosed or self-reported to have T2D

(3) Serum triglycerides ≥150 mg/dL (1·7 mmol/L) or on cholesterol medications
(4) Waist circumference ≥102 cm in men and ≥88 cm in women for Caucasians
(5) Waist circumference ≥90 cm in men and ≥80 cm in women for Asians
(6) HDL-C level <40 mg/dL (1·0 mmol/L) for men and <50 mg/dL (1·3 mmol/L) for women

Data extraction and transformation was conducted on UKB research analysis platform (RAP) through JupyterLab and Google Colab. UKB also contains a comprehensive plasma nuclear magnetic resonance (NMR) biomarker data encompassing 249 measurements related to lipoprotein lipids, fatty acids, as well as small molecules like amino acids, ketones, and glycolysis metabolites.[95] Clinical outcomes are defined as shown in Table S1 and composite CVD outcome which encompasses CVD defined by ICD10 codes, self-reported conditions, mortality due to CVD, and coronary revascularization procedure as shown in Table S2. Missing data was dropped during analysis. The TWB phenotypic data were pre-processed using pipeline similar to that used for the UKB data, and outliers were removed with Peirce's criterion.

Unsupervised clustering for mets clusters

The k-means clustering is a centroid-based unsupervised learning technique that divides a dataset into various clusters by minimizing the within-cluster variances.[96] In our study, we applied the k-means clustering algorithm to the z-scores of MetS criteria, which include waist circumference, mean arterial pressure (MAP), serum glucose, triglyceride, and HDL cholesterol. This allowed us to identify MetS sub-clusters. The optimal number of clusters (k) was determined through the elbow method and silhouette coefficient.[97] For improved initialization of the algorithm, we chose k-means++ over naive k-means.[98] We also significantly increased the maximum number of iterations to ensure the exploration of feature space for elevated accuracy. For the implementation, we utilized the sklearn.cluster.KMeans() class from the scikit-learn module in Python.[99] In addition to that, we have also performed sex-specific clustering for UKB. A k-means model was computed with the UKB data, and this model was used to predict the cluster memberships of the TWB data.

Genetic QC and GWAS

Genotyping and imputation of UKB and TWB were performed as previously described.[91,92] Individuals with ambiguous sex (different sex and genetic sex), sex chromosome aneuploidy, ten or more third-degree relatives identified, non-white British ancestry, who are outliers in heterozygosity and missing rates were removed. For TWB, similar QC of individuals were performed but as TWB consists of purely homogeneous Han Chinese ancestry, no filtering based on ancestry was conducted. GWAS of each MetS clusters with healthy control were performed through two-stage REGENIE v3.1.1 [101]. Individuals with high genotype missingness (>10%) were also removed. Variants (both genotype and imputation) with high genotype missingness (>10%), low Hardy-Weinberg equilibrium (HWE) *p*-value ($<1x10^{-15}$), minor allele count (MAC) < 20, minor allele frequency (MAF) < 0·01, and sample missing rate were filtered out prior to step 1 and 2 of REGENIE[100] using PLINK 2.0 v1.0.6. Imputation information score was set as >0.8.

GWAS of each MetS clusters with healthy control were performed through REGENIE v3.1.1 [101]. Step 1 standard logistic regression was conducted with leave-one out cross validation and size of the genotype blocks of 1000 markers. Step 2 was conducted through Firth logistic regression on variants with *p*-value <0.01 from the standard logistic regression and size of the genotype blocks of 200. Covariates of age, age,[2] first 20 genetic principal components, and sex were included in both steps of REGENIE. Genome-wide significance was determined as *p*-value $<1X10^{-9}$, multiple testing correction for six GWAS. PLINK and REGENIE are conducted through RAP Swiss Army Knife v4.7.1 for UKB.

Post-GWAS functional mapping and annotation was conducted through FUMA SNP2GENE and GENE2FUNCTION.[101] Independent significant SNPS were defined with r^2 threshold of ≥0.6 and genome-wide significant *p*-value of $1 X 10^{-9}$, a Bonferroni multiple correction for six GWAS from the standard threshold of $5 X 10^{-8}$; and lead SNPs with a further r^2 threshold of ≥0·1. For TWB, a least stringent genome-wide significant *p*-value of $5 X 10^{-8}$ was used instead. All candidate SNPs were annotated using built-in ANNOVAR with UKB release 2b 10k White British as reference panel and with 1000 genomes Phase 3 East Asian reference panel. Annotated SNPs were mapped through positional mapping (physical distances of 10kb, expression quantitative trait locus (eQTL) mapping (SNPs that likely affect expression of genes up to 1Mb), and chromatin interaction mapping with FDR threshold $\leq 1 X 10^{-6}$. For eQTL mapping, only GTEx v8 tissue types were selected with eQTL maximum *p*-value $\leq 1 X 10^{-3}$. Gene mapping is filtered based on functional annotation with CADD score ≥12·37, RegulomeDB score ≥7, and maximum state of chromatin ≤7 from all tissue/cell types available. Tissue-expression analysis of prioritized genes were done through FUMA[101] GENE2FUNCTION using data from GTEx v8 53 specific tissue types. In addition to that, FUMA also performed MAGMA gene and gene-set analysis with gene windows of 10kb.

Genotypic traits comparison among MetS clusters

We compared the independent significant single nucleotide polymorphisms (SNPs) and mapped genes amongst the MetS clusters, and defined SNPs and genes unique to each cluster as cluster-specific genotypes. For numerical comparison of the different MetS clusters, we employed Jaccard and cosine similarity coefficients on both independent significant SNPs and mapped genes. We also conducted pairwise genotype comparison (SNPs and genes) of the five MetS clusters with GWAS of all MetS to examine if the GWAS of MetS clusters can reveal novel findings. Mapped genes from MetS clusters were also compared to that from GWAS of five MetS components (glucose, systolic blood pressure, triglycerides, waist circumference, and high-density lipoprotein-cholesterol) from Lind.[102]

Precision drug repurposing

FUMA GENE2FUNCTION assigned Drug IDs to UniProt IDs of genes that are targets of drugs, enabling identification of cluster-specific drug repurposing targets for each endotype. GREP (Genome for REPositioning drugs) identified drugs and drug classes enriched for genes prioritized from GWAS results categorized in gene-sets by Anatomical Therapeutic Chemical (ATC) classification system.[103] GREP utilized two major drug databases, Drug Bank[104] and Therapeutic Target Database.[105] Fisher's exact tests were performed to inspect whether the genes associated with each endotype were enriched in genes targeted by drugs in gene-sets according to ATC classification system.

Precision disease risk prediction

To test the hypothesis that homogeneity of cases can improve the performance of polygenic risk score prediction in complex disease, we constructed and compared PRS models for MetS endotypes with all MetS. The dataset was split 50:50 into a baseline dataset and a target dataset for all MetS and MetS clusters. The baseline dataset was used to conduct a GWAS with the exact same parameters, and resulting GWAS summary statistics were used to construct PRS models with PRSice2 (v1.0.2). The remaining target dataset was used for PRS model testing to compare the performance.

In order to reduce the influence of sample sizes on the performance of PRS models, we randomly sampled 3,800 cases and 38,000 controls for all MetS and MetS clusters, and repeated the analysis process as that for the entire sample size of cases. 3,800 cases were selected based on the smallest MetS cluster, MetS C5. In brief, PRS models were constructed by averaging the dosage of each variant associated with the traits (all MetS or MetS endotypes), weighted by the effect size from the corresponding GWAS summary statistics. Furthermore, we estimated the SNP-based heritability (h^2) of all MetS and MetS clusters using the LD Score regression (ldsc) software (v1.0.1)[106] with LD Scores from the European ancestry samples of the 1000 Genomes Project.[107]

QUANTIFICATION AND STATISTICAL ANALYSIS

To compare the phenotypic traits among different MetS clusters, quantitative traits were compared through one-way analysis of variance (-ANOVA) and post-hoc analysis with Tukey's HSD test. The associations of different MetS clusters with clinical outcomes were analysed through age and sex adjusted multivariate logistic regression models with healthy individuals as comparator group; further adjusted for T2D status. The statistical significance of the associations with MetS clusters was evaluated using a p-value <0.001 based on a Bonferroni adjustment for performing 50 tests (25 clinical outcomes and two models: T2D-unadjusted and T2D-adjusted). Relative risks and adjusted relative risk calculation using Poisson regression with robust error variances were also calculated for incidence of multiple common cancers. We further investigated the role of menopausal status in subset of female individuals from UKB through menopausal status adjusted and unadjusted logistic regression. For TWB, as only self-reported diagnosis was available, comparable self-reported clinical outcomes (coronary artery disease, arrythmia, cardiomyopathy, stroke, chronic kidney disease, depression, and cancer) were analysed using a similar pipeline.