scientific reports

OPEN

The overlap of genetic susceptibility to schizophrenia and cardiometabolic disease can be used to identify metabolically different groups of individuals

Rona J. Strawbridge[1,2,3]✉, Keira J. A. Johnston[1,4,5], Mark E. S. Bailey[5], Damiano Baldassarre[6,7], Breda Cullen[1], Per Eriksson[3], Ulf deFaire[8], Amy Ferguson[1,9], Bruna Gigante[3], Philippe Giral[10], Nicholas Graham[1], Anders Hamsten[3], Steve E. Humphries[11], Sudhir Kurl[12], Donald M. Lyall[1], Laura M. Lyall[1], Jill P. Pell[1], Matteo Pirro[13], Kai Savonen[14,15], Andries J. Smit[16], Elena Tremoli[7], Tomi-Pekka Tomainen[17], Fabrizio Veglia[7], Joey Ward[1], Bengt Sennblad[18] & Daniel J. Smith[1]

Understanding why individuals with severe mental illness (Schizophrenia, Bipolar Disorder and Major Depressive Disorder) have increased risk of cardiometabolic disease (including obesity, type 2 diabetes and cardiovascular disease), and identifying those at highest risk of cardiometabolic disease are important priority areas for researchers. For individuals with European ancestry we explored whether genetic variation could identify sub-groups with different metabolic profiles. Loci associated with schizophrenia, bipolar disorder and major depressive disorder from previous genome-wide association studies and loci that were also implicated in cardiometabolic processes and diseases were selected. In the IMPROVE study (a high cardiovascular risk sample) and UK Biobank (general population sample) multidimensional scaling was applied to genetic variants implicated in both psychiatric and cardiometabolic disorders. Visual inspection of the resulting plots used to identify distinct clusters. Differences between these clusters were assessed using chi-squared and Kruskall-Wallis tests. In IMPROVE, genetic loci associated with both schizophrenia and cardiometabolic disease (but not bipolar disorder or major depressive disorder) identified three groups of individuals with distinct metabolic profiles. This grouping was replicated within UK Biobank, with somewhat less distinction between metabolic profiles. This work focused on individuals of European ancestry and is unlikely

[1]Institute of Health and Wellbeing, University of Glasgow, Room 111, Public Health, 1 Lilybank Gardens, Glasgow G12 8RZ, UK. [2]Health Data Research, London, UK. [3]Cardiovascular Medicine Unit, Department of Medicine Solna, Karolinska Institute, Stockholm, Sweden. [4]Deanery of Molecular, Genetic and Population Health Sciences, College of Medicine and Veterinary Medicine, University of Edinburgh, Edinburgh, Scotland, UK. [5]School of Life Sciences, College of Medical, Veterinary & Life Sciences, University of Glasgow, Glasgow, Scotland, UK. [6]Department of Medical Biotechnology and Translational Medicine, Università degli Studi di Milano, Milan, Italy. [7]Centro Cardiologico Monzino, IRCCS, Milan, Italy. [8]Cardiovascular and Nutritional Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden. [9]Usher Institute, University of Edinburgh, Edinburgh, UK. [10]Service Endocrinologie-Metabolisme, Groupe Hôpitalier Pitie-Salpetriere, Unités de Prévention Cardiovasculaire, Assistance Publique - Hopitaux de Paris, Paris, France. [11]Centre for Cardiovascular Genetics, Institute Cardiovascular Science, University College London, London, UK. [12]Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland. [13]Internal Medicine, Angiology and Arteriosclerosis Diseases, Department of Clinical and Experimental Medicine, University of Perugia, Perugia, Italy. [14]Foundation for Research in Health Exercise and Nutrition, Kuopio Research Institute of Exercise Medicine, Kuopio, Finland. [15]Department of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital, Kuopio, Finland. [16]Department of Medicine, University Medical Center Groningen and University of Groningen, Groningen, The Netherlands. [17]Public Health and Clinical Nutrition, Department of Medicine, University of Eastern Finland, Kupiou, Finland. [18]Department of Cell and Molecular Biology, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Uppsala University, Uppsala, Sweden. ✉email: rona.strawbridge@glasgow.ac.uk

to apply to more genetically diverse populations. Overall, this study provides proof of concept that common biology underlying mental and physical illness may help to stratify subsets of individuals with different cardiometabolic profiles.

Individuals with serious mental illness (such as schizophrenia (SCZ), major depressive disorder (MDD) and bipolar disorder (BD)) have a reduced life expectancy (10–15 years for BD, 15–20 years for SCZ[1]). This is likely due to the well-established increased prevalence of cardiovascular and metabolic disorders compared to the general population. For example, obesity is up to 3.5-fold higher in those with SCZ[2], type 2 diabetes is ~ twofold higher in those with MDD, BD or SCZ[2], and cerebrovascular disease is increased by up to 3.3-fold in those with BD[2]. Understanding this increased risk and identifying individuals at highest risk of metabolic and cardiovascular disease are important priority areas for researchers and healthcare providers.

Historically, the increased risk and prevalence of cardiometabolic disease (CMD) has been attributed to social determinants and lifestyle factors (including poor diet, sedentary behaviour, alcohol and substance use) that co-exist with serious mental illness and effects of psychotropic medication[2], however there is growing evidence that there might be common biological mechanisms underlying both mental and psychiatric illness. As genetic data is stable over an individual's lifetime, and not influenced by disease course, genetic approaches are ideal for investigation of common biology in comorbid conditions. The identification of genetic variants robustly associated with a wide range of psychiatric and cardiometabolic phenotypes by international genetics consortia has enabled the exploration of relationships between psychiatric and cardiometabolic conditions.

Genome-wide genetic correlations between psychiatric and cardiometabolic traits provide evidence for underlying common biology. Correlations have been described between depression and obesity (rg = 0.12) or cardiovascular disease (rg = 0.42)[3]. Evidence of causal relationships between psychiatric and cardiometabolic traits have also been described[1,4,5]. However, the mechanisms involved have yet to be uncovered and therefore this knowledge has had no clinical impact.

Here we tested whether a novel approach using multi-dimensional scaling (MDS) of genetic variation associated with psychiatric and cardiometabolic disorders could aid stratification of individuals into groups with differing cardiometabolic risk profiles.

Results

The IMPROVE and UK Biobank studies.
The demographic characteristics of the IMPROVE, UK Biobank subsets 1 (UKB1) and 2 (UKB2) are provided in Table 1. At baseline, individuals in IMPROVE (a European high cardiovascular-risk cohort) were older, more overweight and more likely to have T2D, hypertension or medication for hypertension or lipid-lowering medication than the UKB subsets (self-reported white British general population cohort). UKB1 and UKB2 were very similar, with lower frequency of hypertension at follow-up in UKB1 (51.5%) compared to UKB2 (62.0%) but slightly larger carotid Intima-media thickness (cIMT, indicative of vessel wall remodelling) measures in UKB2 to UKB1. Despite different proportions of UKB1 and UKB2 completing the mental health questionnaire, the frequencies of BD, MDD and GAD were similar.

Figure 1 provides a schematic overview of the analysis procedure.

SCZ-CM loci can identify metabolically distinct groups of individuals in IMPROVE.
When using IMPROVE and single nucleotide polymorphisms (SNPs) with a minor allele frequency (MAF) > 1%, implicated in both SCZ and CMD (SCZ-CMD), plotting the first two multi-dimensional scaling components (C1 and C2) demonstrated 3 groups of individuals (by visual inspection) (Fig. 2a). Separation was predominantly due to C1, and whilst C1 is nominally significantly correlated with latitude (rho = − 0.036, p = 0.0339), the clustering is not being driven by latitude (Supplementary Fig. 1). SNPs with MAF as low as 1% might differ across populations (even within the same ancestry grouping), therefore robustness to MAF threshold also assessed. When using MAF > 5% showed additional groups (Fig. 2b), whereas MAF > 10% showed similar groups to MAF > 1% (Fig. 2c). Assignment to groups was consistent using MAF > 1% and MAF > 10% (Supplementary Table 1). The three groups appear to have modest differences in cardiometabolic profiles (Table 2): Group 3 had a significantly lower frequency of hypertension (group 3: 74% vs groups 1 or 2: 80% or 81% respectively, P = 0.004) and lower fastest progression of cIMT (group 3: 0.156 mm vs groups 1 or 2: 0.176 mm or 0.166 mm, P = 0.002). This is surprising given the (non-significant) higher rates of smoking in this group. Group 2 had (non-significantly) lower rates of T2D than the other groups (group 2: 25% vs groups 1 or 3: 28%). Similar groups were observed using T-distributed Stochastic Neighbour Embedding (tSNE) or principal component analyses (PCA, Supplementary Methods), with the majority of individuals being consistently grouped together (Supplementary Figs. 2 and 3, respectively).

This result appears specific to SCZ-CMD SNP subset; no separation into groups was observed when using MDD-CMD SNPs, irrespective of the MAF filter used (Fig. 2d–f). For BD-CMD SNPs (Fig. 2g–i), grouping is apparent at MAF > 1%, but not when MAF > 5% or 10% were considered.

Validation of method and sensitivity testing of clustering in UKB1.
In order to assess whether MDS analysis of SCZ-CMD SNPs could reproducibly identify three groups of individuals, validation of the method was attempted in UKB1. Firstly, to directly replicate the analysis conducted in IMPROVE (Fig. 3a,b), the post-filtering SNPs from IMPROVE were used (Fig. 3c); however the grouping is not convincing as there is little separation between the groups. Secondly, to assess robustness of the method to differences in MAF and LD structure between populations, the SCZ-CMD SNPs were filtered for MAF and LD in UKB1. As noted in Fig. 1, the majority of SNPs included in the two approaches were the same. Unsurprisingly, the SNPs that differed were

	IMPROVE	UKB1	UKB2
Nmax	3300	2202	20,182
Male (%)	1695 (51.4)	1042 (47.3)	9759 (48.4)
Baseline			
Age (years)	64.2 (5.4)	55.7 (7.6)	55.2 (7.5)
Weight (kg)	76.7 (15.1)	76.1 (14.5)	76.9 (14.8)
Waist (cm)	94 (13)	88 (12)	88 (13)
Hip (cm)	102 (10)	102 (8)	102 (8)
WHR	0.92 (0.09)	0.86 (0.09)	0.86 (0.09)
BMI (kg/m^2)	27.3 (4.2)	26.4 (3.9)	26.6 (4.2)
SBP (mmHg)	142 (18)	136 (18)	136 (18)
DBP (mmHg)	82 (10)	82 (10)	82 (10)
SBP* (mmHg)	151 (21)	138 (19)	138 (19)
DBP* (mmHg)	88 (11)	83 (11)	83 (11)
T2D	880 (26.7)	46 (2.1)	45 (2.2)
HTN	2634 (79.8)	974 (45.0)	8765 (45.2)
HTN medication	1904 (57.7)	332 (15.2)	2820 (14.0)
Lipid-lowering medication	1623 (49.2)	172 (18.4)	1677 (18.9)
ISH	0 (0)	21 (1.0)	259 (1.3)
IMTmean (mm)	0.891 (0.199)	na	na
IMTmax (mm)	2.037 (0.813)	na	na
Current smoking	498 (15.1)	139 (6.3)	1234 (6.1)
Former smoking	1216 (36.9)	746 (33.9)	6656 (33.0)
Follow-up			
Age	66.7 (5.4)	61.8 (7.5)	63.2 (7.5-x)
Weight	na	75.6 (14.6)	76.3 (15.0)
Waist	na	86 (12)	88 (12)
Hip	na	100 (8)	101 (9)
WHR	na	0.86 (0.08)	0.87 (0.09)
BMI	na	26.4 (4.0)	26.5 (4.4)
SBP	na	138 (20)	137 (18)
DBP	na	83 (11)	79 (10)
SBP*	na	140 (20)	141 (20)
DBP*	na	81 (11)	81 (11)
T2D	na	79 (3.6)	867 (4.3)
HTN	na	1121 (51.5)	12,414 (62.0)
HTN medication	na	466 (21.2)	4568 (22.7)
Lipid-lowering medication	na	408 (22.3)	4557 (26.4)
ISH	119 (3.6)	43 (2.0)	333 (1.7)
IMTmean (mm)	na	0.672 (0.119)	0.682 (0.125)
IMTmax (mm)	na	0.888 (0.182)	0.914 (0.297)
Progression of IMTmean	0.0186 (0.032)	na	na
Progression of IMTmax	0.0439 (0.163)	na	na
Current smoking	na	99 (4.6)	708 (3.5)
Former smoker	na	742 (34.1)	6800 (33.9)
MHQ			
Nmax (% of group)	na	1528 (69.4)	10,079 (49.9)
BD	na	28 (1.8)	196 (1.4)
MDD	na	410 (29.9)	3512 (28.6)

Table 1. Demographic characteristics of IMPROVE and UKB participants. Where: *, adjusted to provide estimates of treatment-naïve levels, as per Ehret etl al; na, not available.

mainly those with MAF < 10%. Using SCZ-CMD and conducting MAF and LD filtering in UKB1, nine groups are evident when using SNPs with MAF > 1% (Fig. 3d), whereas three groups are observed when using SNPs with MAF > 10% (Fig. 3e). When comparing the metabolic profiles of the 3 groups, no significant differences
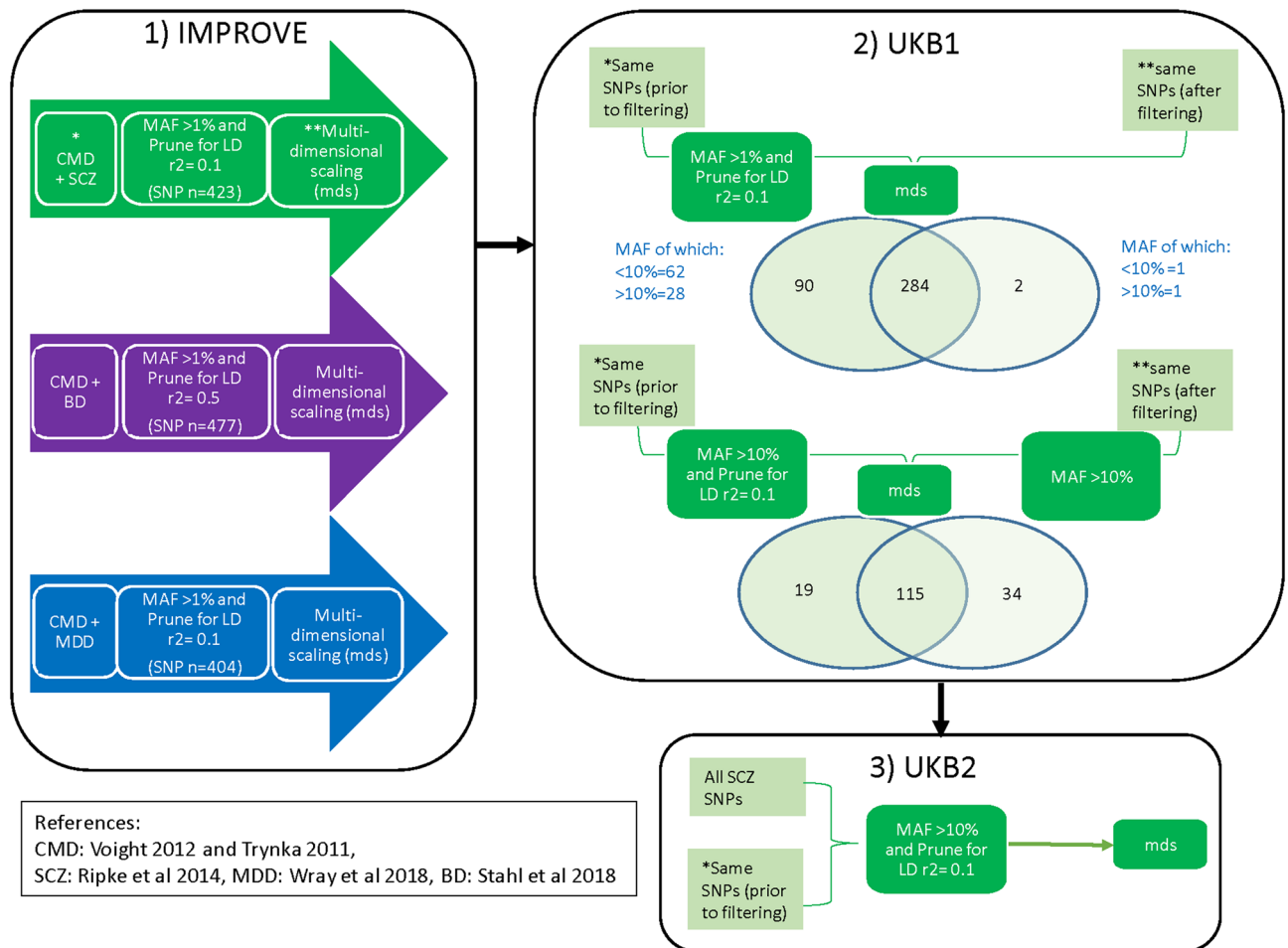
Figure 1. Schematic of the analysis procedure used to identify clusters.

were seen (Fig. 4a and Table 3). This is unsurprising, given that it is a smaller cohort with a lower cardiovascular burden.

Validation of metabolic differences between clusters in UKB2. In an attempt to replicate the clustering and validate the metabolic differences between groups, the larger UKB2 subset was analysed. As filtering with MAF > 10% and 1% gave similar clusters, filtering with MAF > 10% was applied as it is more likely to generalise to other populations. Again, three major groups were identified (Fig. 4b), similar to those identified in IMPROVE and UKB1. Additional clusters between the major three groups were apparent, but they account for ~ 7% of the studied population, and were omitted from the groups.

Consistent with the IMPROVE study, clinically modest (and statistically significant) differences were observed in baseline SBP, SDP adjusted for blood-pressure medication, and frequency of hypertension and T2D (Table 3). These effects were not observed at follow-up, potentially due to lifestyle or medications changes in response to baseline observations. It was also noted that the frequency of MDD but not BD differed between the groups. The number of SCZ in UKB2 is too low to provide meaningful statistics.

Impact of MDD/BD on clusters. As phenotypes and genetic loci for SCZ overlap with those for MDD and BD, it is perhaps unsurprising to see that the clusters include different proportions of individuals with MDD. To investigate whether these individuals were driving the clustering, the process was repeated in those without BD/MDD separately from those with these diagnoses (using SNPs with MAF > 10%). In those without mental illness, similar to the overall UKB2, there were there main groups, intermediate clusters accounting for 7.4% of the sample (Fig. 5a). In those with mental illness the three clusters were observed, with better between-group separation and only 1.3% of the sample being ungrouped (Fig. 5b). Small but significant differences between groups were observed for blood pressure measures and rates of hypertension, in both those with and without mental illness (Supplementary Table 2). These results suggest that this method is applicable to the general population, as well as those with increased genetic burden for mental illness.

All genetic loci associated with SCZ do not identify clusters in UKB. To determine whether it is common biology (ie. Overlap in loci for SCZ and CMD) per se, rather than SCZ in general that drives the clus-
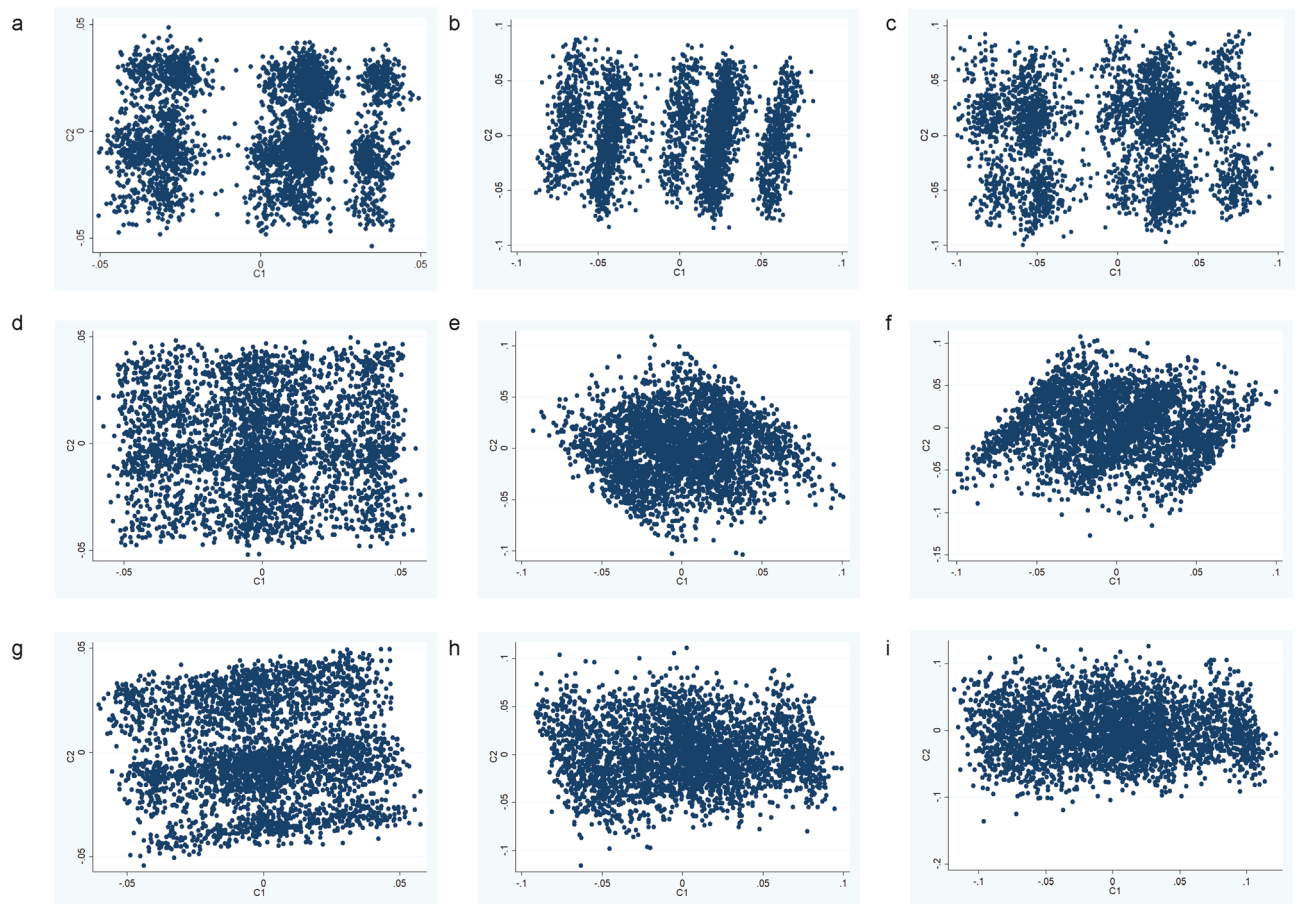
Figure 2. Results of MDS analysis in IMPROVE, using the loci in common between CMD and SCZ with (**a**) MAF > 1%, (**b**) MAF > 5% or (**c**) MAF > 10%; CMD and MDD with (**d**) MAF > 1%, (**e**) MAF > 5% or (**f**) MAF > 10%; CMD and BD with (**g**) MAF > 1%, (**h**) MAF > 5% or (**i**) MAF > 10%. Each data point is an individual therefore the individuals who are closer together are more genetically similar.

tering, the same procedure was followed using all SNPs in loci associated with SCZ in UKB2, with the same MAF and LD filtering being applied prior to MDS analysis. As shown in Supplementary Fig. 4, SNPs in loci associated with SCZ do not separate individuals into groups. A further "negative control" experiment was conducted in UKB2. When repeating the analysis using the genetic loci (Supplementary Table 4) associated with eye colour[6,7], there was no evidence of subgroups (Supplementary Fig. 5). These results confirm that it is the overlap of SCZ and CMD loci (rather than a methodological artefact), and therefore probably common biological mechanisms, which are driving the clustering.

Discussion

This study provides proof of principle that, using the genetic overlap between SCZ and cardiometabolic disorders, subsets of European ancestry individuals with different metabolic profiles can be identified. These findings support the existence of mechanisms common to SCZ and blood pressure regulation.

The discovery cohort IMPROVE deliberately recruited to identify genes and biomarkers associated with the risk of cardiovascular diseases, at a time when psychiatric disorders were typically excluded from non-psychiatric studies, therefore only a portion of the spectrum of psychiatric genetic burden is represented. In contrast, UKB1 and UKB2 are general population cohorts and therefore have a wider spectrum of both psychiatric and cardiometabolic disorder genetic burden, although it is recognised that the recruitment skews this distribution towards to the healthier segment of the population[8]. It is therefore both striking that the grouping was present in IMPROVE, and unsurprising that the blood pressure and hypertension differences between groups were more modest in UKB2 than those in IMPROVE.

It is worth noting that similar groups were observed in the IMPROVE cohort, using three different methods and (where applicable) exploring a variety of parameter settings. This suggests that the grouping is robust. The metabolic profiles of the groups did not completely agree between the 3 cohorts, however the repeated observation of between-group differences in T2D and blood pressure/hypertension deserves further attention. If the method can be refined to better identify whether an individual is at increased risk of either hypertension or T2D would be of immense value. Even if the method is only robust in high CMD-risk populations (such as those with family history, multiple risk factors or psychiatric diagnoses), it could be of clinical importance.

	1	2	3	P**
N	1222 (36.2)	1629 (48.2)	526 (15.6)	
Men (%)	581 (47.6)	842 (51.7)	268 (51.0)	0.954
Age (years)	64.2 (5.4)	64.2 (5.4)	64.4 (5.4)	0.558
BMI (kg/m^2)	27.4 (4.3)	27.2 (4.2)	27.0 (4.2)	0.259
Waist (cm)	94.5 (12.8)	93.8 (12.4)	93.9 (12.6)	0.298
Waist_hip	0.92 (0.09)	0.92 (0.09)	0.94 (12.6)	0.859
SBP (mmHg)	142 (18)	142 (19)	140 (18)	0.070
DBP (mmHg)	82 (10)	82 (10)	82 (10)	0.599
HTN	982 (80.4)	1321 (81.1)	389 (74.0)	**0.004**
HTN medication	694 (56.8)	805 (58.1)	291 (55.3)	0.421
SBP* (mmHg)*	151 (20)	151 (21)	148 (21)	0.080
DBP* (mmHg)*	88 (11)	88 (11)	87 (12)	0.368
NSAIDs	253 (20.7)	260 (18.8)	102 (19.4)	0.622
Current smoking	172 (14.1)	246 (15.1)	87 (16.5)	0.614
Pack years	10.7 (17.2)	10.7 (15.9)	12.3 (19.7)	0.191
T2D	342 (28.0)	414 (25.4)	145 (27.6)	0.252
Lipid-lowering medication	585 (47.9)	828 (50.9)	256 (48.8)	0.159
Framingham risk score	0.27 (0.16)	0.27 (0.16)	0.27 (0.16)	0.743
Cardiac event	74 (6.1)	94 (5.8)	22 (4.2)	0.208
Baseline				
CC-IMTmean	0.740 (0.130)	0.745 (0.147)	0.749 (0.155)	0.962
IMTmean	0.895 (0.198)	0.888 (0.199)	0.888 (0.205)	0.296
CC-IMTmax	1.184 (0.368)	1.197 (0.400)	1.222 (0.456)	0.705
IMTmax	2.051 (0.812)	2.026 (0.803)	2.040 (0.849)	0.502
IMTmeanmax	1.258 (0.295)	1.252 (0.298)	1.247 (0.306)	0.300
Diameter	7.847 (0.854)	7.834 (0.862)	7.796 (0.839)	0.441
Progression				
CC-IMTmean	0.009 (0.029)	0.009 (0.027)	0.006 (0.026)	0.193
IMTmean	0.019 (0.032)	0.019 (0.032)	0.017 (0.033)	0.508
CC-IMTmax	0.016 (0.096)	0.017 (0.093)	0.007 (0.098)	0.281
IMTmax	0.049 (0.164)	0.042 (0.156)	0.035 (0.178)	0.209
IMTmeanmax	0.026 (0.053)	0.026 (0.051)	0.023 (0.054)	0.240
IMTfastest	0.176 (0.149)	0.166 (0.139)	0.156 (0.147)	**0.002**
Diameter	0.003 (0.030)	0.005 (0.033)	0.003 (0.029)	0.749

Table 2. Demographic characteristics of the IMPROVE participants, by cluster (MAF > 10%). Highlighted in bold are the significant (p < 0.05) differences between groups. Where: *, adjusted to provide estimates of treatment-naïve levels as per Ehret et al.; Statistical analyses compared levels or frequncies across groups 1, 2 and 3. Ungrouped (.) were omitted from the analyses). **P for Pearsons chi square for categorical variables and Kruskal–Wallis for continuous variables; na, not available.

It is interesting that the analyses using BD and MDD genetic loci did not enable clustering of individuals in the same way as was observed for SCZ, particularly given that BD and SCZ demonstrate an overlap in genetic loci. There are several possible explanations for this, most notably the ability to identify genetic loci for each mental illness: SCZ is clinically a more severe phenotype with diagnostic criteria that are relatively specific (for example psychotic episodes). In comparison, MDD spans a wide spectrum severity, with phenotypic heterogeneity potentially diluting or obscuring some true genetic effects. Whilst BD can be considered an intermediate (some symptoms more severe than MDD, most are less severe than for SCZ) diagnostic criteria for MDD and BD overlap to a large degree as both involve episodes of depression, meaning that there is potential for misdiagnosis and therefore dilution of genetic effects for either trait. Another explanation is that the mechanisms leading to CMD in SCZ differ from those in MDD or BD, with processes that are represented on the CardioMetabo and Immuno chips failing to capture some pathological mechanisms. With this in mind, the finding of different frequencies of MDD in the groups was not anticipated, as the MDD genetics did not achieve any form of grouping, and the overlap of MDD and SCZ genetics is modest. However, MDD is highly heterogeneous, therefore it would be of interest to further explore whether there are any differences between the MDD cases in each group, specifically whether any of the groups corresponds to the recently proposed atypical depression subtype[9,10].

Genetic correlation analyses have begun to explore the common biology and causal relationships between psychiatric and cardiometabolic diseases[1,3,10], however these methods assume that the entire genome influences both sets of traits. The small to moderate correlations could suggest that it is only a portion of the genome that has common effects. In contrast, the current study focuses on only the parts of the genome that have been implicated
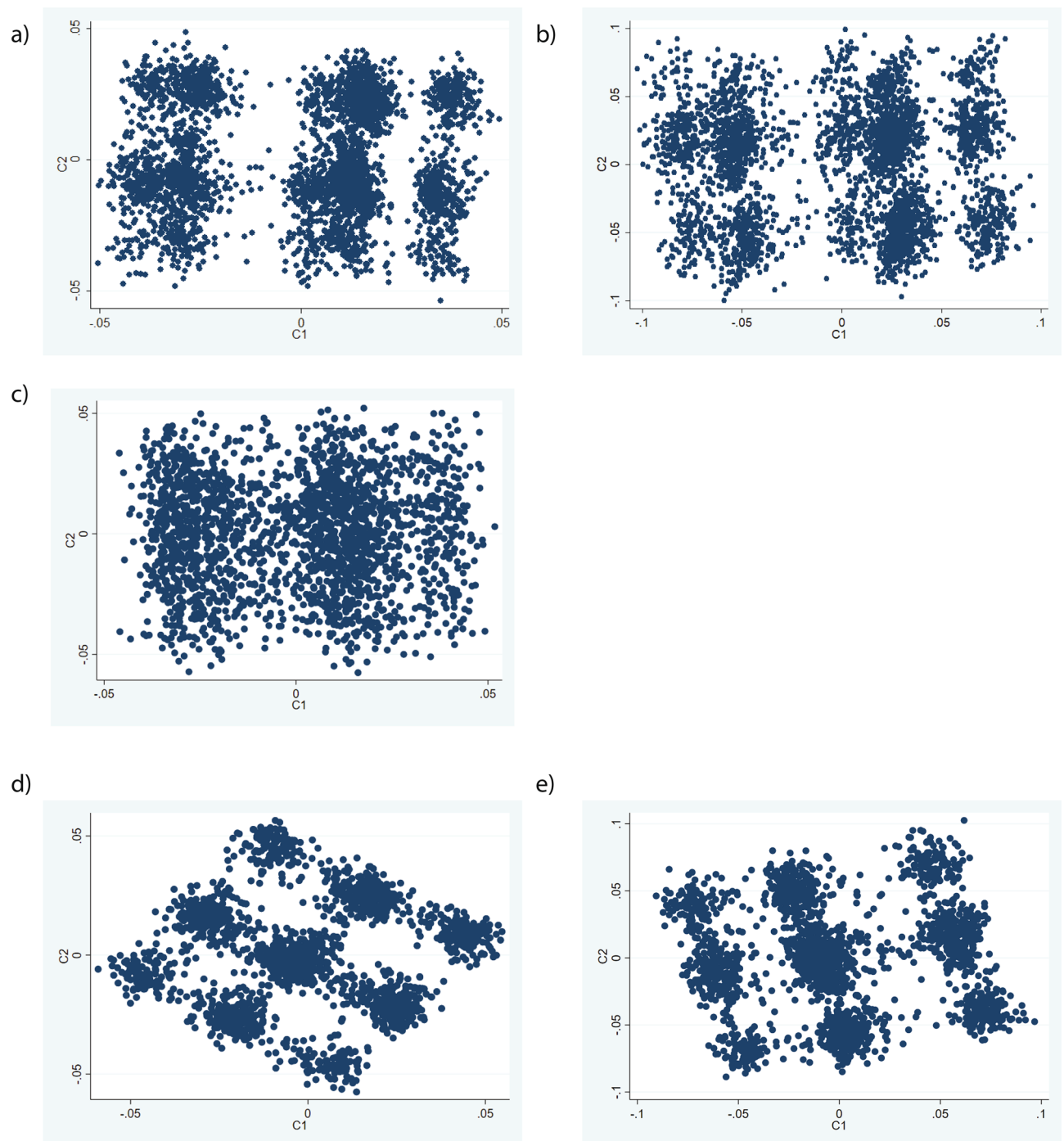
Figure 3. Sensitivity testing in UKB1. For comparison, IMPROVE MDS analysis using (**a**) MAF > 1% and (**b**) MAF > 10%. MDS analysis in UKB1 using (**c**) the same post-filtering SNPs as for IMPROVE, (**d**) the same pre-filtering SNPs with MAF > 1% in UKB1 and (**e**) the same pre-filtering SNPs with MAF > 10% in UKB1. Each data point is an individual therefore the individuals who are closer together are more genetically similar.

in both psychiatric and CMD. Whilst this study does not bring us any closer to understanding the mechanisms underlying the common pathological mechanisms, it does suggest that exploration of the SCZ-CMD loci could have clinical utility, irrespective of mechanistic understanding.

One limitation is that these analyses were conducted in individuals of European ancestry and as SNPs were filtered by MAF and linkage disequilibrium, it is not possible to generalise them to other populations. Indeed, to apply current information from European ancestry individuals to additional ancestry groups has the potential to be misleading and is certainly incomplete. Whilst there is a recognised need[11] and growing efforts around the world to explore genetics of disease in non-European ancestry individuals, it will take time to gain full insight into the genetic architecture of diseases in these ancestry groups.

Figure 4. Comparison of the three clusters identified in (**a**) UKB1 and (**b**) UKB2 (lower panel). Each data point is an individual therefore the individuals who are closer together are more genetically similar.

	UKB1 (N=2,202)					UKB2 (N=20,181)				
Group	1	2	3		P**	1	2	3		P**
N	443 (20.1)	1054 (47.9)	622 (28.2)	83 (3.8)		5972 (29.6)	9926 (45.7)	3558 (17.6)	1425 (7.1)	
Male (%)	205 (46.3)	499 (47.3)	300 (48.2)	38 (45.8)	0.788	2837 (47.5)	4459 (48.3)	1745 (49.0)	718 (50.4)	0.323
Baseline										
Age (years)	55.6 (7.7)	56.0 (7.4)	55.4 (7.7)	54.6 (7.9)	0.121	55.1 (7.4)	55.1 (7.5)	55.4 (7.5)	55.2 (7.5)	0.172
Weight (kg)	76.0 (14.2)	75.7 (14.2)	76.6 (15.2)	76.9 (14.9)	0.805	76.7 (15.0)	77.0 (14.7)	77.0 (14.6)	77.6 (15)	0.352
Waist (cm)	87 (13)	87 (12)	88 (13)	88 (12)	0.883	88 (13)	88 (13)	88 (13)	89 (13)	0.418
Hip (cm)	102 (8)	101 (8)	102 (8)	102 (8)	0.581	102 (8)	102 (8)	102 (8)	102 (8)	0.681
WHR	0.86 (0.09)	0.86 (0.09)	0.86 (0.09)	0.86 (0.09)	0.974	0.86 (0.09)	0.86 (0.09)	0.86 (0.09)	0.86 (0.09)	0.271
BMI (kg/m^2)	26.5 (3.9)	26.3 (3.9)	26.4 (4.0)	26.4 (3.7)	0.492	26.6 (4.4)	26.6 (4.2)	26.6 (4.1)	26.8 (4.2)	0.506
SBP (mmHg)	135 (17)	136 (18)	135 (18)	133 (17)	0.822	135 (17)	136 (18)	136 (18)	136 (18)	**0.002**
DBP (mmHg)	81 (10)	81 (10)	81 (10)	80 (11)	0.822	81 (10)	82 (10)	82 (10)	82 (10)	0.060
SBP* (mmHg)	138 (20)	138 (19)	138 (20)	134 (17)	0.774	137 (19)	138 (20)	138 (19)	138 (20)	**0.016**
DBP* (mmHg)	83 (11)	83 (11)	83 (11)	81 (11)	0.550	83 (11)	83 (11)	83 (11)	83 (11)	0.164
T2D	14 (3.2)	18 (1.7)	13 (2.1)	1 (1.2)	0.208	154 (2.6)	192 (2.1)	67 (1.9)	32 (2.3)	**0.044**
HTN	197 (44.9)	457 (44.2)	285 (46.3)	35 (44.3)	0.931	2491 (43.6)	4097 (46.3)	1534 (44.6)	643 (46.6)	**0.005**
HTN medication	75 (17.1)	158 (15.1)	91 (14.7)	8 (9.8)	0.395	817 (13.7)	1273 (13.9)	515 (14.5)	215 (15.2)	0.530
Lipid-lowering medication	43 (23.5)	80 (17.8)	45 (17.0)	4 (11.1)	0.174	509 (19.7)	741 (18.3)	292 (18.4)	135 (20.1)	0.344
ISH	4 (0.9)	10 (1.0)	6 (1.0)	1 (1.2)	0.361	81 (1.4)	118 (1.3)	47 (1.3)	13 (0.9)	0.920
Current smoking	29 (10.1)	63 (9.3)	45 (9.3)	7 (8.4)	0.900	375 (9.4)	586 (9.5)	193 (8.03)	80 (5.6)	0.092
Former smoking	185 (41.9)	436 (41.4)	252 (36.4)	32 (38.6)	0.072	2341 (39.3)	3622 (39.3)	1340 (37.8)	587 (41.2)	0.221
Follow-up										
Age (years)	61.7 (7.6)	62.0 (7.3)	61.5 (7.6)	60.9 (7.6)	0.069	61.2 (7.4)	63.2 (7.5)	63.4 (7.5)	63.2 (7.5)	0.190
Weight (kg)	75.8 (14.3)	75.3 (14.3)	75.8 (15.4)	77.4 (14.5)	0.770	76.1 (15.2)	76.4 (14.9)	76.3 (14.8)	76.9 (15.1)	0.225
Waist (cm)	86 (12)	86 (12)	86 (12)	88 (11)	0.943	88 (13)	88 (12)	88 (12)	89 (12)	0.474
Hip (cm)	101 (8)	100 (8)	100 (8)	101 (8)	0.402	101 (9)	101 (9)	101 (8)	101 (9)	0.620
WHR	0.85 (0.08)	0.86 (0.08)	0.86 (0.08)	0.87 (0.08)	0.768	0.87 (0.09)	0.87 (0.09)	0.87 (0.09)	0.87 (0.08)	0.797
BMI (kg/m^2)	26.6 (4.0)	26.3 (4.0)	26.3 (4.0)	26.8 (3.7)	0.457	26.5 (4.5)	26.6 (4.4)	26.5 (4.2)	26.7 (4.3)	0.511
SBP (mmHg)	137 (17)	138 (18)	137 (18)	134 (16)	0.332	137 (18)	138 (18)	137 (18)	137 (18)	0.385
DBP (mmHg)	79 (10)	79 (10)	78 (10)	78 (9)	0.381	78 (10)	79 (19)	78 (10)	78 (10)	0.171
SBP* (mmHg)	140 (20)	141 (20)	140 (21)	136 (17)	0.428	141 (20)	141 (20)	141 (20)	141 (20)	0.466
DBP* (mmHg)	81 (11)	81 (11)	81 (12)	80 (10)	0.370	81 (11)	81 (11)	81 (11)	81 (11)	0.316
T2D	22 (5.0)	32 (3.0)	23 (3.7)	2 (2.4)	0.186	267 (4.5)	387 (4.2)	149 (4.2)	64 (4.5)	0.682
HTN	225 (51.6)	556 (53.3)	304 (49.4)	36 (43.4)	0.127	3647 (61.6)	5727 (62.6)	2151 (61.0)	889 (62.8)	0.192
HTN medication	95 (21.5)	222 (21.2)	136 (21.9)	13 (15.7)	0.978	1328 (22.3)	2111 (23.0)	805 (22.7)	324 (22.9)	0.649
Lipid-lowering medication	94 (25.2)	192 (21.9)	107 (20.9)	15 (21.2)	0.341	1374 (26.8)	2056 (26.0)	795 (26.1)	332 (27.2)	0.608
ISH	7 (1.6)	24 (2.3)	8 (1.3)	4 (4.9)	0.491	112 (1.9)	152 (1.7)	53 (1.5)	16 (1.1)	0.334
IMTmean (mm)	0.673 (0.114)	0.674 (0.118)	0.670 (0.122)	0.663 (0.140)	0.475	0.680 (0.123)	0.682 (0.126)	0.686 (0.127)	0.686 (0.127)	0.205
IMTmax (mm)	0.884 (0.172)	0.891 (0.183)	0.886 (0.182)	0.880 (0.217)	0.817	0.911 (0.201)	0.914 (0.208)	0.921 (0.212)	0.916 (0.204)	0.242
Current smoking	17 (6.1)	45 (6.7)	40 (8.2)	2 (2.4)	0.456	221 (5.6)	326 (5.4)	108 (4.6)	53 (3.7)	0.183
Former smoking	177 (40.5)	412 (39.7)	247 (35.8)	26 (31.3)	0.179	2231 (37.7)	3439 (37.6)	1279 (36.4)	559 (39.2)	0.346
MHQ										
Nmax (% of group)	319 (72.0)	739 (70.1)	462 (74.3)	62 (74.7)	0.289	4285 (71.7)	6636 (66.9)	2549 (71.6)	1008 (70.7)	0.803
bd	3 (0.9)	17 (2.3)	5 (1.1)	3 (4.8)	0.108	55 (1.3)	93 (1.4)	31 (1.2)	17 (1.7)	0.748
mdd	89 (32.4)	178 (28.3)	133 (32.5)	10 (17.9)	0.357	1110 (30.7)	1593 (28.0)	572 (26.6)	237 (28.2)	**0.002**

Table 3. Demographic characteristics of the UKB1 and UKB2 participants, by cluster. Highlighted in bold are the significant (p < 0.05) differences between groups. Where: *, adjusted to provide estimates of treatment-naïve levels; na, not available; Statistical analyses compared levels or frequencies across groups 1, 2 and 3. Ungrouped (.) were omitted from the analyses). **P for Pearson's chi square for categorical variables and Kruskal–Wallis for continuous variables.

Another limitation is that the CardioMetabo and Immuno chips do not include all loci implicated in cardio-metabolic disorders. Since these chips were described (2012 and 2011 respectively), many more loci involved in many more processes have been identified. However, as more and more samples are available for GWAS analyses, loci are being identified with smaller and smaller effect sizes. Therefore whilst not all possible information is captured by using the CardioMetabo and Immuno chips, the loci with the largest effects are represented.
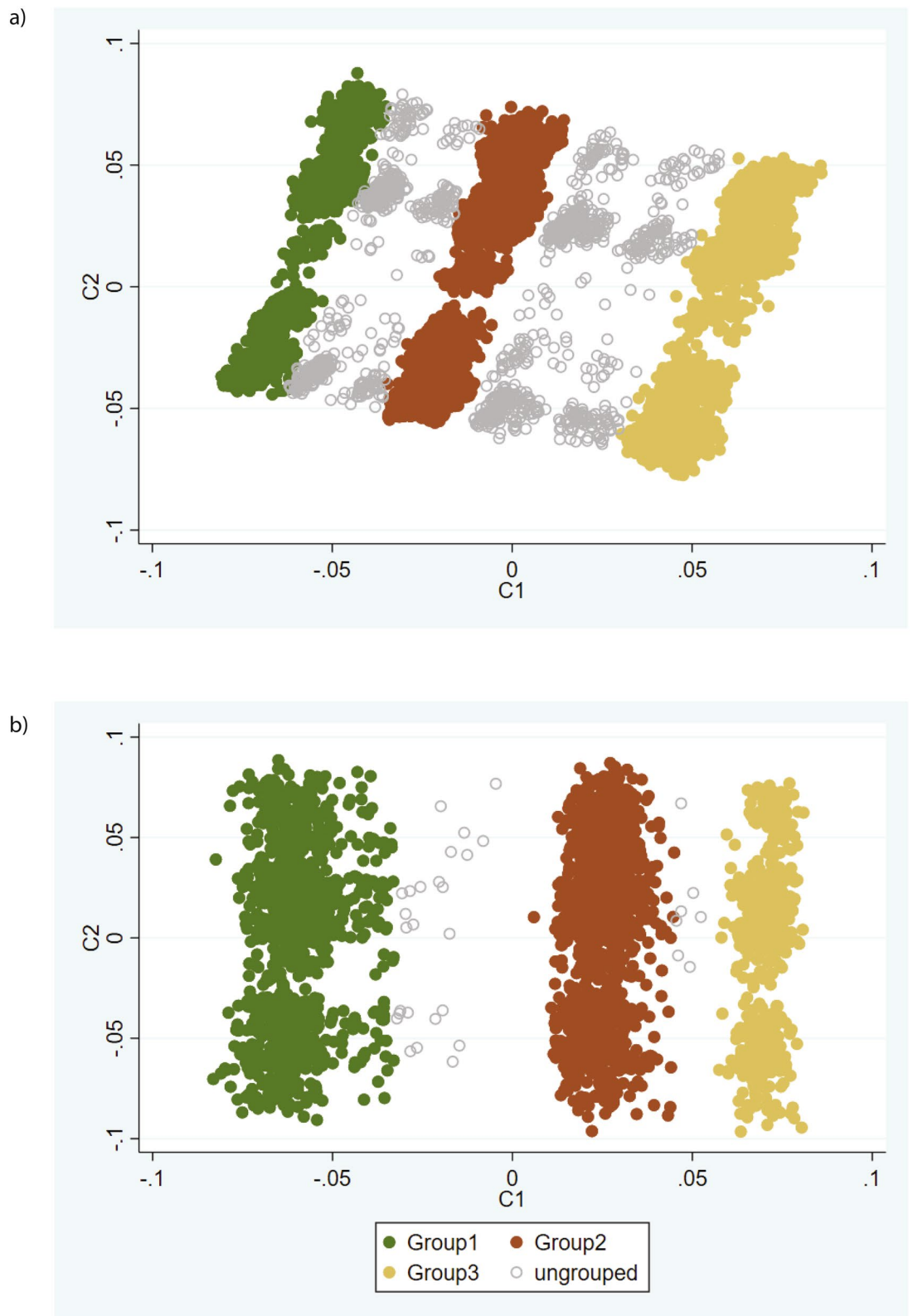
Figure 5. Comparison of the three clusters identified in UKB2 in individuals (**a**) without and (**b**) with mental illness. Each data point is an individual therefore the individuals who are closer together are more genetically similar.

In conclusion, this study provides proof of concept that common biology underlying mental and physical illness is probable and can distinguish subsets of individuals with differing metabolic profiles, even if full understanding of mechanisms is lacking. Given that large-scale genotyping is not available to healthcare providers and the differences between groups are subtle, there is currently limited potential for translation of this into clinical practice. Further investigation with longitudinal datasets, particularly in high CVD risk populations, would define whether or not there is potential for clinical value in this method.

Methods

Cohorts: phenotyping and genotyping.

The IMPROVE study has been described previously[12,13]. In short, 3700 individuals aged between 54–79 years with high CVD risk profiles (the presence of at least 3 classical CVD risk factors, including family history of CVD, type 2 diabetes, hypertension, hyperlipidaemia and smoking) were recruited from seven centres in Finland, Sweden, the Netherlands, France and Italy. At baseline, individuals completed lifestyle and medical questionnaires and anthropometric measures taken. Blood was sampled for DNA extraction and clinical biochemistry and stored for further biochemical analyses. Detailed ultra-sound examination of the carotid intima-media thickness (cIMT) was conducted at baseline, 15 months and 30 months. Linear regression using all data points was used to calculate progression of cIMT. Mental illness was not assessed; however it is believed that if there is mental illness in this cohort it is likely to be subclinical. All participants provided written informed consent and the study was conducted in accordance with the Helsinki Declaration. Ethical approval was granted by the Regional Ethics Review Boards at Karolinska Institutet, Stockholm Sweden, the Groupe Hôpitalier Pitie-Salpetriere, Paris, France, the Comitato Etico delle Aziende Sanitarie della regione Umbria, Perugia, Italy, the Ospedale Niguarda Ca´Granda, Milano, Italy, the University Hospital Groningen, Groningen, the Netherlands, the Hospital District of Northern Savo, Kuopio, Finland and the University of Eastern Finland, Kuopio, Finland.

The IMPROVE study was genotyped on the Illumina Cardio-Metabo[14] and Immuno chips[15], therefore cardiometabolic disorders (including immune and inflammatory components) were well represented. Standard quality control procedures were conducted, namely exclusion of SNPs for low call rate (< 95%) and deviation from Hardy–Weinberg Equilibrium ($p < 1 \times 10^{-6}$) and exclusion of samples for low call rate (< 95%), sex-mismatch, cryptic relatedness. Quality control was conducted on each chip separately, followed by a further round of quality control on the combined chip.

The UK Biobank (UKB) has been described previously[16,17]. Approximately 500,000 volunteers aged 39–73 years were recruited from 22 centres across the UK. At baseline, detailed questionnaires on sociodemographic factors, lifestyle factors and medical history were completed by all individuals. Measurements of anthropometric variables were recorded and blood samples were taken for DNA extraction. Subsequently (4–8 years after baseline), subsets of participants were invited for follow-up measurements and extensive imaging. All participants provided written informed consent and ethical approval was granted by the NHS national Research Ethics Service. This work was conducted under projects #6533 (Smith) and #1755 (Pell).

Ultrasound measurement of cIMT was conducted in a pilot phase of ~ 2500 individuals (henceforth denoted as UKB1) followed by a subsequent phase including ~ 22,000 individuals (denoted UKB2) using the same recruitment and measurement protocol. cIMT measurements were generally consistent with the measurements available in IMPROVE. A mental health/thoughts and feelings questionnaire was also completed by a subset of participants, which enabled estimation of life history of MDD and BD. For both UKB1 and UKB2, 73% of participants completed the mental health questionnaire.

Genome-wide genotyping was conducted and standard quality control procedures were applied by the UK Biobank team[18]. Imputation was conducted using the Haplotype reference consortium and 1000 Genomes with standard pre- and post- imputation quality controls being applied by the UK Biobank team (further information is provided in[18]).

Multi-dimensional scaling (MDS) to identify clusters.

Genome-wide genetic loci reported to be associated with SCZ[19], MDD[20] and BD[21] were identified. SNPs within these (SCZ, MDD or BD) loci which were present on the CardioMetabo and Immuno chips were selected[14,15] (denoted SCZ-CM SNPs, MDD-CM SNPs or BD-CM SNPs, respectively). SNPs with MAF > 1% were included (Supplementary Table 3). A schematic diagram of the analyses steps is provided in Fig. 1.

In IMPROVE, each set of SNPs (SCZ-CM SNPs, MDD-CM SNPs or BD-CM SNPs) were pruned by pairwise LD (parameters 50, 5, 0.1) using PLINK[22]. Individuals with > 1% missing genetic data were excluded prior to clustering.

Clustering was performed using multi-dimensional scaling, implemented in PLINK, using default settings. Multidimensional scaling essentially measures similarity between individuals, in this case using the patterns of genetic variation as the assessment criteria[23,24]. Individuals with similar genetic sequences are deemed more similar to each other than those with less similar genetic sequences. Clustering was also conducted using tSNE and PCA (Supplementary Methods).

Subsequently in UKB1, SCZ-cardiometabolic SNPs only were used and individuals with > 1% missing genetic data were excluded prior to clustering. MDS analyses was conducted using either exactly the same SNPs as were used in IMPROVE (ie SCZ-CM SNPs after filtering for MAF and LD in IMPROVE) or SCZ-CM SNPs with filtering for MAF and LD being done in UKB1.

Finally, in UKB2, Individuals with > 1% missing genetic data were excluded prior to clustering. MDS analysis was conducted on SCZ-CM SNPs with filtering for MAF and LD in UKB2, or on all SCZ SNPs after MAF filtering and pruning in UKB2.

The first two MDS components (C1 and C2) were plotted for visual assessment.

Choosing a negative control experiment is not straight forward, as current evidence suggests that most genetic variants are highly pleiotropic and that complex traits overlap with each other to a large degree. Despite some overlap with CMD or SCZ-related traits, SNPs in genetic loci associated with eye colour were used as a negative control experiment. The analysis was conducted in UKB2 with MAF > 10% filtering and pruning as described above.

Statistical analyses. In IMPROVE, Spearmans rank correlation coefficients were used to assess the relationship between the MDS components and latitude. For IMPROVE, UKB1 and UKB2, Differences between groups were assessed by Pearsons chi squared test for categorical values and Kruskal–Wallis test for continuous variables. All statistical analyses were conducted in Stata (version 11.0). The threshold for significance was set at $p < 0.05$. No adjustment for multiple testing was applied, because these analyses are exploratory rather than definitive and secondly because most of the cardiometabolic phenotypes tested are interrelated and thus are not independent tests.

Data availability

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

References

1. So, H. C., Chau, K. L., Ao, F. K., Mo, C. H. & Sham, P. C. Exploring shared genetic bases and causal relationships of schizophrenia and bipolar disorder with 28 cardiovascular and metabolic traits. *Psychol. Med.* **49**, 1286–1298. https://doi.org/10.1017/S0033291718001812 (2019).
2. De Hert, M. *et al.* Physical illness in patients with severe mental disorders. I. Prevalence, impact of medications and disparities in health care. *World Psychiatry* **10**, 52–77. https://doi.org/10.1002/j.2051-5545.2011.tb00014.x (2011).
3. Amare, A. T., Schubert, K. O., Klingler-Hoffmann, M., Cohen-Woods, S. & Baune, B. T. The genetic overlap between mood disorders and cardiometabolic diseases: a systematic review of genome wide and candidate gene studies. *Transl. Psychiatry* **7**, e1007. https://doi.org/10.1038/tp.2016.261 (2017).
4. Tang, B., Yuan, S., Xiong, Y., He, Q. & Larsson, S. C. Major depressive disorder and cardiometabolic diseases: a bidirectional Mendelian randomisation study. *Diabetologia* **63**, 1305–1311. https://doi.org/10.1007/s00125-020-05131-6 (2020).
5. Mulugeta, A., Zhou, A., King, C. & Hypponen, E. Association between major depressive disorder and multiple disease outcomes: a phenome-wide Mendelian randomisation study in the UK Biobank. *Mol. Psychiatry* **25**, 1469–1476. https://doi.org/10.1038/s41380-019-0486-1 (2020).
6. Adhikari, K. *et al.* A GWAS in Latin Americans highlights the convergent evolution of lighter skin pigmentation in Eurasia. *Nat. Commun.* **10**, 358. https://doi.org/10.1038/s41467-018-08147-0 (2019).
7. Jonnalagadda, M. *et al.* A genome-wide association study of skin and iris pigmentation among individuals of South Asian ancestry. *Genome Biol. Evol.* **11**, 1066–1076. https://doi.org/10.1093/gbe/evz057 (2019).
8. Fry, A. *et al.* Comparison of sociodemographic and health-related characteristics of UK biobank participants with those of the general population. *Am. J. Epidemiol.* **186**, 1026–1034. https://doi.org/10.1093/aje/kwx246 (2017).
9. Milaneschi, Y., Lamers, F., Bot, M., Drent, M. L. & Penninx, B. W. Leptin dysregulation is specifically associated with major depression with atypical features: evidence for a mechanism connecting obesity and depression. *Biol. Psychiatry* **81**, 807–814. https://doi.org/10.1016/j.biopsych.2015.10.023 (2017).
10. Milaneschi, Y. *et al.* Genetic association of major depression with atypical features and obesity-related immunometabolic dysregulations. *JAMA Psychiatry* **74**, 1214–1225. https://doi.org/10.1001/jamapsychiatry.2017.3016 (2017).
11. Peterson, R. E. *et al.* Genome-wide association studies in ancestrally diverse populations: opportunities, methods, pitfalls, and recommendations. *Cell* **179**, 589–603. https://doi.org/10.1016/j.cell.2019.08.051 (2019).
12. Baldassarre, D. *et al.* Measurements of carotid intima-media thickness and of interadventitia common carotid diameter improve prediction of cardiovascular events: results of the IMPROVE (Carotid Intima Media Thickness [IMT] and IMT-progression as predictors of vascular events in a high risk European Population) study. *J. Am. Coll. Cardiol.* **60**, 1489–1499. https://doi.org/10.1016/j.jacc.2012.06.034 (2012).
13. Baldassarre, D. *et al.* Cross-sectional analysis of baseline data to identify the major determinants of carotid intima-media thickness in a European population: the IMPROVE study. *Eur. Heart J.* **31**, 614–622. https://doi.org/10.1093/eurheartj/ehp496 (2010).
14. Voight, B. F. *et al.* The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS Genet.* **8**, e1002793. https://doi.org/10.1371/journal.pgen.1002793 (2012).
15. Trynka, G. *et al.* Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nat. Genet.* **43**, 1193–1201. https://doi.org/10.1038/ng.998 (2011).
16. Matthews, P. M. & Sudlow, C. The UK Biobank. *Brain* **138**, 3463–3465. https://doi.org/10.1093/brain/awv335 (2015).
17. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779. https://doi.org/10.1371/journal.pmed.1001779 (2015).
18. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209. https://doi.org/10.1038/s41586-018-0579-z (2018).
19. Schizophrenia Working Group of the Psychiatric Genomics, C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427, https://doi.org/10.1038/nature13595 (2014).
20. Wray, N. R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681. https://doi.org/10.1038/s41588-018-0090-3 (2018).
21. Stahl, E. A. *et al.* Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat. Genet.* **51**, 793–803. https://doi.org/10.1038/s41588-019-0397-8 (2019).
22. 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).
23. Zhang, Z. & Takane, Y. Multidimensional Scaling. *International Encyclopedia of Education (Third Edition)*, 304–311 (2010).
24. Jaworska, N. & Chupetlovska-Anastasova, A. A review of multidimensional scaling (MDS) and its utility in various psychological domains. *Tutorials Quantitative Methods Psychol.* **5**, 1–10 (2009).

Acknowledgements

Author contributions

Study conception, design, data analysis (R.J.S., B.S.) or acquisition (D.B., U.dF., A.H., S.E.H., E.T., T.P.T.), interpretation (R.J.S., K.J.A.J., P.E., B.G., S.E.H., D.M.L., L.M.L., B.S., D.J.S.), manuscript preparation or revision (R.J.S., K.J.A.J., M.E.S.B., D.B., B.C., P.E., U.dF., A.F., B.G., P.G., N.G., A.H., S.E.H., S.K., D.M.L., L.M.L., M.P., J.P.P., K.S., B.S., A.J.S., E.T., T.P.T., F.V., J.W., D.J.S.).

Competing interests

The authors declare no competing interests.

Additional information

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