medRxiv preprint doi: https://doi.org/10.1101/2020.06.23.20138271; this version posted June 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

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 Baldasarre[6,7], Breda Cullen [1], Per Eriksson[3], Ulf DeFaire[8], Amy Ferguson[1,9] , Bruna Gigante[8] on behalf of the IMPROVE study¤, Philippe Giral[10], Nicholas Graham[1], Anders Hamsten[3], Steve E. Humphries[11], Sudhir Kurl[12], Donald M. Lyall[1], Laura M. Lyall[1], Matteo Pirro[13], Jill P. Pell[1], Kai Savonen[14,15], Bengt Sennblad[16], Andries J. Smit[17], Elena Tremoli[6], Tomi-Pekka Tomainen[18], Fabrizio Veglia[6], Joey Ward[1], Daniel J. Smith[1]

[1]Institute of Health and Wellbeing, University of Glasgow, Glasgow, UK
[2]Health Data Research, UK
[3]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, Scotland, UK
[5]School of Life Sciences, College of Medical, Veterinary & Life Sciences, University of Glasgow, Scotland, UK
[6]Centro Cardiologico Monzino, IRCCS, Milan, Italy
[7]Department of Medical Biotechnology and Translational Medicine, Università degli Studi di Milano, Milan, Italy
[8]Cardiovascular Medicine Unit, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden.
[9]Usher Institute, University of Edinburgh, UK
[10]Assistance Publique - Hopitaux de Paris; Service Endocrinologie-Metabolisme, Groupe Hôpitalier Pitie-Salpetriere, Unités de Prévention Cardiovasculaire, 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 Cell and Molecular Biology, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Uppsala University, Uppsala, Sweden
[17]Department of Medicine, University Medical Center Groningen and University of Groningen, Groningen, The Netherlands
[18]Public Health and Clinical Nutrition, Department of Medicine, University of Eastern Finland, Kupiou, Fnland

¤additional contributors to the IMPROVE Study Group are listed in the Supplementary Data.

Brief title: Metabolic differences between genetically-defined subsets of individuals

*** Corresponding author:**

Dr Rona J. Strawbridge
Address: Institute of Health and Wellbeing, University of Glasgow, Room 111, Public Health, 1 Lilybank Gardens, Glasgow, G12 8RZ, UK.
Telephone: +44 141 330 3482 Email: rona.strawbridge@glasgow.ac.uk

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.23.20138271; this version posted June 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Abstract

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 researcher. We explored whether genetic variation could identify individuals with different metabolic profiles. Loci previously associated with schizophrenia, bipolar disorder and major depressive disorder were identified from literature and those overlapping loci genotyped on the Illumina CardioMetabo and Immuno chips (representing cardiometabolic processes and diseases) were selected. In the IMPROVE study (high cardiovascular risk) and UK Biobank (general population) multidimensional scaling was applied to genetic variants implicated in both mental and cardiometabolic illness. Visual inspection of the resulting plots used to identify distinct clusters. Differences between clusters were assessed using chi-squared and Kruskall-Wallis tests. In IMPROVE, genetic loci associated with both cardiometabolic disease and schizophrenia (but not bipolar or major depressive disorders) identified three groups of individuals with distinct metabolic profiles. The grouping was replicated in UK Biobank, albeit with less distinction between metabolic profiles. This study provides proof of concept that common biology underlying mental and physical illness can identify subsets of individuals with different cardiometabolic profiles.

Key words

Genetics, Schizophrenia, type 2 diabetes, blood pressure, hypertension, cardiovascular disease, obesity, clustering

Introduction

Individuals with serious mental illness (such as schizophrenia (SCZ), major depressive disorder (MDD) and bipolar disorder (BD)) have a reduced life expectancy (10-15years for BD, 15-20years 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 ~2-fold 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.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.23.20138271; this version posted June 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

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 (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) (Figure 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 (SFigure 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 (Figure 2b), whereas MAF >10% showed similar groups to MAF >1. Assignment to groups was consistent using MAF >1% and MAF >10% (STable 1). The three groups appear to have modest differences in cardiometabolic profiles (Table 2): Group 3 had a significantly lower frequency of hypertension (P=0.004) and lower fastest progression of cIMT (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.

No separation into groups was observed when using MDD-CM SNPs, irrespective of the MAF filter used (Figure 2d-f). For BD-CM SNPs (Figure 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-CM 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 (Figure 3a and b), the post-filtering SNPs from IMPROVE were used (Figure 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-CM SNPs were filtered for MAF and LD in UKB1. As noted in Figure 1, the majority of SNPs included in the two approaches were the same. Unsurprisingly, the SNPs that differed were mainly those with MAF<10%. Using SCZ-CM and conducting MAF and LD filtering in UKB1, nine groups are evident when using SNPs with MAF >1% (Figure 3d), whereas three groups are observed when using SNPs with MAF >10% (Figure 3e). When comparing the metabolic profiles of

medRxiv preprint doi: https://doi.org/10.1101/2020.06.23.20138271; this version posted June 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

the 3 groups, no significant differences were seen (Figure 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 (Figure 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.

Significant (but clinically modest) differences were observed in baseline SBP, SDP adjusted for blood-pressure medication, and frequency of hypertension and T2D (Table3). 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 (Figure 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 (Figure 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 (STable 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 clustering, 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 SFigure 2, SNPs in loci associated with SCZ do not separate individuals into groups, although nine closely grouped clusters are visible when plotting MDS components 3 and 4. This confirms that it is the overlap of SCZ and CMD loci, 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 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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.23.20138271; this version posted June 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

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 [6]. 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.

That the metabolic profiles of the groups did not completely agree between the 3 cohorts was disappointing, 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 would be of clinical importance.

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 [7, 8].

Genetic correlation analyses have begun to explore the common biology and causal relationships between psychiatric and cardiometabolic diseases [1, 3, 8], 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 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. Another limitation is that the CardioMetabo and Immuno chips do not include all loci implicated in cardiometabolic 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.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.23.20138271; this version posted June 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

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, 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 [9, 10]. 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 borad for each recruitment centre.

The IMPROVE study was genotyped on the Illumina Cardio-Metabo [11] and Immuno chips [12], 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 ($p< 1x10^{-6}$) 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 [13, 14]. 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 [15]. Imputation was conducted using the Haplotype reference consortium and

medRxiv preprint doi: https://doi.org/10.1101/2020.06.23.20138271; this version posted June 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

1000 Genomes with standard pre- and post- imputation quality controls being applied by the UK Biobank team (further information is provided in [15-17]).

Multi-dimensional scaling (MDS) to identify clusters

Genome-wide genetic loci reported to be associated with SCZ [18], MDD [19] and BD [20] were identified. SNPs within these (SCZ, MDD or BD) loci which were present on the CardioMetabo and Immuno chips were selected [11, 12] (denoted SCZ-CM SNPs, MDD-CM SNPs or BD-CM SNPs, respectively). SNPs with MAF >1% were included (Stable 3). A schematic diagram of the analyses steps is provided in Figure 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 [21]. 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 [22, 23]. Individuals with similar genetic sequences are deemed more similar to each other than those with less similar genetic sequences.

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. Ideally a negative control experiment would be included, however as current evidence suggests that most genetic variants are highly pleiotropic and that complex traits overlap with each other to a large degree. Therefore choosing a "negative control" trait is not straightforward.

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 on reasonable request.

References

1. So, H.C., et al., *Exploring shared genetic bases and causal relationships of schizophrenia and bipolar disorder with 28 cardiovascular and metabolic traits.* Psychol Med, 2019. **49**(8): p. 1286-1298.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.23.20138271; this version posted June 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

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, 2011. **10**(1): p. 52-77.
3. Amare, A.T., et al., *The genetic overlap between mood disorders and cardiometabolic diseases: a systematic review of genome wide and candidate gene studies.* Transl Psychiatry, 2017. **7**(1): p. e1007.
4. Tang, B., et al., *Major depressive disorder and cardiometabolic diseases: a bidirectional Mendelian randomisation study.* Diabetologia, 2020. **63**(7): p. 1305-1311.
5. Mulugeta, A., et al., *Association between major depressive disorder and multiple disease outcomes: a phenome-wide Mendelian randomisation study in the UK Biobank.* Mol Psychiatry, 2020. **25**(7): p. 1469-1476.
6. Fry, A., et al., *Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population.* Am J Epidemiol, 2017. **186**(9): p. 1026-1034.
7. Milaneschi, Y., et al., *Leptin Dysregulation Is Specifically Associated With Major Depression With Atypical Features: Evidence for a Mechanism Connecting Obesity and Depression.* Biol Psychiatry, 2017. **81**(9): p. 807-814.
8. Milaneschi, Y., et al., *Genetic Association of Major Depression With Atypical Features and Obesity-Related Immunometabolic Dysregulations.* JAMA Psychiatry, 2017. **74**(12): p. 1214-1225.
9. 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, 2012. **60**(16): p. 1489-99.
10. 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, 2010. **31**(5): p. 614-22.
11. Voight, B.F., et al., *The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits.* PLoS Genet, 2012. **8**(8): p. e1002793.
12. Trynka, G., et al., *Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease.* Nat Genet, 2011. **43**(12): p. 1193-201.
13. Matthews, P.M. and C. Sudlow, *The UK Biobank.* Brain, 2015. **138**(Pt 12): p. 3463-5.
14. 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, 2015. **12**(3): p. e1001779.
15. Bycroft, C., et al., *The UK Biobank resource with deep phenotyping and genomic data.* Nature, 2018. **562**(7726): p. 203-209.
16. Biobank, U., *Genotype imputation and genetic association studies of UK Biobank.* 2015.
17. Biobank, U., *Genotyping of 500,000 UK Biobank participants. Description of sample processing workflow and preparation of DNA for genotyping.* 2015.
18. Schizophrenia Working Group of the Psychiatric Genomics, C., *Biological insights from 108 schizophrenia-associated genetic loci.* Nature, 2014. **511**(7510): p. 421-7.
19. Wray, N.R., et al., *Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression.* Nat Genet, 2018. **50**(5): p. 668-681.
20. Stahl, E.A., et al., *Genome-wide association study identifies 30 loci associated with bipolar disorder.* Nat Genet, 2019. **51**(5): p. 793-803.
21. Purcell, S., et al., *PLINK: a tool set for whole-genome association and population-based linkage analyses.* Am J Hum Genet, 2007. **81**(3): p. 559-75.
22. Zhang, Z. and Y. Takane, *Multidimensional Scaling.* International Encyclopedia of Education (Third Edition), 2010: p. Pages 304-311.
23. Jaworska, N. and A. Chupetlovska-Anastasova, *A Review of Multidimensional Scaling (MDS) and its Utility in Various Psychological Domains.* Tutorials in Quantitative Methods for Psychology, 2009. **5**(1): p. 1-10.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.23.20138271; this version posted June 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Acknowledgements

The authors thank all participants and staff of the IMPROVE and UK Biobank studies. This work uses data provided by patients and collected by the NHS as part of their care and support. IMPROVE was supported by the European Commission (Contract number: QLG1-CT-2002-00896), the Swedish Heart-Lung Foundation, the Swedish Research Council (projects 8691 and 09533), the Knut and Alice Wallenberg Foundation, the Foundation for Strategic Research, the Stockholm County Council (project 592229), the Strategic Cardiovascular and Diabetes Programmes of Karolinska Institutet and Stockholm County Council, the European Union Framework Programme 7 (FP7/2007-2013) for the Innovative Medicine Initiative under grant agreement n° IMI/115006 (the SUMMIT consortium), the Academy of Finland (Grant #110413), the British Heart Foundation (RG2008/08, RG2008/014) and the Italian Ministry of Health (Ricerca Corrente). The UK Biobank was established by the Wellcome Trust, Medical Research Council, Department of Health, Scottish Government and Northwest Regional Development Agency. UK Biobank has also had funding from the Welsh Assembly Government and the British Heart Foundation. Data collection was funded by UK Biobank.
RoJS is supported by a UKRI Innovation-HDR-UK Fellowship (MR/S003061/1). LML is supported by the JMAS Sim Fellowship for depression research from the Royal College of Physicians of Edinburgh. AF is supported by an MRC Doctoral Training Programme Studentship at the University of Glasgow (MR/K501335/1). KJAJ is supported by an MRC Doctoral Training Programme Studentship at the Universities of Glasgow and Edinburgh. DJS acknowledges the support of a Lister Prize Fellowship (173096) and MRC Mental Health Data Pathfinder Award (MC_PC_17217).

Author Contributions

Study conception, design, data analysis (RJS) or acquisition (DB, UdF, AH, SHE, TPT), interpretation (RJS, KJAJ, PE, BG, SEH, DML, LML, BS, DJS), manuscript preparation or revision (RJS, KJAJ, MESB, DB, BC, PE, UdF, AF, BG, PG, NG, AH, SEH, SK, DML, LML, MP, JPP, KS, BS, AJS, TPT, FV, JW, DJS).

Additional Information

The authors have no competing interests to declare.

Figure Legends

Figure 1: Schematic of the analysis procedure used to identify clusters.

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.

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.

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.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.23.20138271; this version posted June 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

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.

Tables

Table 1: Descriptive characteristics of the cohorts used in this study

Table 2: Demographic characteristics of the IMPROVE participants, by cluster

Table 3: Demographic characteristics of the UKB1 and UKB2 participants, by cluster

medRxiv preprint doi: https://doi.org/10.1101/2020.06.23.20138271; this version posted June 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Table 1: Demographic characteristics of IMPROVE and UKB participants

		IMPROVE	UKB1	UKB2
	Nmax	3300	2202	20182
	Male (%)	1695 (51.4)	1042 (47.3)	9759 (48.4)
baseline	age	64.2 (5.4)	55.7 (7.6)	55.2 (7.5)
	weight	76.7 (15.1)	76.1 (14.5)	76.9 (14.8)
	waist	94 (13)	88 (12)	88 (13)
	hip	102 (10)	102 (8)	102 (8)
	whr	0.92 (0.09)	0.86 (0.09)	0.86 (0.09)
	bmi	27.3 (4.2)	26.4 (3.9)	26.6 (4.2)
	sbp	142 (18)	136 (18)	136 (18)
	dbp	82 (10)	82 (10)	82 (10)
	sbp*	151 (21)	138 (19)	138 (19)
	dbp*	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_meds	1904 (57.7)	332 (15.2)	2820 (14.0)
	lipid-lowering	1623 (49.2)	172 (18.4)	1677 (18.9)
	ish	0 (0)	21 (1.0)	259 (1.3)
	mm_mean_imt	0.891 (0.199)	na	na
	mm_max_imt	2.037 (0.813)	na	na
	current smoking	498 (15.1)	139 (6.3)	1234 (6.1)
	Former smoker	1216 (36.9)	746 (33.9)	6656 (33.0)
	age	66.7 (5.4)	61.8 (7.5)	63.2 (7.5)
	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)	12414 (62.0)
	htn_meds	na	466 (21.2)	4568 (22.7)
	lipid-lowering	na	408 (22.3)	4557 (26.4)
	ish	119 (3.6)	43 (2.0)	333 (1.7)
	mm_mean_imt	na	0.672 (0.119)	0.682 (0.125)
	mm_max_imt	na	0.888 (0.182)	0.914 (0.297)
	progression of mean_imt	0.0186 (0.032)	na	na
	progression of max_imt	0.0439 (0.163)	na	na
	current smoking	na	99 (4.6)	708 (3.5)
followup	Former smoker	na	742 (34.1)	6800 (33.9)
	Nmax (% of group)	na	1528 (69.4)	10079 (49.9)
	bd	na	28 (1.8)	196 (1.4)
MHQ	mdd	na	410 (29.9)	3512 (28.6)

medRxiv preprint doi: https://doi.org/10.1101/2020.06.23.20138271; this version posted June 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

gad	na	99 (8.9)	1026 (10.2)

Where: *, adjusted to provide estimates of treatment-naïve levels, as per Ehret etl al; na, not available

medRxiv preprint doi: https://doi.org/10.1101/2020.06.23.20138271; this version posted June 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Table2: Demographic characteristics of the IMPROVE participants, by cluster (MAF>10%)

		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/m2)		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
hypertension		982 (80.4)	1321 (81.1)	389 (74.0)	**0.004**
on anti-hypertensive meds		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
smokers		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		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

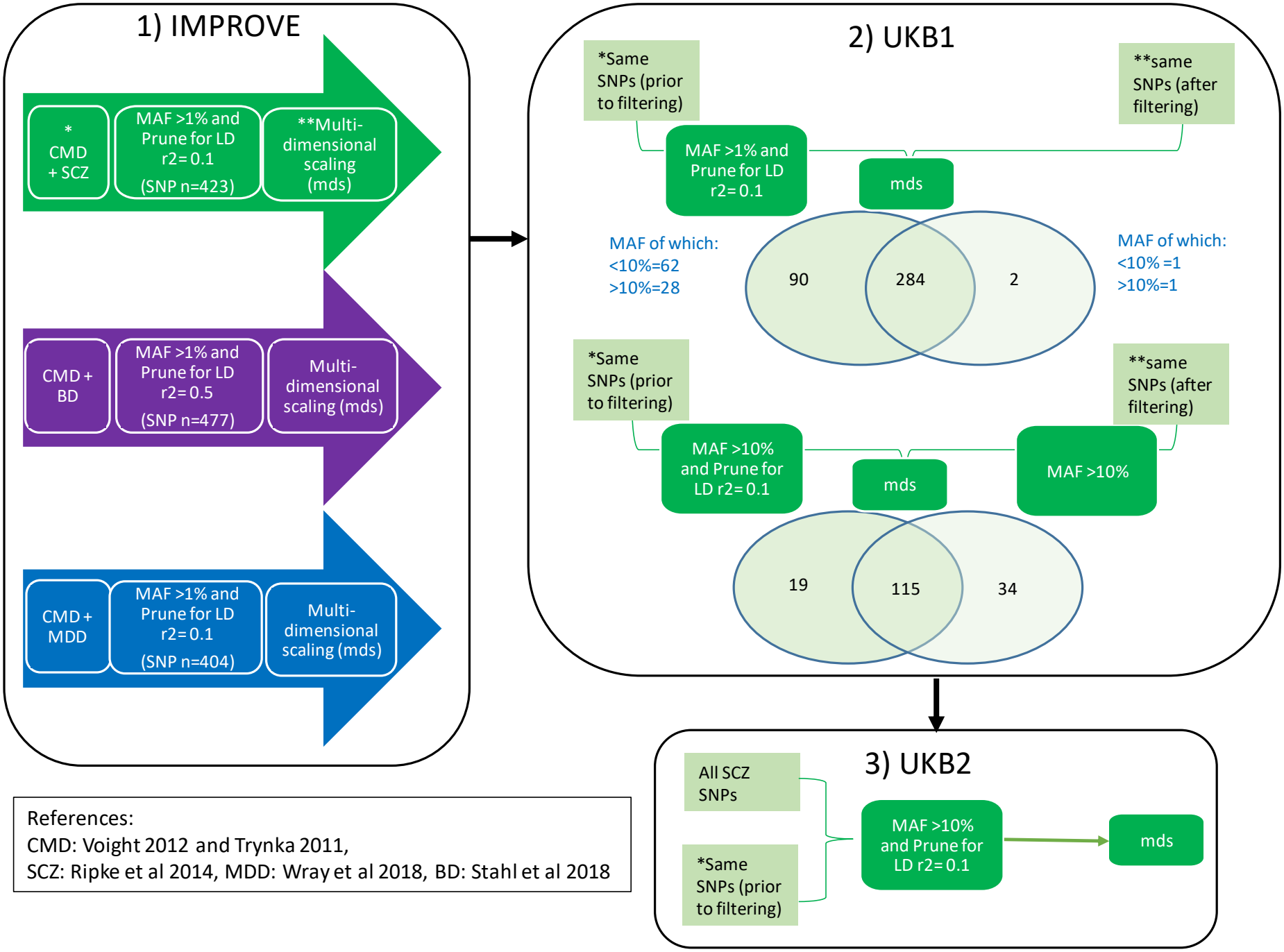
Where: *, adjusted to provide estimates of treatment-naïve levels as per Ehret et al; ** P for Pearsons chi square for categorical variables and Kruskal-Wallis for continuous variables; na, not available

Table3: Demographic characteristics of the UKB1 and UKB2 participants, by cluster

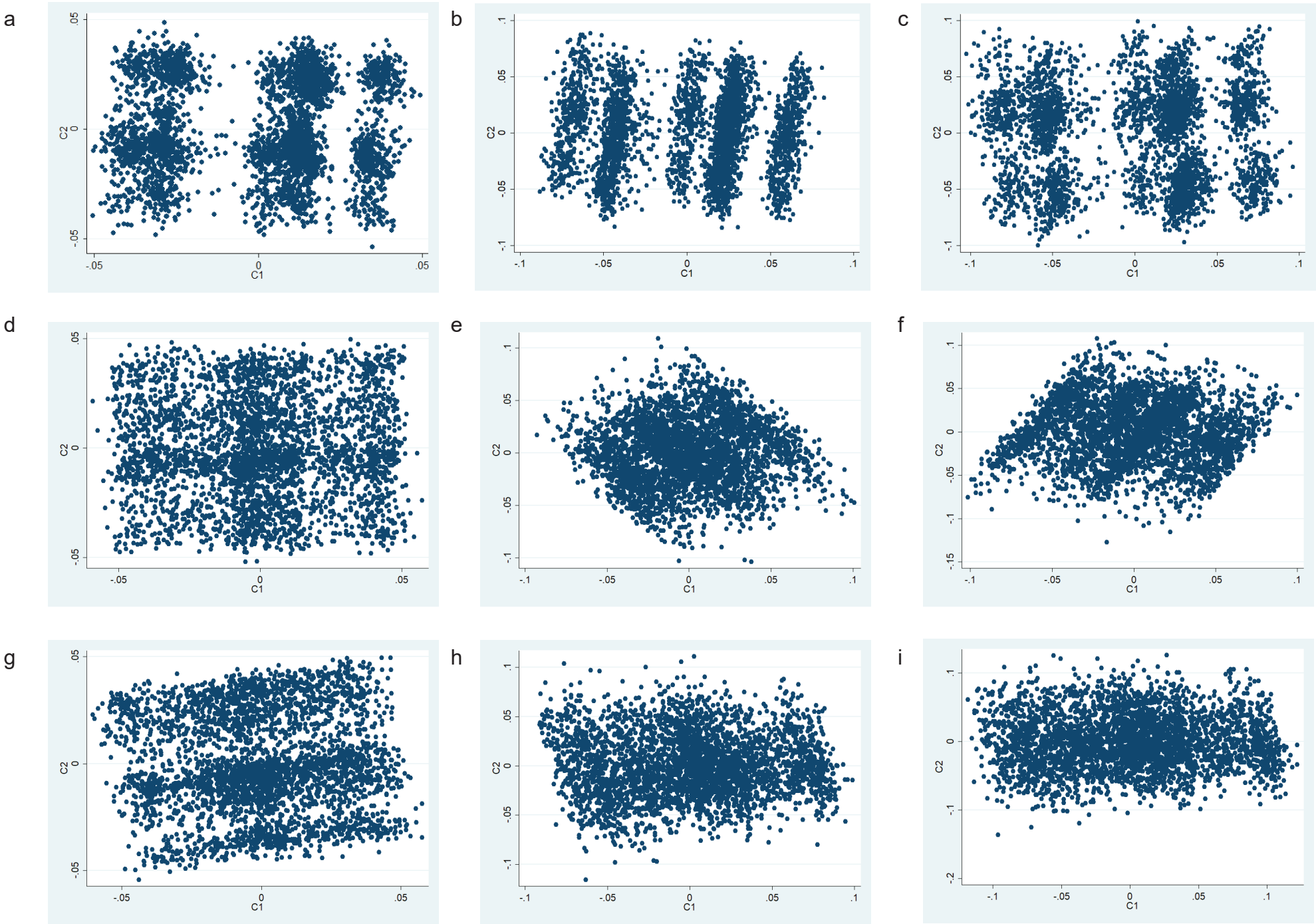
		UKB1 (N=2,202)					UKB2 (N=20181)				
	Group	1	2	3	.	P	1	2	3	.	P
	N	443 (20.1)	1054 (47.9)	622 (28.2)	83 (3.8)		5972 (29.6)	9926 (45.7)	3558 (17.6)	1425 (7.1)	
	Male (%)	205 (46.3)	499 (47.3)	300 (48.2)	38 (45.8)	0.788	2837 (47.5)	4459 (48.3)	1745 (49.0)	718 (50.4)	0.323
baseline	age	55.6 (7.7)	56.0 (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	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	87 (13)	87 (12)	88 (13)	88 (12)	0.883	88 (13)	88 (13)	88 (13)	89 (13)	0.418
	hip	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	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	135 (17)	136 (18)	135 (18)	133 (17)	0.822	135 (17)	136 (18)	136 (18)	136 (18)	**0.002**
	dbp	81 (10)	81 (10)	81 (10)	80 (11)	0.822	81 (10)	82 (10)	82 (10)	82 (10)	0.060
	sbp_adj	138 (20)	138 (19)	138 (20)	134 (17)	0.774	137 (19)	138 (20)	138 (19)	138 (20)	**0.016**
	dbp_adj	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_meds	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	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
	Ever smoker	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
followup	age	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	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	86 (12)	86 (12)	86 (12)	88 (11)	0.943	88 (13)	88 (12)	88 (12)	89 (12)	0.474
	hip	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	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	137 (17)	138 (18)	137 (18)	134 (16)	0.332	137 (18)	138 (18)	137 (18)	137 (18)	0.385
	dbp	79 (10)	79 (10)	78 (10)	78 (9)	0.381	78 (10)	79 (19)	78 (10)	78 (10)	0.171

	sbp_adj	140 (20)	141 (20)	140 (21)	136 (17)	0.428	141 (20)	141 (20)	141 (20)	141 (20)	0.466
	dbp_adj	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_meds	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	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
	mm_mean_imt	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
	mm_max_imt	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
	Ever smoker	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**

Where: *, adjusted to provide estimates of treatment-naïve levels; na, not available

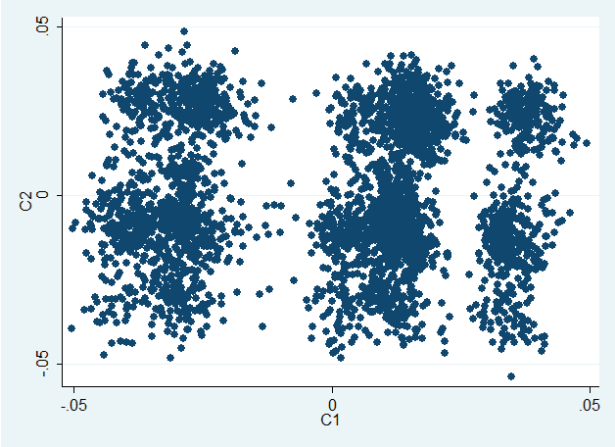


References:
CMD: Voight 2012 and Trynka 2011,
SCZ: Ripke et al 2014, MDD: Wray et al 2018, BD: Stahl et al 2018

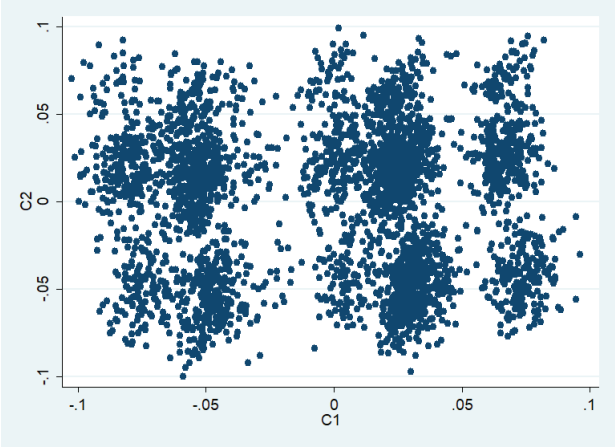

a
b
c
d
e
f
g
h
i
C1
C2

medRxiv preprint doi: https://doi.org/10.1101/2020.06.23.20138271; this version posted June 24, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

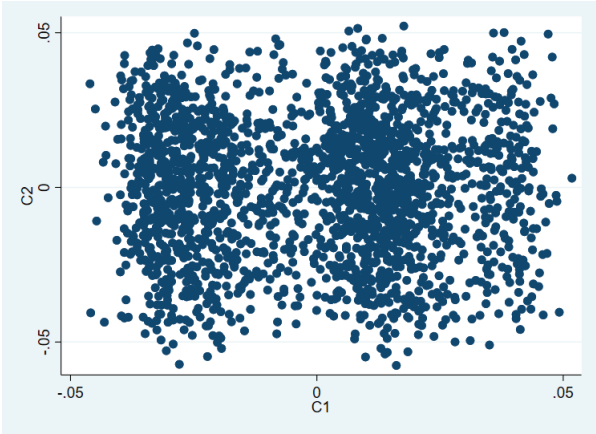
a)



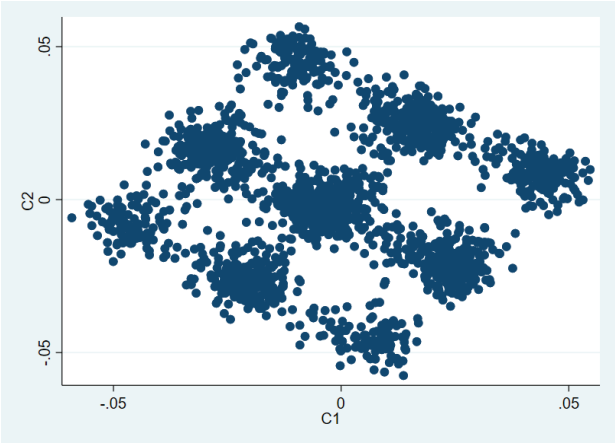
b)



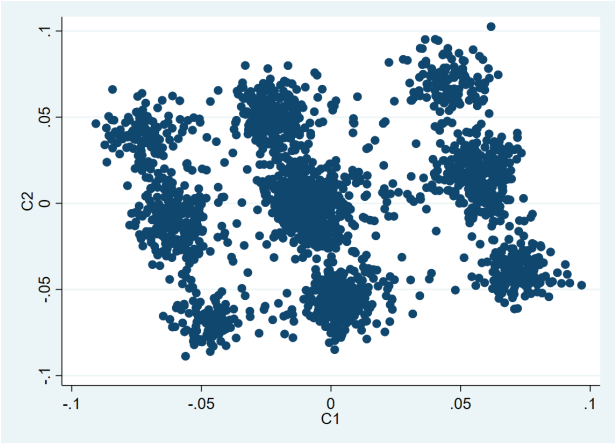
c)



d)

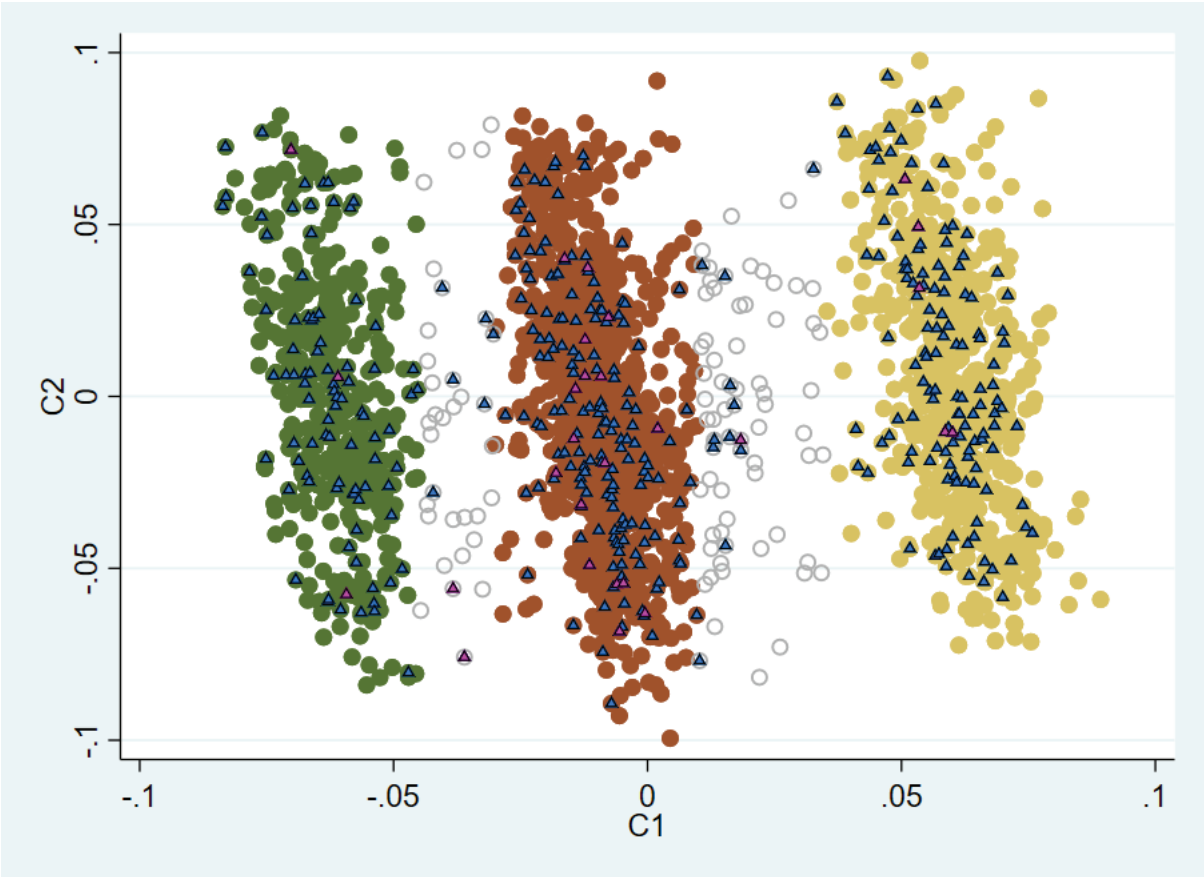


e)

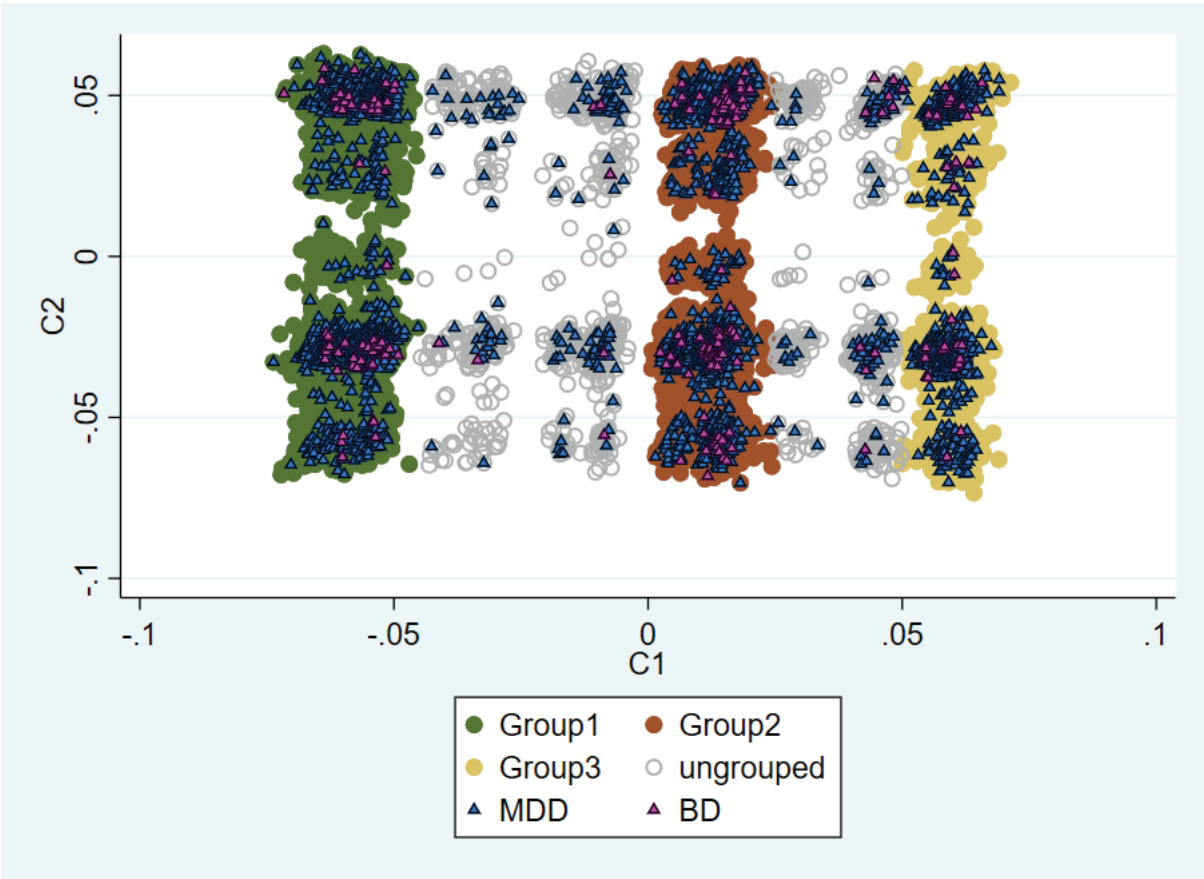


medRxiv preprint doi: https://doi.org/10.1101/2020.06.23.20138271; this version posted June 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

a)

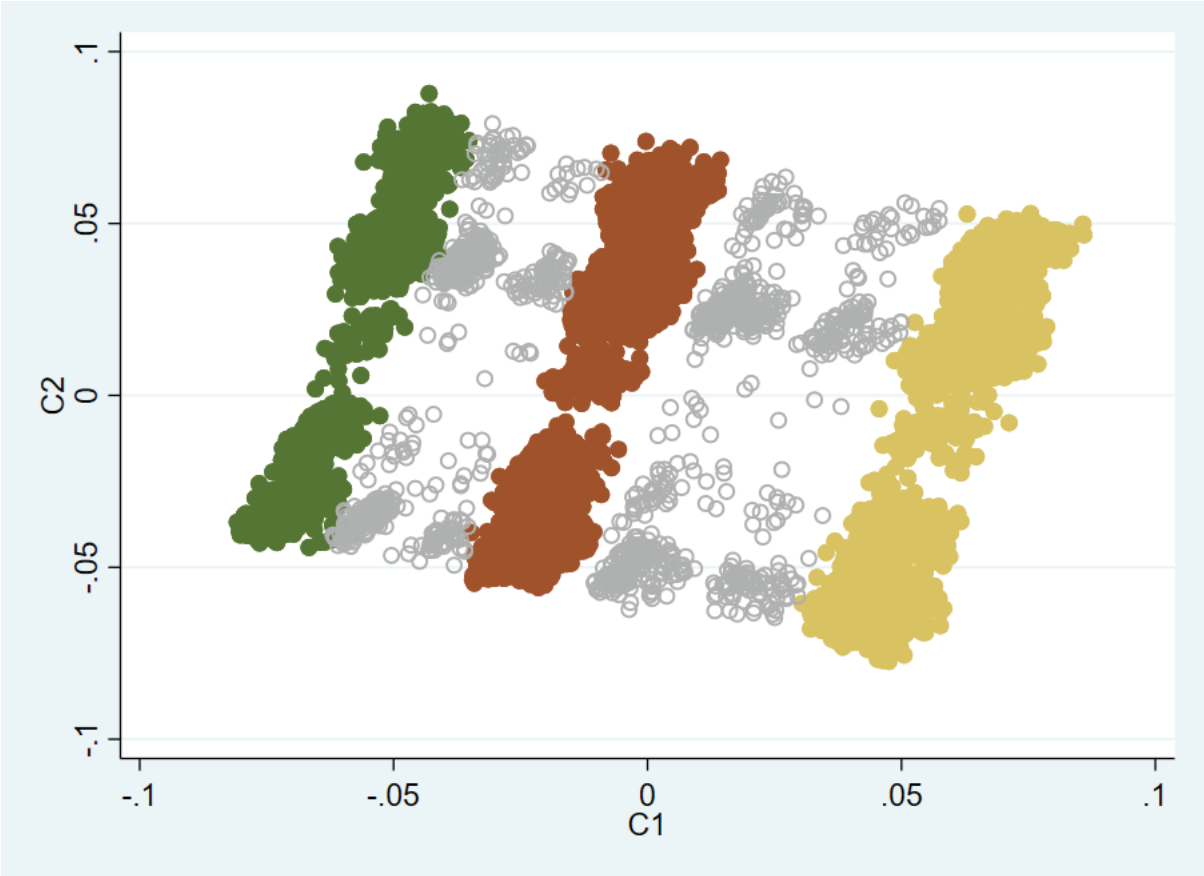
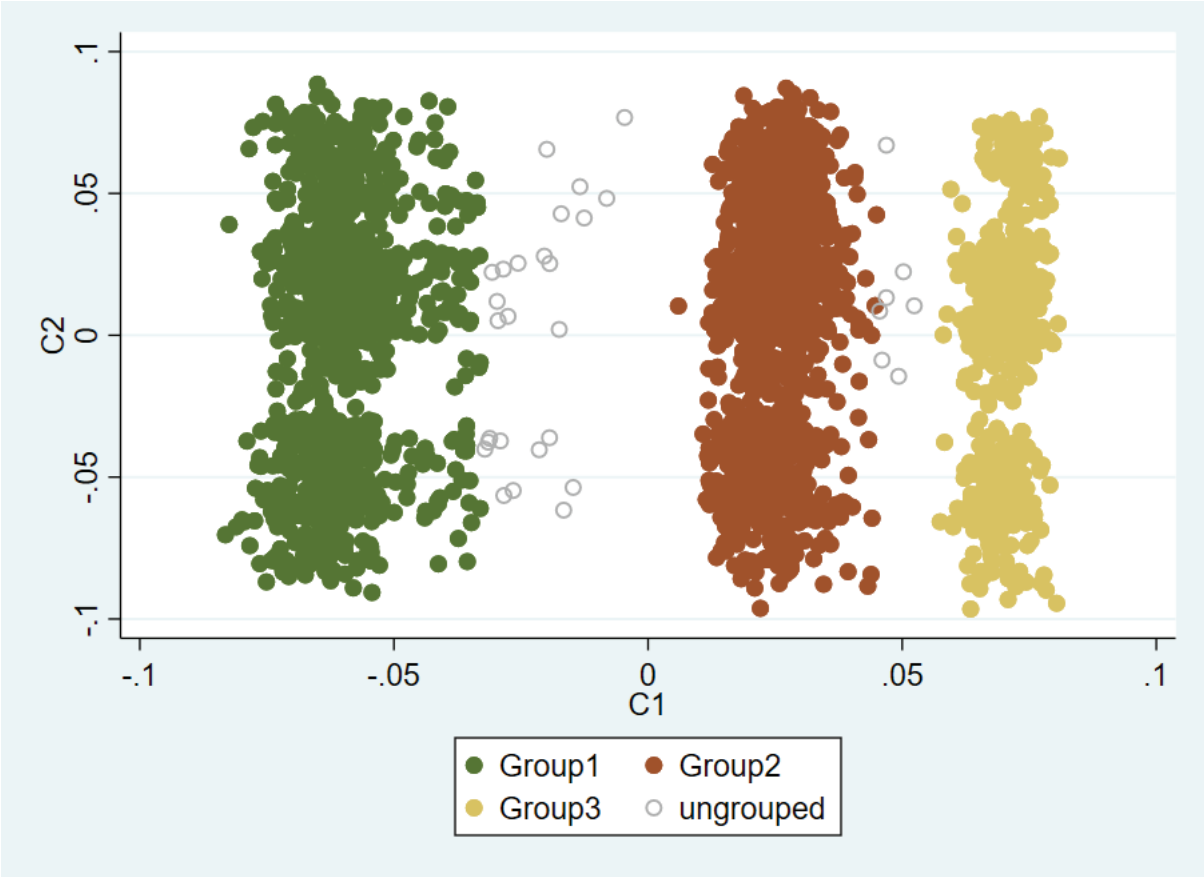


b)



medRxiv preprint doi: https://doi.org/10.1101/2020.06.23.20138271; this version posted June 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

a)



b)