The role of liver fat in cardiometabolic diseases is highlighted by genome-wide association study of MRI-derived measures of body composition

Dennis van der Meer, Ph.D.[1,2*], Tiril P. Gurholt, Ph.D.[1] Ida E. Sønderby, Ph.D.[1,3,4], Alexey A. Shadrin, Ph.D.[1], Guy Hindley, M.D.[1,5] Zillur Rahman, Ph.D.[1], Ann-Marie G. de Lange, Ph.D.[1,6,7], Oleksandr Frei, Ph.D.[1,8], Olof D. Leinhard, Ph.D.[9,10,11], Jennifer Linge, M.Sc.[9,10], Rozalyn Simon, Ph.D.[10,11], Lars T. Westlye, Ph.D.[1,4,12], Sigrun Halvorsen M.D. Ph.D.[13], Anders M. Dale, Ph.D.[14], Tom H. Karlsen M.D. Ph.D.[15,16], Tobias Kaufmann, Ph.D.[1,17] & Ole A. Andreassen, M.D. Ph.D.[1,4]

[1]*Norwegian Centre for Mental Disorders Research (NORMENT), Division of Mental Health and Addiction, Oslo University Hospital & Institute of Clinical Medicine, University of Oslo, Oslo, Norway*
[2]*School of Mental Health and Neuroscience, Faculty of Health, Medicine and Life Sciences, Maastricht University, The Netherlands*
[3]*Department of Medical Genetics, Oslo University Hospital, Oslo, Norway*
[4]*K.G. Jebsen Centre for Neurodevelopmental Disorders, University of Oslo*
[5]*Psychosis Studies, Institute of Psychiatry, Psychology and Neurosciences, King's College London, United Kingdom*
[6]*LREN, Centre for Research in Neurosciences, Dept. of Clinical Neurosciences, Lausanne University Hospital (CHUV) and University of Lausanne, Lausanne, Switzerland*
[7]*Dept. of Psychiatry, University of Oxford, Oxford, UK*
[8]*Centre for Bioinformatics, Department of Informatics, University of Oslo, Oslo, Norway*
[9]*AMRA Medical, Linköping, Sweden*
[10]*Division of Diagnostics and Specialist Medicine, Department of Health, Medicine and Caring Sciences, Linköping University, Linköping, Sweden*
[11]*Center for Medical Image Science and Visualization (CMIV), Linköping University, Linköping, Sweden*
[12]*Department of Psychology, University of Oslo, Oslo, Norway*
[13]*Department of Cardiology, Oslo University Hospital Ullevål, and University of Oslo, Oslo, Norway*
[14]*Center for Multimodal Imaging and Genetics, University of California at San Diego, La Jolla, CA 92037, USA*
[15]*Department of Transplantation Medicine, Division of Surgery, Inflammatory Diseases and Transplantation, Oslo University Hospital Rikshospitalet, Oslo, Norway.*
[16]*Research Institute for Internal Medicine, Division of Surgery, Inflammatory Diseases and Transplantation, Oslo University Hospital Rikshospitalet and University of Oslo, Oslo, Norway.*
[17]*Department of Psychiatry and Psychotherapy, University of Tübingen, Germany*

* Corresponding author. E-mail: d.v.d.meer@medisin.uio.no

Conflicts of interest

OAA has received speaker's honorarium from Lundbeck and is a consultant to HealthLytix. JL, and ODL are employed by and stockholders in AMRA Medical, and RS was previously employed by AMRA medical. THK received consultancy fees from Intercept and Engitix and speaker fees from Novartis, Gilead and AlfaSigma. AMD is a Founder of and holds equity in CorTechs Labs, Inc, and serves on its Scientific Advisory Board. He is a member of the Scientific Advisory Board of Human Longevity, Inc. and receives funding through research agreements with General Electric Healthcare and Medtronic, Inc. The terms of these arrangements have been reviewed and approved by UCSD in accordance with its conflict of interest policies. All other authors report no potential conflicts of interest.

Financial support

The authors were funded by the Research Council of Norway (276082, 213837, 223273, 204966/F20, 229129, 249795/F20, 225989, 248778, 249795, 298646, 300767), the South-Eastern Norway Regional Health Authority (2013-123, 2014-097, 2015-073, 2016-064, 2017-004, 2017-112, 2019-101, 2020-060 (IES), 2022-080), Stiftelsen Kristian Gerhard Jebsen (SKGJ-MED-021), The European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (ERC Starting Grant, Grant agreement No. 802998), ERA-Net Cofund through the ERA PerMed project 'IMPLEMENT', and National Institutes of Health (R01MH100351, R01GM104400). This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant agreement No 847776847776 (CoMorMent).

Author contributions

D.v.d.M., T.G., I.E.S. and O.A.A. conceived the study; D.v.d.M., T.G. and T.K. pre-processed the data. D.v.d.M. performed all analyses, with conceptual input from O.A.A.; All authors contributed to interpretation of results; D.v.d.M. drafted the manuscript and all authors contributed to and approved the final manuscript.

Abstract

Background & Aims: Obesity and associated morbidities, metabolic associated liver disease (MAFLD) included, constitute some of the largest public health threats worldwide. Body composition and related risk factors are known to be heritable and identification of their genetic determinants may aid in the development of better prevention and treatment strategies. Recently, large-scale whole-body MRI data has become available, providing more specific measures of body composition than anthropometrics such as body mass index. Here, we aimed to elucidate the genetic architecture of body composition, by conducting the first genome-wide association study (GWAS) of these MRI-derived measures.

Methods: We ran both univariate and multivariate GWAS on fourteen MRI-derived measurements of adipose and muscle tissue distribution, derived from scans from 34,036 White European UK Biobank participants (mean age of 64.5 years, 51.5% female).

Results: Through multivariate analysis, we discovered 108 loci with distributed effects across the body composition measures and 256 significant genes primarily involved in immune system functioning. Liver fat stood out, with a highly discoverable and oligogenic architecture and the strongest genetic associations. Comparison with 21 common cardiometabolic traits revealed both shared and specific genetic influences, with higher mean heritability for the MRI measures (h^2=.25 vs. .16, $p=1.4x10^{-6}$). We found substantial genetic correlations between the body composition measures and a range of cardiometabolic diseases, with the strongest correlation between liver fat and type 2 diabetes (r_g=.48, $p=1.6x10^{-22}$).

Conclusions: These findings show that MRI-derived body composition measures complement conventional body anthropometrics and other biomarkers of cardiometabolic health, highlighting

81 the central role of liver fat, and improving our knowledge of the genetic architecture of body
82 composition and related diseases.

83 **Keywords:** genome-wide association study; body composition; liver fat; whole-body MRI

Introduction

Obesity and associated cardiometabolic diseases are currently considered one of the largest global public health concerns[1,2]. Over one-third of the United States adult population qualifies for a diagnosis of metabolic syndrome,[3] characterized by excessive visceral adiposity, insulin resistance, hypertension, low high-density lipoprotein cholesterol, and hypertriglyceridemia.[4,5] Metabolic syndrome substantially increases the risk of coronary artery disease, type 2 diabetes, cancer, and metabolic associated fatty liver disease (MAFLD, previously described as non-alcoholic fatty liver disease[6]).[7–11] Body composition is also associated with brain structure and brain disorders.[12,13] An improved understanding of the genetic and biological determinants of body composition is needed to provide insights into the complex interplay between metabolic factors, prevent and treat multiple highly prevalent conditions, and improve public health outcomes.[2,10]

Body composition is partly determined by a complex constellation of interacting metabolic processes and inter-organ cross-talk that may become dysregulated and lead to metabolic syndrome.[14] In susceptible individuals, excessive energy intake, stored as visceral adipose tissue, combined with insulin resistance, leads to heightened lipolysis and release of free fatty acids.[15] Increased free fatty acid flux to the liver results in hypertriglyceridemia, which in turn contributes to dyslipidemia and atherosclerosis. Lipolysis in visceral adipose tissue further promotes insulin resistance and gluconeogenesis and increases pro-inflammatory reactions that exacerbate endothelial dysfunction and hypertension.[15] This is reflected in heightened levels of pro-inflammatory markers among individuals with metabolic syndrome.[16] Muscle mass is also a determinant of cardiometabolic health,[17] as skeletal muscle constitutes the largest insulin-sensitive tissue in the body and is the primary site for insulin-stimulated glucose utilization.[18] Still, the

nature and extent of overlap between these different determinants of cardiometabolic functioning remain unclear.

Measures of localized adipose tissue, liver fat and regional muscle volume can now be accurately extracted from whole-body MRI scans.[19–22] MRI-based body tissue quantification offers more sensitive proxies of cardiometabolic health than body anthropometrics such as waist circumference and body mass index (BMI)[23], which also lack a direct connection to pathophysiology.[5,24] Measures of regional adipose tissue, most accurately and comprehensively identified through MRI,[25,26] show independent associations with cardiometabolic diseases and improve risk prediction beyond body anthropometrics.[27–29]

In addition to social and physical environmental factors,[30] genetically determined individual differences play a significant role in regulating body composition.[31–33] Cardiometabolic risk factors have both unique and shared genetic correlates.[34] Much less is known about the genetics of specific MRI-derived body composition measures.[35] We aimed to map the unique and shared genetic architectures across the MRI-derived body composition to provide a holistic understanding of the interplay between different tissue types and their role in metabolic syndrome and cardiometabolic health.

Results

We conducted GWASs of fourteen MRI-derived muscle and adipose tissue distribution measures and investigated the genetic link to conventional cardiometabolic risk factors. We included six measures of adipose tissue distribution: abdominal subcutaneous adipose tissue, visceral adipose tissue, abdominal fat ratio, anterior and posterior thigh muscle fat infiltration, and liver protein density fat fraction. Additionally, we investigated three measures related to thigh muscle tissue, namely anterior and posterior thigh muscle volume and weight-to-muscle ratio. We further

129 analyzed visceral and abdominal adipose tissue, and anterior and posterior muscle volume, divided
130 by standing height in meters squared, and total thigh muscle volume z-score (sex-, height-, weight-
131 and BMI-invariant).[36] See Table 1 for an overview of these measures, and the Methods section for
132 protocols and definitions. Given a total of fourteen individual measures, we set the univariate
133 GWAS significance threshold at $\alpha=5*10^{-8}/14=3.6*10^{-9}$. Our sample for the main analyses
134 consisted of 34,036 White European participants of the UK Biobank (UKB), with a mean age of
135 64.5 years (standard deviation (SD) 7.4 years), 51.5% female. We pre-residualized all measures
136 for age, sex, test center, and the first twenty genetic principal components to control for population
137 stratification.

138 *Table 1. MRI-derived measures of body composition included in this study, together with the*
139 *available sample size and number of loci discovered through univariate GWAS.*

Measure	*Abbreviation*	*N*	*# loci*
Abdominal subcutaneous adipose tissue	ASAT	33979	1
Visceral adipose tissue	VAT	33989	2
Anterior thigh muscle volume	ATMV	33415	7
Posterior thigh muscle volume	PTMV	33459	10
Anterior thigh muscle fat infiltration (%)	ATMFI	33347	19
Posterior thigh muscle fat infiltration (%)	PTMFI	33392	25
Weight-muscle-ratio	WMR	33406	1
Abdominal fat ratio	AFR	33375	1
Liver proton density fat fraction (%)	LPDFF	33674	8
VAT/height^2	VATi	33007	2
ASAT/height^2	ASATi	32998	1
ATMV/height^2	ATMVi	32442	5
PTMV/height^2	PTMVi	32483	0
Total thigh muscle volume z-score	TTMVz	32401	5

140

141 *Univariate GWAS*

142 Univariate GWASs on the individual measures revealed a total of 87 loci, including 54 unique,
143 surpassing the study-wide significance threshold of $3.6*10^{-9}$. Two loci stood out with highly

144 significant p-values, on chromosome 19 (lead rs58542926, $p=8.8x10^{-113}$) and chromosome 22
145 (lead rs738409, $p=4.6x10^{-166}$), both identified in the GWAS on liver fat. Using converging
146 positional, eQTL, and chromatin interaction information (see Methods), we mapped these loci to
147 genes previously coupled to MAFLD (rs738409: *PNPLA3*, *SAMM50*, *PARVB*)[37] as well as
148 inflammatory processes and cancer (rs58542926: *CD99*).[38] The Supplementary Information (SI)
149 contains Manhattan plots and overviews of all loci discovered together with mapped genes.

150 Additionally, we assessed the generalization of the discovered loci in a hold-out set of 5,081
151 non-White European UKB participants with identical processing steps. Of the 84 lead SNPs
152 available in this set, 82 had effects in the same direction as the main analyses (97.6%, sign-test
153 $p<1x10^{-16}$). Thus, our results suggest a cross-ethnicity generalization of these genetic associations
154 with MRI-derived measures of body composition, despite the known high variability of body
155 anthropometrics across ethnicities.[5,24]

156 In total, we identified eight study-wide significant loci for liver fat, validating those found in a
157 previous smaller GWAS.[35] Gene-based analysis through Multi-marker Analysis of GenoMic
158 Annotation (MAGMA) identified 35 genome-wide significant genes, including the three primary
159 MAFLD genes (*TM6SF2* $p=7.2*10^{-16}$, *PNPLA3* $p=1.0*10^{-14}$, and *TMC4-MBOAT7* $p=2.1*10^{-8}$),[39–
160 42] further confirming the strong biological validity of this liver fat measure and its connection to
161 MAFLD. Functional annotation of the set of 35 genes revealed differential expression in the liver,
162 pancreas, and subcortical brain regions and significant enrichment among Gene Ontology (GO)
163 biological processes specifically related to lipid homeostasis and metabolic processes. The SI
164 further contains results of gene set enrichment analyses for each individual measure.

165 Next, we estimated the polygenicity and effect size variance ('discoverability') by fitting a
166 Gaussian mixture model of null and non-null effects to the GWAS summary statistics using

MiXeR.[43,44] The results are summarized in Figure 1a, depicting the estimated proportion of genetic variance explained by discovered SNPs for each measure as a function of sample size. This illustrates that body MRI measures generally show genetic architectures similar to e.g. brain MRI measures, characterized by high polygenicity.[45,46] However, the notable exception is liver fat, with substantially lower polygenicity and higher discoverability than the other measures, in line with the relatively few highly significant associations we identified through the GWAS.

Figure 1b visualizes the phenotypic and genetic correlations between each pair of measures, confirming a strong structure and a subdivision between adipose- and muscle-related measures. SNP-based heritability ranged from 19% to 34% (all $p<1x10^{-16}$); see the diagonal of Figure 1b.

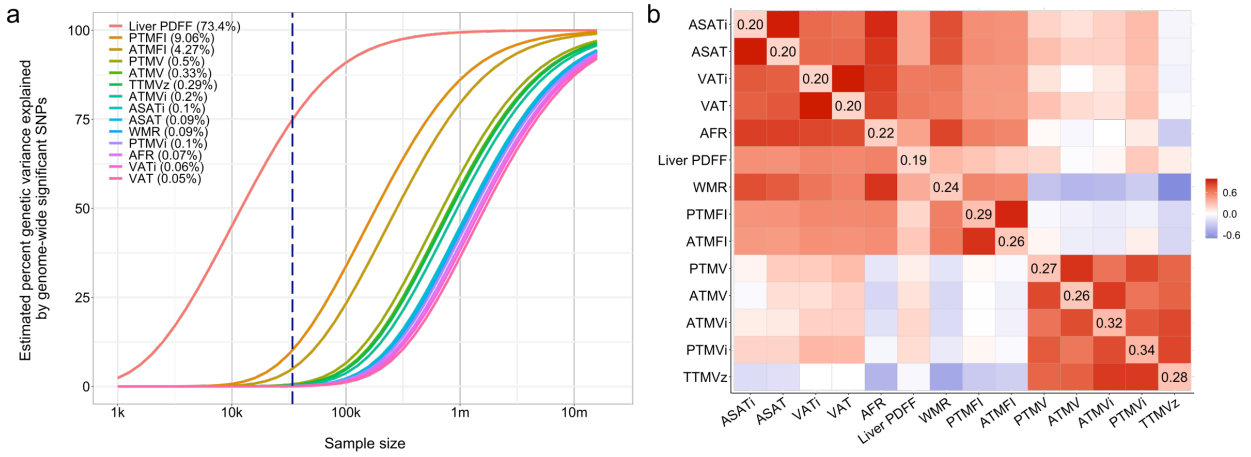

Figure 1. Comparison of the genetic architecture of individual body composition measures. a) *The relation between genetic variance explained by genome-wide significant hits (y-axis) and sample size (x-axis) for each measure (solid colored lines). The vertical dashed blue line marks the current sample size, with the corresponding percent genetic variance explained indicated between brackets in the legend.* ***b)*** *Correlation between the measures, with phenotypic correlation shown in bottom triangle and genetic correlation in the upper triangle, and heritability on the diagonal. Abbreviations: ASAT=abdominal subcutaneous adipose tissue, VAT=visceral adipose tissue, AFR= abdominal fat ratio, WMR=weight-muscle-ratio, ATMV=anterior thigh muscle volume, PTMV=Posterior thigh muscle volume, ATMFI=anterior thigh muscle fat infiltration, PTMFI=posterior thigh muscle fat infiltration, Liver PDFF=liver proton density fat fraction, TTMVz=total thigh muscle volume z-score, i=index, referring to a measure divided by standing height*^{2}.

177 *Multivariate GWAS*

178 Gene variants are likely to have distributed effects across these measures, as they are correlated
179 components of the same biological system. We therefore also jointly analyzed all measures
180 through the Multivariate Omnibus Statistical Test (MOSTest),[47] which increases statistical power
181 in a scenario of shared genetic signal across the univariate measures.[47–49] After applying a rank-
182 based inverse normal transformation, we performed MOSTest on the residualized measures,
183 yielding a multivariate association with 9.1 million SNPs included.

184 MOSTest revealed 108 significant independent loci across all MRI-derived measures (see
185 Figure 2a and SI). Figure 2b visualizes the significance of the association between the individual

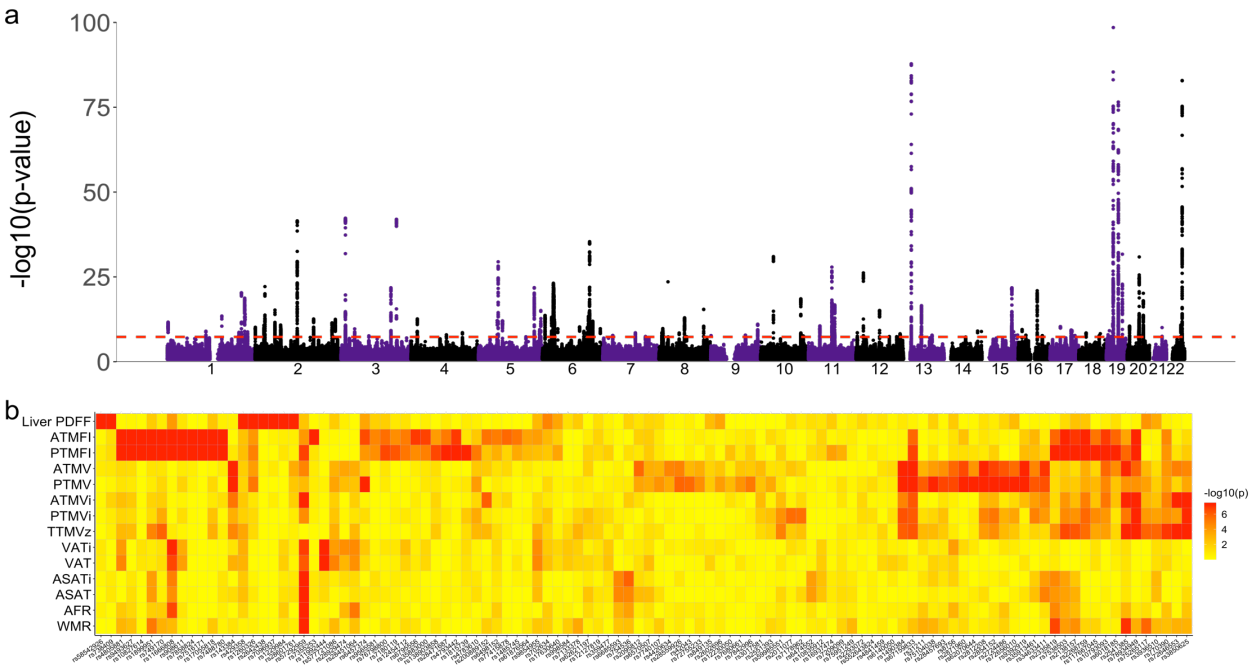

Figure 2. Multivariate locus discovery a) *Manhattan plot of the multivariate GWAS on all MRI-derived body composition measures, with the observed -log10(p) of each SNP shown on the y-axis. The x-axis shows the relative genomic location, grouped by chromosome, and the red dashed line indicates the whole-genome significance threshold of $5x10^{-8}$. The y-axis is clipped at -log10(p)=75.* ***b)*** *Heatmap showing -log10(p) of the association between the lead variants of MOSTest-identified independent loci (x-axis) and each of the individual MRI measures (y-axis). The values are capped at 7.5 ($p=5x10^{-8}$).*

186 measures and each of the 108 loci, illustrating the presence of many shared but also specific genetic
187 variants.

188 MAGMA identified 256 significant genes after multiple comparison correction (α=.05/18,203),
189 with highly significant differential expression in the liver, pancreas, heart, muscle, and several
190 other tissues (Figure 3). Coupling the significant genes to the Reactome database[50] indicated most
191 prominent associations with the adaptive immune system and cytokine signaling ($p<1*10^{-16}$), see
192 Supplementary Data and Supplementary Figure 2 for an overview.

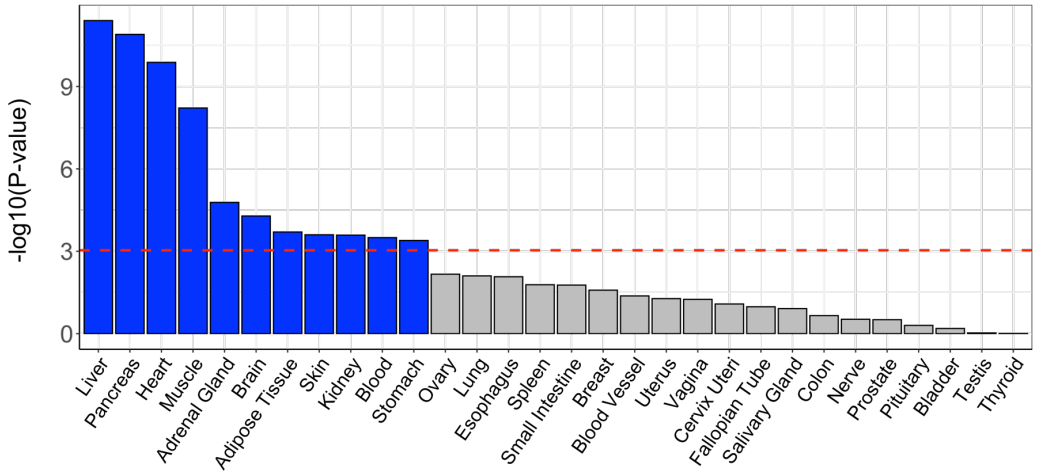

Figure 3. *Tissue-specific differential expression of the set of significant genes identified through the multivariate GWAS on MRI-derived measures of body composition. The red-dotted line indicates the multiple comparisons-corrected significance threshold.*

193

194 *Comparison of genetic architecture with cardiometabolic risk factors*

195 We additionally analyzed a set of 21 measures of anthropometric and cardiometabolic factors (e.g.,
196 BMI, triglycerides, cholesterol, blood pressure; see Table 2), which were available for up to
197 412,316 White European UKB participants. Through multivariate GWAS on this separate set of
198 measures in the full UKB sample, we found 1134 genome-wide significant loci with $\alpha=5\times10^{-8}$ (list
199 provided in SI). Of the 108 loci identified through the primary multivariate analysis of MRI-

derived body composition measures, 94 (87%) were significant in this secondary analysis in a larger sample. This indicates that these sets of measures overall are influenced by the same network of biological processes.

The heritability of the MRI-derived measures (mean h^2=.25) was significantly higher than the body anthropometrics and other biomarkers (mean h^2=.16), p=1.4x10^{-6}. As shown in Figure 4, these measures generally showed higher genetic correlations with the MRI-derived measures of adipose tissue than the muscle-related measures. Further, BMI, hip/waist circumference, and waist-to-hip-ratio were genetically correlated with nearly all body MRI measures.

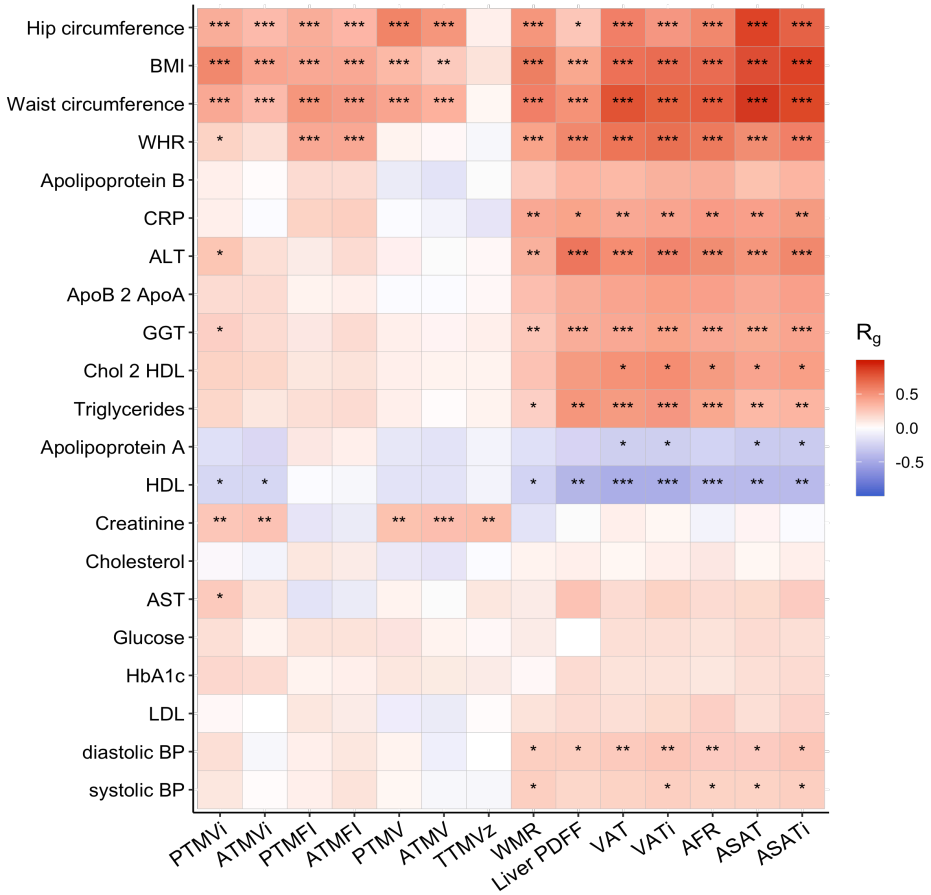

Figure 4. *Genetic correlations of the MRI-derived body composition measures with standard anthropometrics and cardiometabolic measures. Abbreviations: BMI=body mass index, WHR=waist-hip ratio, CRP=C-reactive protein, ALT=alanine aminotransferase, GGT= gamma-glutamyl transferase, HDL=high-density lipoproteins, AST=aspartate aminotransferase, HbA1c=glycated hemoglobin, LDL=low-density lipoproteins, BP=blood pressure.*

Genetic correlation with cardiovascular, metabolic and mental disorders

Next, we analyzed the genetic overlap of the MRI-derived measures with medical conditions previously linked to cardiometabolic health, selecting recent GWAS with adequate power.[51–57] As shown in Figure 5a, the strongest association across all measures was found for liver fat, with a genetic correlation of 0.48 ($p=1.6x10^{-22}$) with type 2 diabetes. Coronary artery disease was found to have highly significant positive genetic correlations with visceral and subcutaneous adipose tissue. Overall, we found weak negative genetic correlations with muscle tissue measures and stronger positive genetic correlations with adipose tissue measures, with two exceptions; anorexia nervosa showed the opposite direction of correlation compared to the other conditions, and there was no discernible pattern for schizophrenia. Genetic correlations with the anthropometric and metabolic measures are provided in Figure 5b for comparison, indicating that the adipose tissue measures are as strong as or stronger correlated with these conditions than the conventional body anthropometrics.

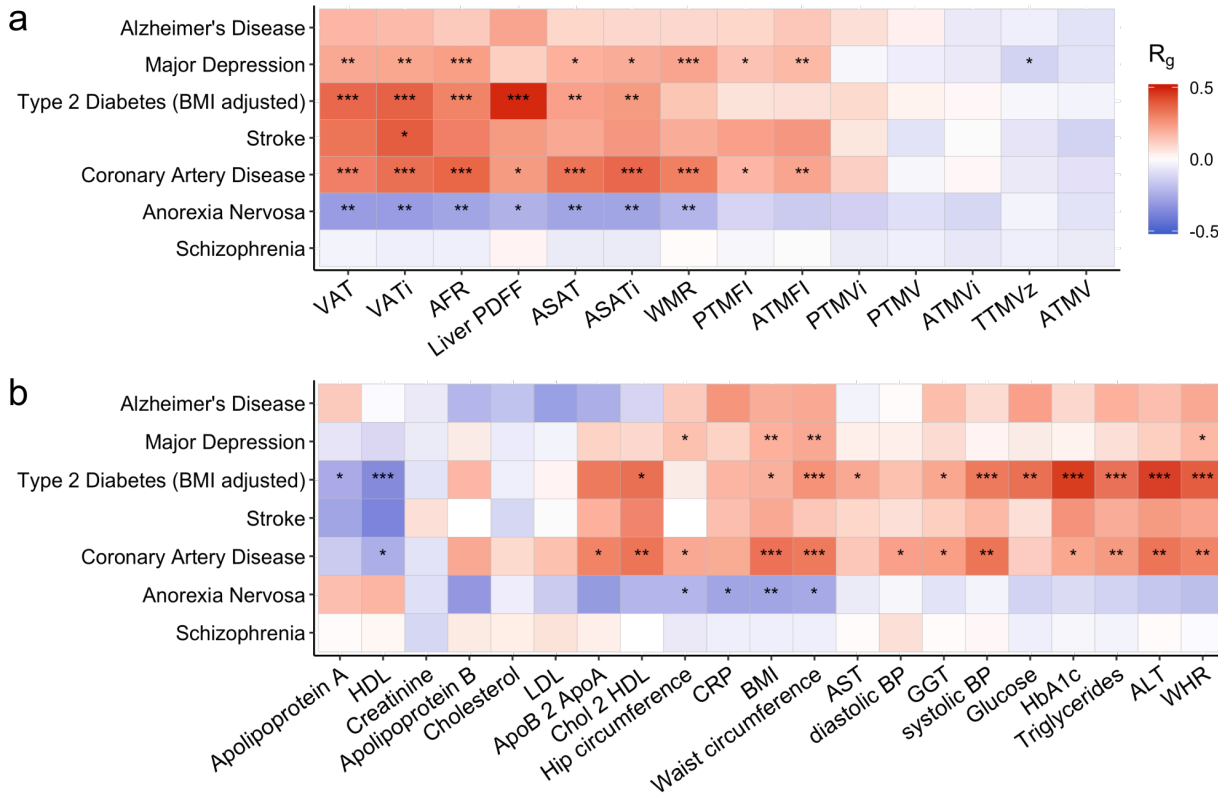

Figure 5. *Genetic correlations of* ***a****) MRI-derived body composition measures, and* ***b****) anthropometric and metabolic measures (x-axis) with conditions linked to poor cardiometabolic health (y-axis). ***p=5x10^{-9}, **p=5x10^{-6}, *p=5x10^{-4}.*

221

Discussion

Here, we reported results from a comprehensive, large-scale GWAS of MRI-derived measures of body composition. Joint analyses of measures of regional adipose and muscle tissue distributions revealed extensive genetic overlap and led to the identification of a large number of shared genetic risk loci across traits. We further showed genetic overlap with body anthropometrics and cardiometabolic measures as well as medical conditions linked to cardiometabolic health. Our findings illustrate how MRI-derived measures can be leveraged to improve our understanding of the biology underlying the metabolic system, identifying liver fat as a particularly promising measure, highlighting the integral role of steatosis and MAFLD in cardiometabolic health.

The genetic correlations of body composition measures with common medical conditions underlined that they may complement conventional measures to better understand cardiometabolic health. Liver fat showed a stronger genetic correlation with type 2 diabetes than conventional measures. The amount of liver fat and its genetic determinants may thus play a central role in type 2 diabetes development, and at a minimum robustly positions MAFLD onto the map of relevant comorbidities of type 2 diabetes alongside cardiovascular disease, kidney disease and diabetic retinopathy. Further, we found significant positive genetic correlations between coronary artery disease and visceral and subcutaneous adipose tissue, adding genetic evidence to the well-established relation between this disease, obesity, and body fat distribution.[58]

Liver fat also stood out from the other measures with regard to its genetic architecture. While all traits investigated were substantially heritable, the genetic discoverability of liver fat was much higher, with an oligogenic architecture as opposed to the polygenic architectures of the remaining traits and other complex biomedical measures.[45] This was reflected in the GWAS yield, with a few highly significant loci coupled to lipid homeostasis explaining the majority of genetic variance for

this measure. These loci should be scrutinized for the biological link between liver fat and cardiometabolic conditions,[59] and may potentially point to fundamental processes that become dysregulated in these diseases. Indeed, all components of metabolic syndrome correlate with liver fat content.[60] Evaluation of MAFLD risk has been recommended for any individual with metabolic syndrome and related morbidities (e.g. type 2 diabetes),[11,60] and the large effects of these liver fat-associated loci even may suggest potential as features for individual risk stratification in MAFLD.[61,62] These findings also attest to the accuracy and clinical relevance of MRI-derived measures of liver fat, and support the notion that MAFLD should be considered an integral component of obesity and metabolic syndrome and a key non-communicable disease.[11]

Another key finding was that the highest number of significant loci were found for muscle fat infiltration in the anterior and posterior thighs, two measures not previously genetically studied. Fatty infiltration of skeletal muscle reduces the muscle mass and strength,[63] and has been implicated in aging and frailty.[64] It has also been coupled to metabolic syndrome[65] and cardiovascular mortality.[66] Recent literature focused on liver disease and its progression have also highlighted the importance of muscle health.[67] Muscle fat infiltration has been linked to higher comorbidity within MAFLD and decreased muscle fat infiltration has been correlated with improvement in steatohepatitis.[68,69] Our findings suggest a strong genetic component to these associations, indicated by the large degree of shared genetic architecture with related diseases. Interestingly, fat accumulation in the muscle arises through specific pathways, including the intramyocellular accumulation of lipid,[63] which is associated with insulin insensitivity and inflammation.[70]

The genetic correlations between the MRI-derived body composition measures indicate partly overlapping biological processes with some unique genetic determinants. The correlation structure

268 further suggests that adipose tissue distribution is genetically largely independent from muscle
269 tissue. However, it should be noted that global correlations underestimate overlap when a mixture
270 of genetic effects in the same and opposing directions cancels each other out.[46] Indeed, adipose
271 and muscle tissue are known to have complex regulatory cross-talk, both releasing metabolism-
272 regulating molecules to maintain a balanced weight-to-muscle ratio.[71] The increased yield from
273 the multivariate GWAS analysis, nearly doubling the number of unique loci discovered, is in line
274 with the hypothesis of strong biological interplay and shared molecular mechanisms. The
275 multivariate GWAS allowed for identifying loci that have distributed effects across the included
276 body composition measures. These may help to explain the complexity of metabolic syndrome
277 and the frequent comorbidity between diseases associated with body composition. Our findings
278 that a substantial portion of the genetic determinants of these measures are related to the immune
279 system fit with a large body of literature indicating that adipose tissue is an active metabolic and
280 endocrine organ that secretes a host of pro- and anti-inflammatory factors, and with the
281 characterization of obesity as a state of chronic low-grade inflammation.[72] Thus, the current
282 genetic findings can form the basis for functional follow-up studies to determine the molecular
283 mechanisms involved in the complex relations between lipids and the immune system.

284 There was high genetic overlap between the sets of MRI-derived measures of body composition
285 and the conventional measures of body anthropometrics and cardiometabolic health, indicating
286 that they tag similar biological processes. The body anthropometrics were correlated with both
287 muscle and adipose tissue, indicating little specificity, in line with the long-standing recognized
288 limitations of these global measures that they fail to distinguish between specific body types that
289 differ widely in risk for disease.[73]

Strengths of this study are the large number of whole-body MRI scans and the use of state-of-the-art, precise body composition measures, including multiple measures not previously investigated. With this, we were able to replicate loci reported earlier in smaller samples and with different measurement protocols.[35] We further combined the study of individual measures with a multivariate approach to genetic discovery, allowing for greater GWAS hit yield and insight into the overall architecture of these complementary indicators of body composition and associated diseases. The findings allow for numerous follow-up investigations; further studies are needed to clarify the causal directions between the measures, and the role of putative moderators such as sex,[74] age, and ethnicity.[75]

To conclude, the high prevalence of cardiometabolic diseases, combined with substantial morbidity and mortality, indicates a strong need for new therapeutic targets. While these diseases are often comorbid, they are treated separately, with this polypharmacy bringing along increased risk of adverse drug reactions.[4] Genetic data is less subjected to reverse causation and confounding than environmental factors. Knowledge about shared and specific genetic determinants is therefore central to develop effective strategies that optimally treat the individual. We showed that accurate MRI-derived measures of liver and regional adipose and muscle tissue characteristics have strong genetic components, with shared influences that can be leveraged to boost discovery. As such, these findings have the potential to significantly enhance our understanding of body composition and related diseases, provide drug targets for MAFLD and related traits, and contribute to combatting a significant, increasing threat to public health.

References

1. Heymsfield, S. B. & Wadden, T. A. Mechanisms, pathophysiology, and management of obesity. *N. Engl. J. Med.* **376**, 254–266 (2017).

2. Ralston, J. & Nugent, R. Toward a broader response to cardiometabolic disease. *Nat. Med.* **25**, 1644–1646 (2019).

3. Aguilar, M., Bhuket, T., Torres, S., Liu, B. & Wong, R. J. Prevalence of the metabolic syndrome in the United States, 2003-2012. *Jama* **313**, 1973–1974 (2015).

4. Mendrick, D. L. *et al.* Metabolic syndrome and associated diseases: from the bench to the clinic. *Toxicol. Sci.* **162**, 36–42 (2018).

5. Eckel, R. H., Alberti, K. G. M. M., Grundy, S. M. & Zimmet, P. Z. The metabolic syndrome. *Lancet* **375**, 181–183 (2010).

6. Eslam, M., Ratziu, V. & George, J. Yet more evidence that MAFLD is more than a name change. *J. Hepatol.* **74**, 977–979 (2021).

7. Grundy, S. M. Metabolic syndrome pandemic. *Arterioscler. Thromb. Vasc. Biol.* **28**, 629–636 (2008).

8. Isomaa, B. O. *et al.* Cardiovascular morbidity and mortality associated with the metabolic syndrome. *Diabetes Care* **24**, 683–689 (2001).

9. Esposito, K., Chiodini, P., Colao, A., Lenzi, A. & Giugliano, D. Metabolic syndrome and risk of cancer: a systematic review and meta-analysis. *Diabetes Care* **35**, 2402–2411 (2012).

10. Diehl, A. M., Farpour-Lambert, N. J., Zhao, L. & Tilg, H. Why we need to curb the emerging worldwide epidemic of nonalcoholic fatty liver disease. *Nat. Metab.* **1**, 1027–1029 (2019).

11. Karlsen, T. H. *et al.* The EASL-Lancet Liver Commission: protecting the next generation

of Europeans against liver disease complications and premature mortality. *Lancet (London, England)* **399**, 61–116 (2022).

12. Gurholt, T. P. *et al.* Population-based body–brain mapping links brain morphology with anthropometrics and body composition. *Transl. Psychiatry* **11**, 1–12 (2021).

13. Firth, J. *et al.* The Lancet Psychiatry Commission: a blueprint for protecting physical health in people with mental illness. *The Lancet Psychiatry* **6**, 675–712 (2019).

14. Priest, C. & Tontonoz, P. Inter-organ cross-talk in metabolic syndrome. *Nat. Metab.* **1**, 1177–1188 (2019).

15. Samson, S. L. & Garber, A. J. Metabolic syndrome. *Endocrinol. Metab. Clin.* **43**, 1–23 (2014).

16. Rosmond, R. Role of stress in the pathogenesis of the metabolic syndrome. *Psychoneuroendocrinology* **30**, 1–10 (2005).

17. Atlantis, E., Martin, S. A., Haren, M. T., Taylor, A. W. & Wittert, G. A. Inverse associations between muscle mass, strength, and the metabolic syndrome. *Metabolism* **58**, 1013–1022 (2009).

18. Stump, C. S., Henriksen, E. J., Wei, Y. & Sowers, J. R. The metabolic syndrome: role of skeletal muscle metabolism. *Ann. Med.* **38**, 389–402 (2006).

19. Leinhard, O. D. *et al.* Quantitative abdominal fat estimation using MRI. in *2008 19th International Conference on Pattern Recognition* 1–4 (IEEE, 2008).

20. Karlsson, A. *et al.* Automatic and quantitative assessment of regional muscle volume by multi-atlas segmentation using whole-body water–fat MRI. *J. Magn. Reson. Imaging* **41**, 1558–1569 (2015).

21. Borga, M. *et al.* Advanced body composition assessment: from body mass index to body

composition profiling. *J. Investig. Med.* **66**, 1–9 (2018).

22. Borga, M. *et al.* Validation of a fast method for quantification of intra-abdominal and subcutaneous adipose tissue for large-scale human studies. *NMR Biomed.* **28**, 1747–1753 (2015).

23. Beck, E., Esser, N., Paquot, N. & Scheen, A. J. Metabolically obese normal-weight individuals and metabolically healthy, but obese, subjects. *Rev. Med. Suisse* **5**, 1644–1646 (2009).

24. Alberti, K. G. M. M. *et al.* Harmonizing the metabolic syndrome: a joint interim statement of the international diabetes federation task force on epidemiology and prevention; national heart, lung, and blood institute; American heart association; world heart federation; international . *Circulation* **120**, 1640–1645 (2009).

25. Shuster, A., Patlas, M., Pinthus, J. H. & Mourtzakis, M. The clinical importance of visceral adiposity: a critical review of methods for visceral adipose tissue analysis. *Br. J. Radiol.* **85**, 1–10 (2012).

26. Browning, L. M. *et al.* Measuring abdominal adipose tissue: comparison of simpler methods with MRI. *Obes. Facts* **4**, 9–15 (2011).

27. Britton, K. A. *et al.* Body fat distribution, incident cardiovascular disease, cancer, and all-cause mortality. *J. Am. Coll. Cardiol.* **62**, 921–925 (2013).

28. Stanley, M. R. I. A. S. G. of the L. A. R. G. G. D. dg108@ columbia. edu K. D. E. Y. J.-E. S. N. A. J. B. L. P.-S. F. X. H. Adipose tissue distribution is different in type 2 diabetes. *Am. J. Clin. Nutr.* **89**, 807–814 (2009).

29. Neeland, I. J. *et al.* Body fat distribution and incident cardiovascular disease in obese adults. *J. Am. Coll. Cardiol.* **65**, 2150–2151 (2015).

30. Agata, K. & Monyeki, M. A. Association between sport participation, body composition, physical fitness, and social correlates among adolescents: the PAHL study. *Int. J. Environ. Res. Public Health* **15**, 2793 (2018).

31. Maes, H. H. M., Neale, M. C. & Eaves, L. J. Genetic and environmental factors in relative body weight and human adiposity. *Behav. Genet.* **27**, 325–351 (1997).

32. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).

33. Lind, L. Genome-Wide Association Study of the Metabolic Syndrome in UK Biobank. *Metab. Syndr. Relat. Disord.* **17**, 505–511 (2019).

34. Lind, L. Genetic Determinants of Clustering of Cardiometabolic Risk Factors in U.K. Biobank. *Metab. Syndr. Relat. Disord.* **18**, 121–127 (2020).

35. Liu, Y. *et al.* Genetic architecture of 11 organ traits derived from abdominal MRI using deep learning. *Elife* **10**, e65554 (2021).

36. Linge, J., Heymsfield, S. B. & Dahlqvist Leinhard, O. On the Definition of Sarcopenia in the Presence of Aging and Obesity—Initial Results from UK Biobank. *Journals Gerontol. Ser. A* **75**, 1309–1316 (2020).

37. Kawaguchi, T. *et al.* Risk estimation model for nonalcoholic fatty liver disease in the Japanese using multiple genetic markers. *PLoS One* **13**, e0185490 (2018).

38. Pasello, M., Manara, M. C. & Scotlandi, K. CD99 at the crossroads of physiology and pathology. *J. Cell Commun. Signal.* **12**, 55–68 (2018).

39. Mancina, R. M. *et al.* The MBOAT7-TMC4 Variant rs641738 Increases Risk of Nonalcoholic Fatty Liver Disease in Individuals of European Descent. *Gastroenterology* **150**, 1219-1230.e6 (2016).

403 40. Anstee, Q. M. & Day, C. P. The genetics of nonalcoholic fatty liver disease: spotlight on
404 PNPLA3 and TM6SF2. in *Seminars in liver disease* **35**, 270–290 (Thieme Medical
405 Publishers, 2015).
406 41. Liu, Y.-L. *et al.* TM6SF2 rs58542926 influences hepatic fibrosis progression in patients
407 with non-alcoholic fatty liver disease. *Nat. Commun.* **5**, 1–6 (2014).
408 42. Romeo, S. *et al.* Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty
409 liver disease. *Nat. Genet.* **40**, 1461–1465 (2008).
410 43. Holland, D. *et al.* Beyond SNP Heritability: Polygenicity and Discoverability of
411 Phenotypes Estimated with a Univariate Gaussian Mixture Model. *bioRxiv* 498550 (2019).
412 doi:10.1101/498550
413 44. Frei, O. *et al.* Bivariate causal mixture model quantifies polygenic overlap between
414 complex traits beyond genetic correlation. *Nat. Commun.* **10**, 2417 (2019).
415 45. Holland, D. *et al.* Beyond SNP heritability: Polygenicity and discoverability of
416 phenotypes estimated with a univariate Gaussian mixture model. *PLOS Genet.* **16**,
417 e1008612 (2020).
418 46. van der Meer, D. *et al.* Quantifying the polygenic architecture of the human cerebral
419 cortex: Extensive genetic overlap between cortical thickness and surface area. *Cereb.*
420 *Cortex* **30**, 5597–5603 (2020).
421 47. van der Meer, D. *et al.* Understanding the genetic determinants of the brain with
422 MOSTest. *Nat. Commun.* **11**, 1–9 (2020).
423 48. Van Der Meer, D. *et al.* The genetic architecture of human cortical folding. *Sci. Adv.* **7**,
424 eabj9446 (2021).
425 49. Shadrin, A. A. *et al.* Vertex-wise multivariate genome-wide association study identifies

780 unique genetic loci associated with cortical morphology. *Neuroimage* **244**, 118603 (2021).

50. Fabregat, A. *et al.* Reactome pathway analysis: a high-performance in-memory approach. *BMC Bioinformatics* **18**, 1–9 (2017).

51. Fuchsberger, C. *et al.* The genetic architecture of type 2 diabetes. *Nature* **536**, 41–47 (2016).

52. Nikpay, M. *et al.* A comprehensive 1000 Genomes–based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121–1130 (2015).

53. Malik, R. *et al.* Low-frequency and common genetic variation in ischemic stroke: The METASTROKE collaboration. *Neurology* **86**, 1217–1226 (2016).

54. Wightman, D. P. *et al.* A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. *Nat. Genet.* **53**, 1276–1282 (2021).

55. Howard, D. M. *et al.* Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.* **22**, 343–352 (2019).

56. Ripke, S., Walters, J. T. R. & O'Donovan, M. C. Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. *medRxiv* 2020.09.12.20192922 (2020). doi:10.1101/2020.09.12.20192922

57. Watson, H. J. *et al.* Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. *Nat. Genet.* **51**, 1207–1214 (2019).

58. Brochu, M., Poehlman, E. T. & Ades, P. A. Obesity, body fat distribution, and coronary artery disease. *J. Cardiopulm. Rehabil. Prev.* **20**, 96–108 (2000).

449 59. Dharmalingam, M. & Yamasandhi, P. G. Nonalcoholic Fatty Liver Disease and Type 2
450 Diabetes Mellitus. *Indian J. Endocrinol. Metab.* **22**, 421–428 (2018).
451 60. Liver, E. A. for the S. of T. & (EASD, E. A. for the S. of D. EASL-EASD-EASO Clinical
452 Practice Guidelines for the management of non-alcoholic fatty liver disease. *Obes. Facts*
453 **9**, 65–90 (2016).
454 61. Bianco, C., Tavaglione, F., Romeo, S. & Valenti, L. Genetic risk scores and
455 personalization of care in fatty liver disease. *Curr. Opin. Pharmacol.* **61**, 6–11 (2021).
456 62. De Vincentis, A. *et al.* Metabolic and genetic determinants for progression to severe liver
457 disease in subjects with obesity from the UK Biobank. *Int. J. Obes. (Lond).* 1–8 (2021).
458 doi:10.1038/s41366-021-01015-w
459 63. Hamrick, M. W., McGee-Lawrence, M. E. & Frechette, D. M. Fatty infiltration of skeletal
460 muscle: mechanisms and comparisons with bone marrow adiposity. *Front. Endocrinol.*
461 *(Lausanne).* **7**, 69 (2016).
462 64. Tuttle, L. J., Sinacore, D. R. & Mueller, M. J. Intermuscular adipose tissue is muscle
463 specific and associated with poor functional performance. *J. Aging Res.* **2012**, (2012).
464 65. Gueugneau, M. *et al.* Skeletal muscle lipid content and oxidative activity in relation to
465 muscle fiber type in aging and metabolic syndrome. *Journals Gerontol. Ser. A Biomed.*
466 *Sci. Med. Sci.* **70**, 566–576 (2015).
467 66. Miljkovic, I. *et al.* Greater skeletal muscle fat infiltration is associated with higher all-
468 cause and cardiovascular mortality in older men. *Journals Gerontol. Ser. A Biomed. Sci.*
469 *Med. Sci.* **70**, 1133–1140 (2015).
470 67. Chakravarthy, M. V, Siddiqui, M. S., Forsgren, M. F. & Sanyal, A. J. Harnessing muscle–
471 liver crosstalk to treat nonalcoholic steatohepatitis. *Front. Endocrinol. (Lausanne).* 982

(2020).

68. Linge, J., Ekstedt, M. & Leinhard, O. D. Adverse muscle composition is linked to poor functional performance and metabolic comorbidities in NAFLD. *JHEP Reports* **3**, 100197 (2021).

69. Nachit, M. *et al.* Muscle fat content is strongly associated with NASH: a longitudinal study in patients with morbid obesity. *J. Hepatol.* **75**, 292–301 (2021).

70. Rivas, D. A. *et al.* Diminished anabolic signaling response to insulin induced by intramuscular lipid accumulation is associated with inflammation in aging but not obesity. *Am. J. Physiol. Integr. Comp. Physiol.* **310**, R561–R569 (2016).

71. Argilés, J. M., López-Soriano, J., Almendro, V., Busquets, S. & López-Soriano, F. J. Cross-talk between skeletal muscle and adipose tissue: a link with obesity? *Med. Res. Rev.* **25**, 49–65 (2005).

72. Francisco, V. *et al.* Obesity, Fat Mass and Immune System: Role for Leptin . *Frontiers in Physiology* **9**, 640 (2018).

73. Baumgartner, R. N., Heymsfield, S. B. & Roche, A. F. Human body composition and the epidemiology of chronic disease. *Obes. Res.* **3**, 73–95 (1995).

74. Poulsen, P., Vaag, A., Kyvik, K. & Beck-Nielsen, H. Genetic versus environmental aetiology of the metabolic syndrome among male and female twins. *Diabetologia* **44**, 537–543 (2001).

75. Shen, W. *et al.* Sexual dimorphism of adipose tissue distribution across the lifespan: a cross-sectional whole-body magnetic resonance imaging study. *Nutr. Metab. (Lond).* **6**, 17 (2009).

76. 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 (2015).

77. Miller, K. L. *et al.* Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat. Neurosci.* **19**, 1523–1536 (2016).

78. West, J. *et al.* Feasibility of MR-based body composition analysis in large scale population studies. *PLoS One* **11**, e0163332 (2016).

79. Beasley, T. M., Erickson, S. & Allison, D. B. Rank-based inverse normal transformations are increasingly used, but are they merited? *Behav. Genet.* **39**, 580–595 (2009).

80. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).

81. Watanabe, K., Taskesen, E., Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).

Acknowledgements

This work was partly performed on the TSD (Tjeneste for Sensitive Data) facilities, owned by the University of Oslo, operated and developed by the TSD service group at the University of Oslo, IT-Department (USIT). (tsd-drift@usit.uio.no). Computations were also performed on resources provided by UNINETT Sigma2 - the National Infrastructure for High-Performance Computing and Data Storage in Norway.

Materials & Correspondence

The data incorporated in this work were gathered from public resources. The code is available via https://github.com/precimed/mostest (GPLv3 license), and GWAS summary statistics are uploaded to the GWAS catalog (https://www.ebi.ac.uk/gwas/). Correspondence and requests for materials should be addressed to d.v.d.meer@medisin.uio.no

Methods

Participants

We made use of data from participants of the UKB population cohort, obtained from the data repository under accession number 27412. The composition, set-up, and data gathering protocols of the UKB have been extensively described elsewhere[76]. For the primary analyses, we selected White Europeans that had undergone the body MRI protocol, with available genetic and complete covariate data (N=34,036, mean age 64.5 years (SD=7.4), 51.5 % female). For the replication analyses, we made use of data from non-White Europeans (N=5081, mean age 63.0 years (SD=7.7), 52.9 % female).

Data collection and pre-processing

Body and liver MRI scans were collected from three scanning sites throughout the United Kingdom, all with identical scanners and protocols. They were acquired on 1.5T Siemens MAGNETOM Aera scanners using a body dual-echo Dixon Vibe protocol and a single-slice multi-echo gradient Dixon acquisition, respectively. The UKB core neuroimaging team has published extensive information on the applied scanning protocols and procedures, which we refer to for more details.[77] We acquired the data as processed by AMRA (Linköping, Sweden; https://www.amramedical.com), subsequently released by UKB. We bridged with UKB project accession #6569 to obtain early access to this data, which was then obtained from the UKB data repositories and stored locally at the secure computing cluster of the University of Oslo.

The AMRA body MRI processing included intensity inhomogeneity correction, non-rigid registration of atlases to acquired image volumes, quantification of fat and muscle composition using a voting scheme, and visual inspection for segmentation accuracy and manual adjustment.[78]

The liver MRI data were processed using a magnitude-based chemical shift technique with a 6-peak lipid model and then registered to the body MRI data and corrected for liver T2* to obtain a T1-weighted measure of liver proton density fat fraction (PDFF). AMRA implements manual quality control of the image/segmentation quality.

Measurement protocols and definitions

We extracted a selection of body composition measures (Table 1; see also UKB online documentation (http://biobank.ctsu.ox.ac.uk/showcase/)). Specifically, we extracted the following measures of adipose tissue: visceral adipose tissue (VAT), defined as the adipose tissue within the abdominal cavity, and abdominal subcutaneous adipose tissue (ASAT), defined as the adipose tissue between the top of the femoral head and the top of T9. We also extracted measures of muscle fat infiltration (MFI) derived from the anterior and posterior thighs, averaged over both legs, and liver proton density fat fraction (PDFF). As measures of muscle tissue, we included anterior and posterior thigh fat-free muscle volume (ATMV and PTMV), and total thigh muscle volume, encompassing the gluteus, iliacus, adductors, hamstrings, quadriceps femoris, and sartorius, normalized to a z-score (TTMVz) that corrects for BMI, age, sex and height.[36] We extracted two ratios from the UKB repository, namely weight-to-muscle ratio (WMR), defined as weight/TTMV, and abdominal fat ratio (AFR), which is (VAT+ASAT)/(VAT+ASAT+TTMV). Additionally, for VAT, ASAT, ATMV, and PTMV, we computed index measures by dividing these measures by the squared standing height in meters (e.g., ASATi is ASAT/height^2). This is done since weight, adipose tissue, and lean tissue compartments scale to approximate height squared.

We subsequently regressed out age, sex, scanner site, and the first twenty genetic principal components from each measure. Following this, we applied rank-based inverse normal

564 transformation[79] to the residuals of each measure, leading to normally distributed measures as
565 input for the GWAS.

566 For the secondary analyses, comparing the set of MRI-derived measures of body composition
567 to measures of cardiometabolic health, we included 21 measures available in the UKB as listed in
568 Table 2.

569 *Table 2. Measures of cardiometabolic health used in the secondary analyses, together with*
570 *abbreviations and available sample sizes.*

Measure	*Abbreviation*	*N*
Cholesterol		389832
High-density lipoproteins	HDL	356802
Low-density lipoproteins	LDL	389115
Triglycerides		389524
Apolipoprotein A		354829
Apolipoprotein B		387945
Cholesterol to HDL		356734
ApoB to ApoA		353056
C-Reactive Protein	CRP	388993
Glucose		356558
Glycated hemoglobin	HbA1c	389728
Alanine aminotransferase	ALT	389701
Aspartate aminotransferase	AST	388406
Gamma-glutamyl transferase	GGT	389634
Creatinine		389637
Body mass index	BMI	407558
Waist circumference		408179
Hip circumference		408134
Waist-to-hip ratio	WHR	408096
Diastolic blood pressure	DBP	36591
Systolic blood pressure	SBP	36591

571

572 *GWAS procedure*

573 We made use of the UKB v3 imputed data, which has undergone extensive quality control
574 procedures as described by the UKB genetics team.[80] After converting the BGEN format to PLINK

binary format, we additionally carried out standard quality check procedures, including filtering out individuals with more than 10% missingness, SNPs with more than 5% missingness, and SNPs failing the Hardy-Weinberg equilibrium test at $p=1*10^{-9}$. We further set a minor allele frequency threshold of 0.005, leaving 9,061,022 SNPs.

We carried out GWAS through the freely available MOSTest software (https://github.com/precimed/mostest). Details about the procedure and its extensive validation have been described previously.[47] GWAS on each of the pre-residualized and normalized measures were carried out using the standard additive model of linear association between genotype vector, g_j, and phenotype vector, y. Independent significant SNPs and genomic loci were identified in accordance with the PGC locus definition, as also used in FUMA SNP2GENE.[81] First, we selected a subset of SNPs that passed genome-wide significance threshold $5x10^{-8}$, and used PLINK to perform a clumping procedure at LD r2=0.6 to identify the list of independent significant SNPs. Second, we clumped the list of independent significant SNPs at LD r2=0.1 threshold to identify lead SNPs. Third, we queried the reference panel for all candidate SNPs in LD r2 of 0.1 or higher with any lead SNPs. Further, for each lead SNP, its corresponding genomic loci were defined as a contiguous region of the lead SNPs' chromosome, containing all candidate SNPs in r2=0.1 or higher LD with the lead SNP. Finally, adjacent genomic loci were merged if separated by less than 250 KB. Allele LD correlations were computed from EUR population of the 1000 genomes Phase 3 data. We made use of the Functional Mapping and Annotation of GWAS (FUMA) online platform (https://fuma.ctglab.nl/) to map significant SNPs from the MOSTest analyses to genes.

MiXeR analysis

We applied a causal mixture model[43,44] to estimate the percentage of variance explained by genome-wide significant SNPs as a function of sample size. For each SNP, i, MiXeR models its additive genetic effect of allele substitution, β_i, as a point-normal mixture, $\beta_i = (1-\pi_1)N(0,0) + \pi_1 N(0,\sigma_\beta^2)$, where π_1 represents the proportion of non-null SNPs (`polygenicity`) and σ_β^2 represents the variance of effect sizes of non-null SNPs (`discoverability`). Then, for each SNP, j, MiXeR incorporates LD information and allele frequencies for 9,997,231 SNPs extracted from 1000 Genomes Phase3 data to estimate the expected probability distribution of the signed test statistic, $z_j = \delta_j + \epsilon_j = N\sum_i \sqrt{H_i} r_{ij}\beta_i + \epsilon_j$, where N is the sample size, H_i indicates heterozygosity of i-th SNP, r_{ij} indicates an allelic correlation between i-th and j-th SNPs, and $\epsilon_j \sim N(0,\sigma_0^2)$ is the residual variance. Further, the three parameters, $\pi_1, \sigma_\beta^2, \sigma_0^2$, are fitted by direct maximization of the likelihood function. Fitting the univariate MiXeR model does not depend on the sign of z_j, allowing us to calculate $|z_j|$ from MOSTest p-values. Finally, given the estimated parameters of the model, the power curve $S(N)$ is then calculated from the posterior distribution $p(\delta_j|z_j, N)$.

Gene-set analyses

We carried out gene-based analyses using MAGMA v1.08 with default settings, which entails the application of a SNP-wide mean model and the use of the 1000 Genomes Phase 3 EUR reference panel. Gene-set analyses were done in a similar manner, restricting the sets under investigation to those that are part of the Gene Ontology biological processes subset (n=7522), as listed in the Molecular Signatures Database (MsigdB; c5.bp.v7.1).