Genome-wide association study of a lipedema phenotype among women in the UK Biobank identifies multiple genetic risk factors

Short Title: Genome-wide association study of lipedema

Yann C. Klimentidis [1,2*], Zhao Chen [1], Manuel L. Gonzalez-Garay [3], Dionysios Grigoriadis [4], Ege Sackey [4], Alan Pittman [4], Pia Ostergaard [4], Karen L. Herbst [5,6]

[1] Department of Epidemiology and Biostatistics, Mel and Enid Zuckerman College of Public Health, University of Arizona, Tucson, AZ, USA

[2] BIO5 Institute, University of Arizona, AZ, USA

[3] College of Medicine, University of Arizona, Tucson, AZ, USA

[4] Genetics Research Centre, Molecular and Clinical Sciences Institute, St George's University of London, London, UK.

[5] TREAT Program, College of Medicine, University of Arizona, Tucson, AZ, USA

[6] Total Lipedema Care, 240 South La Cienega Blvd, Beverly Hills, CA, 90211

*** Corresponding Author:** Department of Epidemiology and Biostatistics, Mel and Enid Zuckerman College of Public Health, University of Arizona, Tucson, Arizona, 85724, 520-621-0147, yann@email.arizona.edu

ABSTRACT

Lipedema is a common disorder characterized by excessive deposition of subcutaneous adipose tissue (SAT) in the legs, hips, and buttocks, mainly occurring in adult women. Although it appears to be heritable, no specific genes have yet been identified. To identify potential genetic risk factors for lipedema, we used bioelectrical impedance analysis and anthropometric data from the UK Biobank to identify women with and without a lipedema phenotype. Specifically, we identified women with both a high percentage of fat in the lower limbs and a relatively small waist, adjusting for hip circumference. We performed a genome-wide association study (GWAS) for this phenotype, and performed multiple sensitivity GWAS. In an independent case/control lipedema study, we attempted to replicate our top hits. We identified 18 significant loci ($p<5 \times 10^{-9}$), several of which have previously been identified in GWAS of waist-to-hip ratio with larger effects in women. Two loci (*VEGFA* and *GRB14-COBLL1*) were significantly associated with lipedema in the independent replication study. Follow-up analyses suggest an enrichment of genes expressed in blood vessels and adipose tissue, among other tissues. Our findings provide a starting point towards better understanding the genetic and physiological basis for lipedema.

Keywords: lipedema, lipoedema, genetic, genome-wide association study, body fat distribution

INTRODUCTION

Lipedema has been recognized since the 1940s as a condition occurring mainly in women and is characterized by the bilateral enlargement of the lower limbs which in many instances exaggerate the female form [1]. It can be debilitating with respect to resulting pain and impaired mobility [2].

There are very few prevalence estimates for lipedema. A prevalence of 11% was observed in a sample of women from a lymphedema clinic in Germany [3], and a prevalence of 9.7% was observed in a small sample of professional German women [4]. Family history has been reported, suggesting that lipedema is heritable [1,5,6]. Further evidence supporting a genetic contribution is that lipedema subcutaneous adipose tissue (SAT) is highly resistant to lifestyle changes such as diet and exercise, occurs after menarche, and occurs nearly exclusively in women. These lines of evidence suggest that environmental (non-genetic) factors may not play a predominant role.

Although lipedema does appear to be heritable, individual genetic loci associated with it have yet to be found. Identifying such loci will provide much-needed insight into the pathophysiology of lipedema, and potentially enable the development of therapeutics, and the identification of individuals at high risk. However, since lipedema is rarely diagnosed, and is not widely recognized, there are currently no large collections of diagnosed lipedema cases with whom genetic risk factors could be identified.

Here, we used body fat percentage and anthropometric measurements from the UK Biobank to classify women into those that appear to have a lipedema phenotype, and those without, in an effort to identify genetic risk factors for lipedema. Given the lack of evidence thus

far for a single highly penetrant gene for lipedema [5], we hypothesized that multiple individual genetic variants across the genome are associated with the lipedema phenotype among adult women.

METHODS

UK Biobank

The UK Biobank is a prospective cohort study of approximately 500,000 adults living in the UK, aged 39 to 70 [7,8]. Participants were measured for a variety of traits and diseases from 2006 to 2010 at 22 centers across the UK. Only women were included in this study, and to minimize the potential confounding effects of ancestry, only women self-identifying European ancestry were included. All participants gave written informed consent, and ethical approval was obtained from the North West Multicentre Research Ethics Committee, the National Information Governance Board for Health & Social Care, and the Community Health Index Advisory Group.

Phenotypic measurements

We used bioelectrical impedance analysis (BIA) data obtained using a Tanita BC418MA segmental body composition analyzer. Women who were pregnant, an amputee, wheelchair bound, unable to stand, or using a pacemaker were not measured (i.e. excluded). We excluded women whose left leg fat % was 30% greater or lesser than their right leg fat %, to exclude those with obvious uni-lateral enlargement of the legs, based on a scatterplot visualization. We averaged the left and right leg fat percentages and used this average in all subsequent analyses. The UK Biobank specify that hip and waist circumferences measurements were obtained in

centimeters (cm) with a Seca 200 cm tape. The tape was first passed around the smallest part of the trunk (natural indent). If no natural indent could be found, the waist measurement was taken at the level of the umbilicus. The hip circumference was measured at the widest part of the hips.

Lipedema definition

We defined lipedema first as having a relatively high leg fat % along with a relatively small waist circumference. We first obtained the residuals from a linear regression in which the average of the left and right leg fat percentages was the outcome, and the independent variables were height, hip circumference, age, and recruitment center. These covariates were included as most leg fat is likely to be located on the upper end of the leg (by the hip), and it is therefore important to take into account the anthropometric characteristics of hip size and overall body size in order to minimize confounding from these characteristics. Furthermore, since lipedema usually involves high levels of fat deposition throughout the leg, and not only at the hip, these adjustments helps us to identify women with a high fat percentage throughout the leg, and not just at the hip level. We then obtained the residuals from a linear regression in which waist circumference was the outcome, and the independent variables were the same as those listed above for leg fat percentage. Women were deemed to be lipedema cases if they were in the >65th percentile of the leg fat percentage residual and in the <51.3rd percentile of the waist circumference residual (these cutoffs were chosen to achieve a prevalence of 10%). Controls were individuals with <=45th leg fat residual percentile residual or those in the >=71.3rd waist circumference residual percentile. The resulting "buffer zone" of individuals was to exclude potential false positives or false negatives. We also considered as a sensitivity analysis a lipedema phenotype at an assumed 5% prevalence. In this case, women were deemed to be

lipedema cases if they were in the >75th percentile of the leg fat percentage residual and in the <46.7th percentile of the waist circumference residual, and controls were individuals with <=65th leg fat residual percentile residual or those in the >=56.7th waist circumference residual percentile.

Leg pain

Since leg pain is an associated feature of lipedema, we used responses to the following question asked at the baseline exam at each UK Biobank recruitment center: "Do you get a pain in either leg on walking?" This question was administered to a subset (~40%) of UK Biobank participants, as it was introduced by the UK Biobank towards the end of recruitment. A lack of an affirmative response to this question was required to define a patient as a control.

Genetic data

The vast majority of UK Biobank participants were genotyped with the Affymetrix UK Biobank Axiom Array (Santa Clara, USA). Approximately 10% of participants were genotyped with the Affymetrix UK BiLEVE Axiom Array. Tens of millions of additional SNP genotypes were obtained through imputation using the Haplotype Reference Consortium [9] and UK10K [10] haplotype data as references. Principal component analysis was performed by the UK Biobank team, using fastPCA software on a set of 147,604 high-quality directly genotyped markers. Individuals with an unusually high heterozygosity rate, a >5% missing rate, or a mismatch between self-reported and genetically-inferred sex were excluded. These and other details regarding the genotyping, imputation, and QC procedures are available elsewhere [8].

SNPs not in Hardy-Weinberg equilibrium ($p<1 \times 10^{-6}$), with a high missingness (>1.5%), a low minor allele frequency (<0.1%), or low imputation quality (info<0.4) were excluded. A total of 16.8 million SNPs were available for analysis, including those on the X chromosome.

Statistical analyses

We tested the association between each SNP with case-control status using linear regression implemented in the BOLT-LMM software [11,12], which implements a linear mixed model regression, including SNPs other than the one tested as random effects, and thereby correcting for population stratification and relatedness. Since BOLT-LMM implements a linear regression model, the effect size estimates for case-control outcomes are unreliable. As previously done in other studies [13,14], we therefore estimated the effect sizes of the genome-wide significant ($p < 5 \times 10^{-9}$) SNPs with logistic regression in R [15]. We included age, hip circumference, recruitment center, genotyping platform, and the first 10 principal components (PCs) as covariates in both the BOLT-LMM and the logistic regression SNP association models. Hip circumference was included to account for overall body size and for proportion of fat in thighs. We performed multiple sensitivity analyses. First, we considered a model without any adjustment for hip circumference. Since there is considerable uncertainty regarding the actual prevalence of lipedema, we also considered a prevalence of 5% instead of 10%. We also performed an analysis only in women with BMI <30 to avoid obesity complicating the analyses. Finally, we considered analyses in which we added into our case definition an affirmative response regarding self-reported leg pain. SNPs were considered genome-wide significant if $p < 5 \times 10^{-9}$. This threshold was chosen instead of the more typical threshold because we performed multiple GWAS analyses, and to optimize our ability to replicate findings in our much smaller

150 replication dataset. To estimate SNP-based heritability, and to estimate genetic correlations of
151 our lipedema phenotype with a wide range of other human traits and diseases, including other
152 body composition and cardiometabolic traits and diseases, we used the online implementation
153 (http://ldsc.broadinstitute.org/) of the LD score regression method [16–18]. Details of each
154 phenotype and GWAS used in these genetic correlations can be found on the aforementioned
155 website. We examined the LD-score regression intercept, as well as the Q-Q plot, to assess
156 genomic inflation. We tested the association of each of the significant SNPs identified with
157 potentially relevant anthropometric measures among all UK Biobank women of European
158 ancestry, using linear regression, including age, chip, 10 PCs, and center as covariates.

159 Gene expression enrichment patterns across different tissues were examined through the
160 web-based platform, Functional Mapping and Annotation of Genome-Wide Association Studies
161 (FUMA GWAS) [19], that uses data from GTEx v7 [20], and the MAGMA gene-based analysis
162 for identification of associated genes [21]. Briefly, the MAGMA gene-based analysis uses all the
163 SNPs in a gene as the unit of analysis to test the association of each gene across the genome with
164 the phenotype. A genome-wide significance threshold of $p < 2.6 \times 10^{-6}$ was used. We also used
165 the FUMA GWAS platform to identify eQTL from GTEx v7 and from a large blood eQTL study
166 [22], by interrogating all top SNPs, and all SNPs in LD ($r^2 > 0.6$) of top SNPs.

167 Pathway over-representation analysis was performed for the unique genes identified by
168 the aforementioned eQTL analysis, using the SOAP/WSDL interface of the
169 ConsensusPathwayDB-human. ConsensusPathwayDB is a project of the Max Planck Institute for
170 Molecular Genetics that integrates multiple interaction networks [23,24].

171

Replication study

We sought to replicate our top findings in a case/control study of clinically diagnosed lipedema cases, as previously described [25]. Briefly, 130 lipedema cases recruited from two specialist UK clinics at St George's University Hospital NHS Trust and the University Hospitals of Derby and Burton NHS Trust were genotyped with Illumina Infinium microarrays. Unaffected females (N=5,848) enrolled in the Understanding Society UK study [26] genotyped with Illumina HumanCoreExome-12 (v1.0) were selected as the control group to the replication cohort controls (European Genome-phenome Archive ID: EGAD00010000890). Details of the 'UK Lipoedema' cohort genotyping can be found elsewhere [25]. Imputation was performed by aligning variants to the 1,000 Genomes [27] reference and the normalized variants were imputed using the Michigan Imputation server [28]. Post imputation quality controls were used to remove low-quality ($r^2 \leq 0.8$) imputed variants before further analyses. The association analysis was performed using a univariate linear mixed model, implemented in GEMMA software (version 0.98.1) [29]. The p-value distribution was assessed using a Quantile-Quantile (Q-Q) plot, and there was no inflation effect observed on the association analysis. Given the 14 statistically significant top hits that we were able to test in the replication study, a significant replication was determined based on a p-value<0.0036, according to a Bonferroni correction.

RESULTS

GWAS

In this study, we defined lipedema cases and controls based on leg fat mass and waist circumference. The characteristics of the different case (n=24,450) and control (n=165,227)

194 groups are shown in Table 1. The inferred lipedema cases tend to have a higher leg fat % and a
195 lower waist-to-hip ratio (WHR).

196 Cases are shown in green in Figure 1 with respect to unadjusted waist circumference and
197 unadjusted leg fat %, and in Supplementary Figure 1 with respect to the residualized values of
198 these variables that were used to determine cases and controls. The LD-score regression intercept
199 (<1.02) and Q-Q plot (Supplementary Figure 2) from the genome-wide association study suggest
200 little evidence of systematic inflation of effect sizes, beyond the polygenic signal associated with
201 the lipedema phenotype. SNP-based heritability for our main model was 5.13% (see
202 Supplementary Table 1).

203 We find 18 loci significantly associated with the primary lipedema phenotype as defined
204 in this study (see Figure 2 and Table 2). Associations of these 18 significant SNPs with relevant
205 anthropometric traits among all UK Biobank European-ancestry women are shown in
206 Supplementary Table 2. In analyses not adjusting for hip circumference, the results are relatively
207 similar, with generally smaller magnitudes of association (see Supplementary Table 3 and
208 Supplementary Figure 3). The same is observed for the other sensitivity analyses (see
209 Supplementary Figures 4-5). The addition of leg pain as a criterion for cases (n=1,724 cases;
210 165,227 controls) resulted in no genome-wide significant loci (Supplementary Figure 6). Only
211 the *LYPLAL1* locus exhibited a 'suggestive level' of significance ($p=4.5 \times 10^{-6}$, see
212 Supplementary Table 4). The gene-based analysis, in which the collection of SNPs in each gene
213 is taken as the unit of analysis instead of each SNP individually, identified 72 genes associated
214 with lipedema (Supplementary Figure 7 and Supplementary Table 5).

215

Genetic correlations with other traits and diseases

We observed significant positive genetic correlations of the lipedema phenotype with body fat and leptin levels (Supplementary Figure 8). We also find nominally significant positive genetic correlations with hip circumference, primary biliary cirrhosis and BMI. We find nominally significant negative genetic correlations with WHR, age at first birth, forced vital capacity, birth weight, and age at menopause (Supplementary Figure 8)

Associations of loci with tissue-specific gene expression levels

Genes implicated by eQTL analysis, and which may provide insight into the genetic regulation of the lipedema phenotype, vary by locus and by tissue. For example, the *RSPO3* top SNP, rs72959041, is associated with the expression level of the *RSPO3* gene in subcutaneous adipose tissue, as are *ZNF664* with the rs11057418 SNP, and *TIPARP* with the rs4680338 SNP (Supplementary Table 6). The top SNP at the *GRB14-COBLL1* locus is associated with expression of several genes in various tissues. A list of these gene expression patterns for the significant (FDR<0.05) eQTLs of the top SNPs can be found in Supplementary Table 6. Upon expanding this analysis to all SNPs in LD (r^2>0.6) with the top SNPs, a more extensive list of genes and tissues was identified (Supplementary Table 7). A list of genes sorted by relevant and highly represented tissues is shown in Supplementary Table 8. Tissue enrichment analyses based on the gene-based (MAGMA) analysis identify arterial blood vessels as the main tissues where these genes may exert their actions (Supplementary Figure 9).

Pathway analysis

Our pathway analysis results show significant enrichment for multiple pathways and Gene Ontology Terms (Supplementary Table 9). Noteworthy are: a) the EGFR1 signaling pathway, a pathway that induces growth, differentiation, migration, adhesion and cell survival through multiple hormone interactions [30]; and b) GO:0003785 actin monomer binding, a regulator of actin cytoskeleton dynamics in cells [31].

Replication

In a replication cohort of phenotyped lipedema patients from the 'UK Lipoedema' study, we were able to obtain results for 14 out of 18 loci (see Table 2). We identified two statistically significant associations also showing the same direction of effect as in UK Biobank: *VEGFA* ($p=5.0 \times 10^{-4}$) and *GRB14-COBLL1* ($p=2.3 \times 10^{-3}$). Two nominally significant loci ($p<0.05$; *ADAMTS9* and *LYPLAL1*) were also directionally consistent.

DISCUSSION

Using an inferred lipedema phenotype based on high leg fat % and small waist in the UK Biobank, we performed the first GWAS of a lipedema phenotype and identified 18 loci across the genome. Two of these loci (*VEGFA* and *GRB14-COBLL1*) were significantly associated with lipedema in an independent case-control study including clinically diagnosed lipedema cases.

Loci in/near *RSPO3*, *GRB14-COBLL1*, *ZNF664-FAM101A* (near *CCDC92*), *VEGFA*, *ADAMTS9*, *LYPLAL1*, *ANKRD55-MAP3K1* have previously been found to be associated with WHR, and importantly, to exhibit stronger effects in women than in men [32–34]. Importantly,

we show that these loci are associated with leg fat % independently of hip circumference, indicating that these associations are not predominately driven by hip circumference or fat around the hip, but rather fat throughout the lower limbs.

Pain in the lower limbs is a common complaint among lipedema patients [2] and when including pain as a criterion in the GWAS, only one of the loci, rs749853052 near *LYPLAL1*, exhibited a 'suggestive level' of significance. We strongly suspect that the reduced sample size when incorporating leg pain information resulted in reduced statistical power to detect significantly associated loci. The *LYPLAL1* variant is over 250 kb downstream of the *LYPLAL1* (lysophospholipase-like 1) gene. The function of this locus is still unknown, although it may act as a triglyceride lipase or lysophospholipase [35,36]. This locus has been associated with WHR, with a stronger effect in women than in men [32,37,38]. It has also been associated with a "favorable adiposity" or gynoid phenotype [39], and with non-alcoholic fatty liver disease [40]. It has also been found to be more abundantly expressed in the subcutaneous adipose tissue of obese compared to lean individuals [41], potentially making it an interesting association to explore in the context of lipedema.

One of the novel loci identified on chromosome 5 is located in the *LINC01184* non-coding gene, and just upstream of the *SLC12A2* gene. Our eQTL analysis revealed that the top SNP at this locus is associated with increased expression of the fibrillin 2 (*FBN2)* gene in the thyroid. The *FBN2* gene is located downstream of the *SLC12A2* gene. This locus was also identified in a GWAS of varicose veins of the lower extremities [42]. The allele associated with decreased risk of varicose veins is in strong LD (r^2>0.93) with the allele associated with increased odds for the lipedema phenotype, suggesting a potential connection between lipedema and varicose veins.

We identified an intronic variant in the *ADAMTSL3* gene which codes for a glycoprotein. ADAMTSL3 localize to the extra-cellular matrix (ECM) [43] where it may modulate the ADAMTS proteinases [44]. ADAMTS proteins are involved in ECM or cell-matrix interactions [44]. As the EMC is important for the regulation of adipocyte expansion and proliferation [45], ADAMTS and ADAMTS-like protein could be important in that process. Interestingly, *ADAMTSL3* has been associated with overall body fat [46] and lean body mass [47].

Finally, another locus worth highlighting is at the *DNAH10-CCDC92-ZNF664* locus. Knockdown of both *DNAH10* and *CCDC92* has previously been shown to result in lowered mRNA levels of the respective gene, and in reduced lipid accumulation in mouse adipocytes, consistent with an impairment of lipid accumulation in peripheral adipose tissues in humans [48]. This locus was also implicated in abnormally high HDL-C levels [49], and in large HDL particles [50], further suggesting the involvement of this locus in adipose tissue growth and its consequences.

Of course, for any of the above, the question remains if the inferred lipedema phenotype has any resemblance to a clinically defined lipedema cohort. When validating 14 SNPs in the 'UK Lipoedema' cohort [25], the *VEGFA* and *GRB14-COBLL1* loci were significantly associated with lipedema. In addition to the well-established association of these loci with WHR [34], they have also been associated with other cardiometabolic traits and diseases [51,52], as well as with a favorable pattern of adiposity [39,53]. Interestingly, other loci identified with favorable adiposity overlap with some of those identified here, such as *ANKRD55/MAP3KI*, *DNAH10/CCDC92*, and *FAM101A* [39,53]. It has been suggested previously that despite higher BMI, lipedema patients have relatively lower risk of type 2 diabetes [54] and the gynoid SAT may protect against cardiovascular risk [55].

Among the strengths of our study are the large sample size that enabled our genetic investigation and identification of associated loci. Although there are other studies of lipedema, they have much smaller samples, and thereby do not have sufficient power to detect the tiny effect sizes that characterize the genetic architecture of most traits that do not have a single-gene cause. Another strength is the availability of relatively detailed body composition measures. Given the small effect sizes typically observed for the genetics of complex traits and diseases, a large sample size increased our probability of detecting loci associated with the lipedema phenotype, at the cost of the quality of the phenotype (i.e. absence of a lipedema diagnosis, see next paragraph). A major limitation of this study is that we could not rely on an actual diagnosis or on a validated classifier of lipedema. Since the recognition and diagnosis of lipedema is in its infancy, and is still very limited, it is currently difficult to obtain large collections of genotyped women with diagnosed lipedema. The binary classifier that we used likely mis-classified a number of lipedema cases and controls. However, the larger sample size and the availability of a replication cohort consisting of diagnosed lipedema cases likely counter-balanced some of the resulting loss in power from the discovery cohort. It is also possible that the 16 loci identified in the UK Biobank discovery cohort that were not replicated in the replication cohort are loci that are not associated with lipedema, but rather with other associated aspects of body shape that are similar to, but not related, to lipedema. In addition, it could be that our replication cohort is still too small to successfully replicate most of the loci, and that only through the study of further cohorts could we confirm the potential role of these loci in lipedema. Another limitation of our study is that it is limited to people of European (mainly British) ancestry, and to individuals between the ages of 40 and 70, many years after lipedema typically initiates, and when the lipedema phenotype is more likely to be confounded by frank obesity.

In conclusion, we have identified 18 loci associated with an inferential lipedema phenotype in adult women of European descent, among which, 2 successfully replicate in a study of clinically diagnosed lipedema cases. Some have previously been identified as female-specific loci for WHR, while others have not previously been linked to body composition phenotypes. In a replication study with clinically diagnosed lipedema cases, we successfully replicate the *VEGFA* and *GRB14-COBLL1* loci, which have previously been associated with fat distribution patterns. We hope that these loci and the genes and tissues that they implicate will provide starting points towards a better understanding of the pathophysiology of lipedema, and eventually to treatment and prevention approaches.

ACKNOWLEDGMENTS

This research was conducted using the UK Biobank Resource under Application Number 15678. We thank the participants and organizers of the UK Biobank. YCK and PO would like to acknowledge support from the Lipedema Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

LDHUB Acknowledgements

We gratefully acknowledge all the studies and databases that made GWAS summary data available: ADIPOGen (Adiponectin genetics consortium), C4D (Coronary Artery Disease Genetics Consortium), CARDIoGRAM (Coronary ARtery DIsease Genome wide Replication

and Meta-analysis), CKDGen (Chronic Kidney Disease Genetics consortium), dbGAP (database of Genotypes and Phenotypes), DIAGRAM (DIAbetes Genetics Replication And Meta-analysis), ENIGMA (Enhancing Neuro Imaging Genetics through Meta Analysis), EAGLE (EArly Genetics & Lifecourse Epidemiology Eczema Consortium, excluding 23andMe), EGG (Early Growth Genetics Consortium), GABRIEL (A Multidisciplinary Study to Identify the Genetic and Environmental Causes of Asthma in the European Community), GCAN (Genetic Consortium for Anorexia Nervosa), GEFOS (GEnetic Factors for OSteoporosis Consortium), GIANT (Genetic Investigation of ANthropometric Traits), GIS (Genetics of Iron Status consortium), GLGC (Global Lipids Genetics Consortium), GPC (Genetics of Personality Consortium), GUGC (Global Urate and Gout consortium), HaemGen (haemotological and platelet traits genetics consortium), HRgene (Heart Rate consortium), IIBDGC (International Inflammatory Bowel Disease Genetics Consortium), ILCCO (International Lung Cancer Consortium), IMSGC (International Multiple Sclerosis Genetic Consortium), MAGIC (Meta-Analyses of Glucose and Insulin-related traits Consortium), MESA (Multi-Ethnic Study of Atherosclerosis), PGC (Psychiatric Genomics Consortium), Project MinE consortium, ReproGen (Reproductive Genetics Consortium), SSGAC (Social Science Genetics Association Consortium) and TAG (Tobacco and Genetics Consortium), TRICL (Transdisciplinary Research in Cancer of the Lung consortium), UK Biobank. We gratefully acknowledge the contributions of Alkes Price (the systemic lupus erythematosus GWAS and primary biliary cirrhosis GWAS) and Johannes Kettunen (lipids metabolites GWAS).

Conflict of interest: The authors declare no conflict of interest.

REFERENCES

1. WOLD LE, HINES EAJ, ALLEN E V. Lipedema of the legs; a syndrome characterized by fat legs and edema. Ann Intern Med. 1951;34: 1243–1250.

2. Dudek JE, Białaszek W, Gabriel M. Quality of life, its factors, and sociodemographic characteristics of Polish women with lipedema. BMC Womens Health. 2021;21: 27. doi:10.1186/s12905-021-01174-y

3. Foldi E, Foldi M. Lipedema. In: Foldi M, Foldi E, editors. Foldi's Textbook of Lymphology. Munich, Germany: Elsevier GmbH; 2006. pp. 417–427.

4. Marshall M, C S-S. Prevalence of lipoedema in professional women in Germany. (Lipoedema-3-study). Phlebologie. 2011;40: 127–34.

5. Child AH, Gordon KD, Sharpe P, Brice G, Ostergaard P, Jeffery S, et al. Lipedema: An inherited condition. Am J Med Genet Part A. 2010;152: 970–976.

6. Forner-Cordero I, Szolnoky G, Forner-Cordero A, Kemény L. Lipedema: An overview of its clinical manifestations, diagnosis and treatment of the disproportional fatty deposition syndrome - systematic review. Clin Obes. 2012;2: 86–95. doi:10.1111/j.1758-8111.2012.00045.x

7. Collins R. What makes UK Biobank special? The Lancet. 2012. pp. 1173–1174.

8. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562: 203–209. doi:10.1038/s41586-018-0579-z

9. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference

panel of 64,976 haplotypes for genotype imputation. Nat Genet. 2016;48: 1279–1283. doi:10.1038/ng.3643

10. Consortium TU. The UK10K project identifies rare variants in health and disease. Nature. 2015;526: 82–90. Available: http://dx.doi.org/10.1038/nature14962

11. Loh P-R, Tucker G, Bulik-Sullivan BK, Vilhjalmsson BJ, Finucane HK, Salem RM, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nat Genet. 2015;47: 284–290. doi:10.1038/ng.3190

12. Loh P-R, Kichaev G, Gazal S, Schoech AP, Price AL. Mixed model association for biobank-scale data sets. bioRxiv. 2017. doi:10.1101/194944

13. Klarin D, Zhu QM, Emdin CA, Chaffin M, Horner S, McMillan BJ, et al. Genetic analysis in UK Biobank links insulin resistance and transendothelial migration pathways to coronary artery disease. Nat Genet. 2017;49: 1392–1397. doi:10.1038/ng.3914

14. Klarin D, Emdin CA, Natarajan P, Conrad MF, Kathiresan S. Genetic Analysis of Venous Thromboembolism in UK Biobank Identifies the ZFPM2 Locus and Implicates Obesity as a Causal Risk FactorCLINICAL PERSPECTIVE. Circ Cardiovasc Genet. 2017;10. doi:10.1161/CIRCGENETICS.116.001643

15. Team RDC. R: A language and environment for statistical computing. 2011.

16. Zheng J, Erzurumluoglu AM, Elsworth BL, Kemp JP, Howe L, Haycock PC, et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. Bioinformatics. 2017;33: 272. doi:10.1093/bioinformatics/btw613

413 17. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of
414 genetic correlations across human diseases and traits. Nat Genet. 2015;47: 1236–1241.
415 doi:10.1038/ng.3406

416 18. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Consortium SWG of the PG,
417 et al. LD Score regression distinguishes confounding from polygenicity in genome-wide
418 association studies. Nat Genet. 2015;47: 291–295. Available:
419 http://dx.doi.org/10.1038/ng.3211

420 19. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and
421 annotation of genetic associations with FUMA. Nat Commun. 2017;8: 1826.
422 doi:10.1038/s41467-017-01261-5

423 20. Consortium TGte. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue
424 gene regulation in humans. Sci . 2015;348: 648–660. doi:10.1126/science.1262110

425 21. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set
426 Analysis of GWAS Data. Tang H, editor. PLOS Comput Biol. 2015;11: e1004219.
427 doi:10.1371/journal.pcbi.1004219

428 22. Westra H-J, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al.
429 Systematic identification of trans eQTLs as putative drivers of known disease
430 associations. Nat Genet. 2013;45: 1238–1243. doi:10.1038/ng.2756

431 23. Kamburov A, Stelzl U, Lehrach H, Herwig R. The ConsensusPathDB interaction
432 database: 2013 update. Nucleic Acids Res. 2013;41: D793-800. doi:10.1093/nar/gks1055

433 24. Herwig R, Hardt C, Lienhard M, Kamburov A. Analyzing and interpreting genome data at

the network level with ConsensusPathDB. Nat Protoc. 2016;11: 1889–1907. doi:10.1038/nprot.2016.117

25. Grigoriadis D, Sackey E, Riches K, van Zanten M, Brice G, England R, et al. Investigation of clinical characteristics and genome associations in the UK Lipoedema cohort. medRxiv. 2021. doi:10.1101/2021.06.15.21258988

26. University of Essex. Institute for Social and Economic Research NSR. Understanding Society: Waves 2-3 Nurse Health Assessment, 2010-2012. [data collection]. 3rd Edition. UK Data Service. SN: 7251; 2014. Available: http://doi.org/10.5255/UKDA-SN-7251-3

27. Consortium GP, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491: 56–65. Available: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=23128226&retmode=ref&cmd=prlinks

28. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nat Genet. 2016;48: 1284–1287. doi:10.1038/ng.3656

29. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. 2012;44: 821–824. doi:10.1038/ng.2310

30. Zeng F, Harris RC. Epidermal growth factor, from gene organization to bedside. Semin Cell Dev Biol. 2014;28: 2–11. doi:10.1016/j.semcdb.2014.01.011

31. Rottner K, Faix J, Bogdan S, Linder S, Kerkhoff E. Actin assembly mechanisms at a

455 glance. J Cell Sci. 2017;130: 3427–3435. doi:10.1242/jcs.206433

456 32. Shungin D, Winkler T, Croteau-Chonka D, Ferreira T, Mägi R, Lakka T, et al. New
457 genetic loci link adipose and insulin biology to body fat distribution. Nature. 2015;518:
458 187–196. doi:10.1038/nature14132.New

459 33. Winkler TW, Justice AE, Graff M, Barata L, Feitosa MF, Chu S, et al. The Influence of
460 Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale
461 Genome-Wide Interaction Study. PLoS Genet. 2015;11: e1005378.
462 doi:10.1371/journal.pgen.1005378

463 34. Pulit SL, Stoneman C, Morris AP, Wood AR, Glastonbury CA, Tyrrell J, et al. Meta-
464 analysis of genome-wide association studies for body fat distribution in 694,649
465 individuals of European ancestry. Hum Mol Genet. 2018 [cited 25 Sep 2018].
466 doi:10.1093/hmg/ddy327

467 35. Burger M, Zimmermann TJ, Kondoh Y, Stege P, Watanabe N, Osada H, et al. Crystal
468 structure of the predicted phospholipase LYPLAL1 reveals unexpected functional
469 plasticity despite close relationship to acyl protein thioesterases. J Lipid Res. 2012;53: 43–
470 50. doi:10.1194/jlr.M019851

471 36. Wang A, Dennis EA. Mammalian lysophospholipases. Biochim Biophys Acta.
472 1999;1439: 1–16.

473 37. Lindgren CM, Heid IM, Randall JC, Lamina C, Steinthorsdottir V, Qi L, et al. Genome-
474 wide association scan meta-analysis identifies three Loci influencing adiposity and fat
475 distribution. PLoS Genet. 2009;5: e1000508. Available: pm:19557161

476 38. Heid IM, Jackson AU, Randall JC, Winkler TW, Qi L, Steinthorsdottir V, et al. Meta-
477 analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual
478 dimorphism in the genetic basis of fat distribution. Nat Genet. 2010;42: 949–960.
479 Available: http://dx.doi.org/10.1038/ng.685

480 39. Yaghootkar H, Lotta LA, Tyrrell J, Smit RAJ, Jones SE, Donnelly L, et al. Genetic
481 Evidence for a Link Between Favorable Adiposity and Lower Risk of Type 2 Diabetes,
482 Hypertension, and Heart Disease. Diabetes. 2016;65: 2448 LP – 2460. Available:
483 http://diabetes.diabetesjournals.org/content/65/8/2448.abstract

484 40. Speliotes EK, Yerges-Armstrong LM, Wu J, Hernaez R, Kim LJ, Palmer CD, et al.
485 Genome-wide association analysis identifies variants associated with nonalcoholic fatty
486 liver disease that have distinct effects on metabolic traits. PLoS Genet. 2011;7: e1001324.
487 doi:10.1371/journal.pgen.1001324

488 41. Steinberg GR, Kemp BE, Watt MJ. Adipocyte triglyceride lipase expression in human
489 obesity. Am J Physiol Endocrinol Metab. 2007;293: E958-64.
490 doi:10.1152/ajpendo.00235.2007

491 42. Shadrina AS, Sharapov SZ, Shashkova TI, Tsepilov YA. Varicose veins of lower
492 extremities: insights from the first large-scale genetic study. bioRxiv. 2018. Available:
493 http://biorxiv.org/content/early/2018/07/13/368365.abstract

494 43. Hall NG, Klenotic P, Anand-Apte B, Apte SS. ADAMTSL-3/punctin-2, a novel
495 glycoprotein in extracellular matrix related to the ADAMTS family of metalloproteases.
496 Matrix Biol. 2003;22: 501–510. doi:10.1016/s0945-053x(03)00075-1

497 44. Mead TJ, Apte SS. ADAMTS proteins in human disorders. Matrix Biol. 2018;71–72:

498 225–239. doi:10.1016/j.matbio.2018.06.002

499 45. Lin D, Chun T-H, Kang L. Adipose extracellular matrix remodelling in obesity and insulin
500 resistance. Biochem Pharmacol. 2016;119: 8–16.
501 doi:https://doi.org/10.1016/j.bcp.2016.05.005

502 46. Rask-Andersen M, Karlsson T, Ek WE, Johansson Å. Genome-wide association study of
503 body fat distribution identifies adiposity loci and sex-specific genetic effects. Nat
504 Commun. 2019;10: 339. doi:10.1038/s41467-018-08000-4

505 47. Zillikens MC, Demissie S, Hsu Y-H, Yerges-Armstrong LM, Chou W-C, Stolk L, et al.
506 Large meta-analysis of genome-wide association studies identifies five loci for lean body
507 mass. Nat Commun. 2017;8: 80. doi:10.1038/s41467-017-00031-7

508 48. Lotta LA, Gulati P, Day FR, Payne F, Ongen H, van de Bunt M, et al. Integrative genomic
509 analysis implicates limited peripheral adipose storage capacity in the pathogenesis of
510 human insulin resistance. Nat Genet. 2017;49: 17–26. Available:
511 http://dx.doi.org/10.1038/ng.3714

512 49. Singaraja RR, Tietjen I, Hovingh GK, Franchini PL, Radomski C, Wong K, et al.
513 Identification of four novel genes contributing to familial elevated plasma HDL
514 cholesterol in humans. J Lipid Res. 2014;55: 1693–1701. doi:10.1194/jlr.M048710

515 50. Chasman DI, Pare G, Mora S, Hopewell JC, Peloso G, Clarke R, et al. Forty-three loci
516 associated with plasma lipoprotein size, concentration, and cholesterol content in genome-
517 wide analysis. PLoS Genet. 2009;5.

518 51. Kasim NB, Huri HZ, Vethakkan SR, Ibrahim L, Abdullah BM. Genetic polymorphisms

associated with overweight and obesity in uncontrolled Type 2 diabetes mellitus. Biomark Med. 2016;10: 403–415. doi:10.2217/bmm-2015-0037

52. Lu Y, Day FR, Gustafsson S, Buchkovich ML, Na J, Bataille V, et al. New loci for body fat percentage reveal link between adiposity and cardiometabolic disease risk. 2016;7: 10495. Available: http://dx.doi.org/10.1038/ncomms10495

53. Huang LO, Rauch A, Mazzaferro E, Preuss M, Carobbio S, Bayrak CS, et al. Genome-wide discovery of genetic loci that uncouple excess adiposity from its comorbidities. Nat Metab. 2021;3: 228–243. doi:10.1038/s42255-021-00346-2

54. Torre YS-D la, Wadeea R, Rosas V, Herbst KL. Lipedema: friend and foe. Horm Mol Biol Clin Investig. 2018;33. doi:10.1515/hmbci-2017-0076

55. Pinnick KE, Nicholson G, Manolopoulos KN, McQuaid SE, Valet P, Frayn KN, et al. Distinct developmental profile of lower-body adipose tissue defines resistance against obesity-associated metabolic complications. Diabetes. 2014;63: 3785–3797. doi:10.2337/db14-0385

FIGURE LEGENDS

Figure 1: Scatter plot of raw, unadjusted waist circumference and leg fat % measurements of females in the UK Biobank, indicating in green the lipedema phenotype cases. The horizontal and vertical lines indicate the mean of leg fat % and waist circumference of entire female sample, respectively.

Figure 2: Manhattan plot of GWAS of the inferred lipedema phenotype from the UK Biobank, at an assumed 10% prevalence. Loci that were successfully replicated in the 'UK Lipoedema' cohort are shown in red. The red horizontal line represents the genome-wide significance p-value threshold of 5×10^{-9}.

544 **Table 1:** Inferred lipedema case and control characteristics in the UK Biobank. Means and
545 standard deviations are shown for each trait in cases and controls in analyses adjusting and not
546 adjusting for hip circumference (HC).

547

	HC-Adjusted		Not HC-Adjusted	
	Cases (n=24,450)	Controls (n=165,227)	Cases (n=24,450)	Controls (n=134,347)
Age (yrs)	56.7 (7.71)	56.6 (7.98)	56.7 (7.8)	56.3 (8.03) **
BMI (kg/m^2)	26.9 (3.85)	27 (5.55) **	28.5 (2.31)	25.7 (5.99) ***
Waist Circumference (cm)	80.1 (7.83)	85.6 (13.7) ***	84.7 (4.07)	81.7 (14.9) ***
Hip Circumference (cm)	103 (7.68)	103 (11.1) ***	106 (5.55)	101 (11.9) ***
WHR	0.78 (0.032)	0.83 (0.078) ***	0.80 (0.049)	0.81 (0.075) ***
Mean Leg Fat %	42.5 (3.53)	39.6 (6.1)***	43.9 (2.03)	38.0 (6.17)***

548 *$p<0.05$, **$p<0.0001$, *** $p<2x10^{-16}$, comparing controls to cases
549 HC: hip circumference; WHR: waist-to-hip ratio

Table 2: List of lead variants showing significant association with the lipedema phenotype (hip circumference (HC) adjusted) from the UK Biobank GWAS, including associations of the corresponding SNPs with lipedema in the UK Lipoedema replication study. The SNPs have been annotated to their nearest gene(s). Effect size (OR) and p-values are shown for both analyses. The SNPs have been sorted in ascending order of p-value from UK Biobank GWAS. The rows with bold font indicates the successfully replicated loci.

				UK Biobank					UK Lipoedema study			
SNP	**Chr**	**BP**	**Nearest gene**	**A1**	**A0**	**A1 Freq**	**OR**	**p-value**	**SNP**	**A1**	**A1 Freq**	**OR**
rs72959041	6	127,454,893	*RSPO3*	G	A	0.95	1.24	2.70E-19	rs72959041	A	0.052	0.9920
rs1128249	**2**	**165,528,624**	***GRB14-COBLL1***	**G**	**T**	**0.61**	**0.92**	**3.00E-16**	**rs1128249**	**T**	**0.385**	**1.0086**
rs11057418	12	124,508,976	*ZNF664-FAM101A*	G	C	0.798	1.1	5.60E-14	rs11057418	C	0.204	0.9998
rs6905288	**6**	**43,758,873**	***VEGFA***	**G**	**A**	**0.429**	**1.07**	**1.80E-13**	**rs6905288**	**A**	**0.550**	**0.9029**
rs6602994	15	84,490,757	*ADAMTSL3*	T	C	0.276	0.92	6.90E-13	rs7164141[a]	G	0.730	1.0033
rs10649697	5	127,432,908	*SLC12A2*	T	TAGA	0.246	0.92	1.30E-12	no proxy found			
rs4616635	3	64,702,275	*ADAMTS9*	C	G	0.723	0.93	4.90E-12	rs4616635	G	0.269	1.0064
rs749853052	1	219,747,226	*LYPLAL1*	TG	T	0.702	0.93	1.50E-11	rs2820443[b]	C	0.291	1.0070
rs536569640	12	123,553,002	*PITPNM2*	TAATA	T	0.726	1.08	1.90E-11	no proxy found			
rs28394864	17	47,450,775	*ZNF652*	G	A	0.538	1.07	1.90E-11	rs28394864	A	0.441	0.9986
rs543302184	10	96,009,182	*PLCE1*	T	TA	0.58	1.07	3.00E-11	no proxy found			
rs11772918	7	46,609,344	intergenic	A	G	0.46	0.93	3.60E-11	rs11772918	G	0.531	0.9986
rs62492368	7	150,537,635	*ABP1/AOC1*	G	A	0.694	1.07	4.00E-11	rs62492368	A	0.309	0.9973
rs5868014	5	55,860,907	*ANKRD55-MAP3K1*	GC	G	0.814	1.08	2.70E-10	rs28650790[c]	T	0.190	0.9933
rs71490394	11	62,370,155	*EML3*	G	A	0.629	1.07	1.20E-09	rs71490394	A	0.351	0.9956
rs28849840	12	50,703,384	*LIMA1-FAM186A*	G	A	0.65	0.94	1.50E-09	rs28849840	A	0.349	0.9969
rs565113908	18	37,904,550	intergenic	T	G	0.999	0.52	1.60E-09	no proxy found			
rs4680338	3	156,794,425	*LEKR1-CCNL1*	C	G	0.595	0.94	2.30E-09	rs4680338	G	0.403	1.0044

A1 refers to effect allele that OR (odds ratio) corresponds to; A1 Freq: effect allele frequency

[a] A allele at rs7164141 in LD (r^2=1.0) with T allele of rs6602994
[b] C allele at rs2820443 in LD (r^2=1.0) with no G nucleotide at rs749853052
[c] C allele at rs28650790 in LD (r^2=1.0) with C allele at rs5868014

Figure 1: Scatter plot of raw, unadjusted waist circumference and leg fat % measurements of females in the UK Biobank, indicating in green the lipedema phenotype cases. The horizontal and vertical lines indicate the mean of leg fat % and waist circumference of entire female sample, respectively.

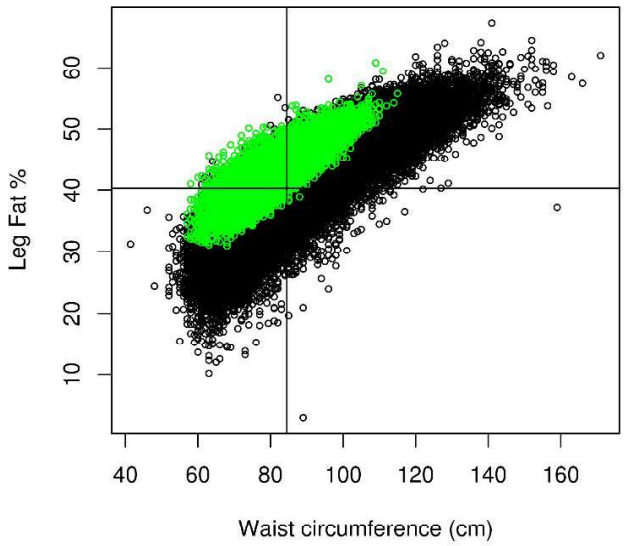

Figure 2: Manhattan plot of GWAS of the inferred lipedema phenotype from the UK Biobank, at an assumed 10% prevalence. Loci that were successfully replicated in the 'UK Lipoedema' cohort are shown in red. The red horizontal line represents the genome-wide significance p-value threshold of 5×10^{-9}.