Genetic and phenotypic links between obesity and extracellular vesicles

Genetic and phenotypic links between obesity and extracellular vesicles

Ranran Zhai[1,3†], Lu Pan[4], Zhijian Yang[1,3], Ting Li[1,3], Zheng Ning[4], Yudi Pawitan[4], James F. Wilson[5,6†], Di Wu[7†*], Xia Shen[1,2,3,4,5†*]

[1] Biostatistics Group, School of Life Sciences, Sun Yat-sen University, Guangzhou, China

[2] State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University, Shanghai, China

[3] Center for Intelligent Medicine Research, Greater Bay Area Institute of Precision Medicine (Guangzhou), Fudan University, Guangzhou, China

[4] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

[5] Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, United Kingdom

[6] MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom

[7] Vesicode AB, Stockholm, Sweden

[†] These authors contributed equally to this work.

* Correspondence should be addressed to:

Prof. Xia Shen, Address: Greater Bay Area Institute of Precision Medicine (Guangzhou), 2nd Nanjiang Road, Nansha District, Guangzhou 511458, China. Tel: +86 138 1115 3502; Email: xia.shen@ed.ac.uk

Dr. Di Wu, Address: Vesicode AB, Nobels väg 16, 17165 Solna, Sweden. Tel: +46 73 919 1116; Email: di.wu@vesicode.com

Abstract

Obesity has a highly complex genetic architecture, making it difficult to understand the genetic mechanisms, despite the large number of discovered loci via genome-wide association studies (GWAS). Omics techniques have provided a better resolution to view this problem. As a proxy of cell-level biology, extracellular vesicles (EVs) are useful for studying cellular regulation of complex phenotypes such as obesity. Here, in a well-established Scottish cohort, we utilized a novel technology to detect surface proteins across millions of single EVs in each individual's plasma sample. Integrating the results with established obesity GWAS, we inferred 78 types of EVs carrying one or two of 12 surface proteins to be associated with adiposity-related traits such as waist circumference. We then verified that particular EVs' abundance is negatively correlated

with body adiposity, while no association with lean body mass. We also revealed that genetic variants associated with protein-specific EVs capture 2-4-fold heritability enrichment for blood cholesterol levels. Our findings provide evidence that EVs with specific surface proteins have phenotypic and genetic links to obesity and blood lipids, respectively, guiding future EV biomarker research.

Introduction

Overweight and obesity is currently major public health issue and most severe in high-income countries. In 2016, more than 1.9 billion adults were overweight, of those over 650 million were obese, and the prevalence of obesity continues to rise (1). Obesity is a risk factor for various chronic diseases such as type 2 diabetes (2) and cardiovascular disease (3), which in turn can lead to reduced quality of life, premature death, or disability (4). Obesity is defined as an excessive accumulation of body fat and is most estimated using body mass index (BMI, kg/m^2) in epidemiological studies and clinical practice. However, BMI cannot distinguish lean from fat mass, nor does it capture information on where in the body the adiposity is concentrated. Abdominal fat, and visceral fat (fat around the internal organs of the body) in particular, has been linked with obesity-related metabolic risk factors, including high total and low-density lipoprotein cholesterol (HDL-c and LDL-c) (5,6), while fat in the lower body (e.g., around the hips) is associated with a protective lipid and glucose profile (7). Thus, proxies of abdominal adiposity, like waist circumference (WC) and waist-to-hip ratio (WHR), are of increasing interest in investigating the mechanism and regulation of obesity and its complications.

Genome-wide association studies (GWAS) have begun to elucidate the genetic architecture of body fat distribution phenotypes. Hundreds of loci associated with WC and WHR have been identified (8–10). However, the underlying biological knowledge gained from GWAS-discovered single nucleotide polymorphisms (SNPs) is, so far, limited. The discovered SNPs have tiny effects, explaining a small proportion of phenotypic variance of these abdominal adiposity-related phenotypes (11). Efforts have been made to integrate GWAS summary statistics with gene expression in adipose tissues to

further characterize obesity-causal genes (12). Integration of GWAS and gene expression quantitative trait loci (eQTL) studies has also achieved success in prioritizing genes in obesity-related pathways (13,14).

To better understand the regulation of abdominal adiposity, adiposity-related biomarkers are also of general interest. In the last decade, extracellular vesicles (EVs) have become essential in biomarker discovery for their emerging roles in physiological and pathological pathways. These nanometer-sized membranous vesicles emitted from nearly all cells have a variety of functions and can regulate target cells' metabolism by conveying genomic material (15) and proteins (16). Recent studies reported that circulating EVs were elevated in obese humans (17), and levels of EVs and specific proteins in EVs were altered in diabetic mice (18). In these studies, EVs were isolated using "EV-specific" markers in bulk, considering the level of EVs or a protein carried by the EVs as an overall measurement. Recently, a new technology named proximity barcoding assay (PBA) was developed to detect multiple surface proteins on single EVs in a high-throughput manner(19). This allows us to quantify EVs with specific surface markers, thus better understanding the heterogeneity of EVs.

Attempting to link specific EVs to obesity, in this study, we aim to identify particular types of EVs associated with abdominal adiposity. We approach this by investigating EVs with specific adiposity-related surface proteins based on established GWAS summary statistics for adiposity-related phenotypes, followed by validation using individuals' body fat distribution measurements. We perform genomic analysis for SNPs associated with the levels of specific EVs. We show that the heritability for plasma LDL-c and total cholesterol levels are enriched at the loci associated with EVs carrying adiposity-related surface proteins.

Results

Prioritizing obesity-related EV surface markers using GWAS summary statistics

The commercial EV surface marker panel available to us consists of 113 proteins targeted by specific antibodies (see Materials and Methods). To identify the associations between the encoding genes of these protein markers and obesity, we started by constructing corresponding gene-level

association scores for obesity-related phenotypes. We collected the established summary-level association data for WC and WHR related phenotypes conducted by the Genetic Investigation of Anthropometric Traits (GIANT) consortium. Those traits include WC and WHR, as well as WC and WHR adjusted or stratified by other factors (BMI, smoking, and physical activity), all using European subjects (details are in **Supplementary Table 1**). We also performed a multi-trait GWAS analysis of those summary statistic data using MultiABEL (20), where we can get combined SNP p-values of those traits. We then used the Pathway Scoring Algorithm (PASCAL) to derive gene-level p-values from SNP-level association statistics(21). Out of the 113 coding genes, 12 were significantly associated with multi-trait waist adiposity GWAS (false discovery rate (FDR) < 0.01) (**Figure 1**).

Measurement quality of candidate EV protein markers

The PBA technique is briefly described in **Figure 2A** (see Materials and Methods). Using the PBA technique, we could detect the abundance of the 113 surface proteins across many single EVs in each plasma sample. Before measuring multiple individual samples, we performed in-depth sequencing quantification (50 million reads per sample) for two independent plasma test samples from two unrelated individuals. We obtained approximately 1.5 million EVs in each of two testing plasma samples, of which ~ 62% were singleton (EVs that only carry one type of protein) (**Supplementary Figure 1A**). To reduce background noise, we conducted data quality-control procedures (see Materials and Methods), which resulted in 61,108 EVs measured with high quality in sample 1 and 73,139 in sample 2.

Most of these EVs (95.6% and 96.4% in samples 1 and 2) carry 1 to 10 different proteins, and the most abundant EVs carry 5 to 6 different proteins (**Supplementary Figure 1B**). Due to multiplex limitations, especially when sequencing depth is not as high when measuring many samples, single EVs carrying many overlapping proteins are challenging to quantify in two independent samples. We thus focused on investigating EVs containing one or two of the obesity-related markers identified above across possible proteins and protein-protein combinations. The abundance of such EVs with specific

marker profiles was consistently measured both between samples (**Figure 2B**) and replicates (**Figure 2C**), suggesting the stability of PBA in measuring these protein markers.

Validating the association between overall protein levels and obesity

To validate the obesity-related EV surface markers prioritized by GWAS association statistics, we conducted the 113-marker single-EV surface protein quantification using PBA in 96 selected individuals in the ORCADES cohort (see Materials and Methods). As a comparison, we also quantified the total abundance of 37 proteins out of the 113 that could be measured in plasma by the Olink Proximity Extension Assay (PEA) panels. For these individuals in the cohort, besides general anthropometric measurements such as height, BMI, WHR, and WC, we also had Dual Energy X-ray Absorptiometry (DEXA) scan data available to quantify their fat distribution in the body. The DEXA phenotypes included total lean mass, trunk fat, trunk lean mass, and visceral fat.

First, we tested the associations between adiposity phenotypes and overall protein levels quantified by both PBA and PEA (**Figure 3A**; see Materials and Methods). Protein abundance on the EV surface had little association with the obesity-related phenotypes in our 113-marker panel. While for two proteins quantified by PEA, levels of VEGFA and CD63 in plasma were associated with body fat distribution-related traits. Among those associations, levels of plasma CD63 have the strongest effect on trunk fat mass (effect size = 1.25, equivalent to 6.5 standard deviation increase of trunk fat per standard deviation of plasma CD63 level, $R^2 = 0.13$, $P = 0.011$). There is a positive correlation between plasma VEGFA level and BMI, supported by another study (22) where they found a positive correlation between circulating VEGF levels and BMI. We then tested the sex-by-protein interaction effect on those phenotypes (**Figure 3B**). Although plasma protein levels showed almost no interaction effects, the overall EV protein levels, especially that of ERBB2, had significant interaction effects on fat distribution traits such as waist circumference: EV ERBB2 levels are more strongly associated with abdominal adiposity in women.

Validating the association between marker-specific EVs and obesity

As well as the overall level of protein on the surface of EVs, the PBA technology can also provide us with surface protein profiles for individual EVs. Here, we tested the associations between the obesity-related phenotypes and the levels of EVs with specific protein profiles (**Figure 4A**; see Materials and Methods). In contrast to overall protein levels on the EV surface, levels of EVs with single-marker specificity showed strong associations with these phenotypes. To be more stringent, we performed FDR calculation across all the statistical tests for the associations regarding total protein levels and specific EVs (p-value distribution of each part of the analysis is given in **Supplementary Figure 2**). Levels of EVs carrying HSPA1A (EV_{HSPA1A}) had the most significant association with BMI (effect size = -1.62, equivalent to 7.60 standard deviation decrease of BMI per standard deviation of plasma EV_{HSPA1A} level, $R^2 = 0.17$, $P = 4.5\times10^{-4}$), meaning that individuals with less of these EVs have significantly higher BMI. Focusing on EVs identified by two protein markers on their surface (double-marker specificity) revealed even stronger associations, for EVs carrying several specific protein-protein combinations had even stronger associations. For instance, levels of $EV_{HSPA1A\ \&\ ICAM1}$ had the strongest association with trunk fat (effect size = -2.31, equivalent to 12.0 standard deviation decrease of trunk fat per standard deviation of plasma $EV_{HSPA1A\ \&\ ICAM1}$ level, $R^2 = 0.27$, $P = 7.1\times10^{-5}$) and visceral fat (effect size = -2.22, equivalent to 0.70 standard deviation decrease of visceral fat per standard deviation of plasma $EV_{HSPA1A\ \&\ ICAM1}$ level, $R^2 = 0.31$, $P = 1.1\times10^{-4}$). These are stronger associations than observed for these traits with EV_{HSPA1A}, but levels of $EV_{HSPA1A\ \&\ ICAM1}$ were less significantly associated with BMI. Similarly, EV_{ITGB8} abundances were not a better indicator of adiposity than its sub-population $EV_{ITGB8\ \&\ VEGFA}$ or $EV_{ITGB8\ \&\ ITGA7}$.

These associations were observed with body fat-related traits, like the trunk and visceral fat, but not with total or trunk lean mass. We also tested the sex-by-EV interaction effects on these phenotypes (**Figure 4B**). The above significant types of EVs tended to have different effects in different sexes. Although the signals were not statistically robust, the results also suggested stronger sex-by-EV interaction effects in women than in men.

EV-associated SNPs capture enriched heritability for blood lipids

Although with limited power, we performed $12 \times (113 - 12) + 66 = 1{,}276$ GWAS of double-protein EV (EV carrying at least one of 12 markers) levels in 96 ORCADES individuals. For each double-protein EV phenotype, we tested the associations of genome-wide SNPs imputed to the 1000 Genomes reference panel (minor allele frequency > 0.2 to avoid inflated false positives due to the small sample size) using a linear model, adjusted for sex, age, population stratification, and other covariates (see Materials and Methods). Despite the small sample size, such GWAS resulted in an association p-value distribution significantly deviating from the null (**Supplementary Figure 3**). With a stringent minor allele frequency cutoff and a Bonferroni-corrected significance threshold of $P < 5 \times 10^{-8}/1276 = 3.9 \times 10^{-11}$, we did not report any specific SNPs associated with the EV phenotypes.

Nevertheless, to understand the role of the EV-associated SNPs in the regulation of obesity, we used stratified LD score regression (23) (S-LDSC) to evaluate heritability enrichment on these SNPs. Seven traits were considered, including BMI, WC, WHR, and four classical blood lipids (24) (high- and low-density lipoprotein cholesterol (HDL-c, LDL-c), total cholesterol, and triglycerides). Using a less stringent threshold in the EV GWAS ($P < 1 \times 10^{-6}$), we annotated the SNPs associated with EVs carrying different types of protein-protein combinations (see Materials and Methods). Heritability enrichment was detected for most annotations. Specifically, we found 2-4-fold heritability enrichment for LDL-c and total cholesterol that is positively correlated with abdominal fat (**Figure 5**). These results indicated that the EVs with particular surface proteins are not only phenotypically but also genetically correlated with obesity and cholesterol metabolism.

Discussion

In this study, we investigated the association between plasma EV levels and body fat distribution, with an emphasis on abdominal fat. We found that the levels of plasma EVs carrying specific surface proteins were negatively associated with adiposity-related measurements, including BMI, WC, WHR, trunk fat, and visceral fat. For some types of EV (e.g., $EV_{HSPA1A\ \&\ ICAM1}$), the association with trunk and visceral fat mass is more significant than BMI, WC, or WHR, implying that those EVs might be more related to or arise from visceral adipose tissue. Notwithstanding our inability to track the tissue or even

cell-type origins of these EVs, to our knowledge, our results for the first time demonstrate that protein-specific EVs in plasma are associated with body fat distribution and obesity in humans.

We were able to identify a few genome-wide significant SNPs associated with certain EVs with specific surface protein markers, despite that the small sample size limited our discovery power. The power of the EV-QTL scan might come from two sources: 1) The PBA technique allowed screening across EVs with many different protein-protein combinations; 2) The double-protein signature on the EVs enhanced the measurement specificity so that a more specific genetic basis for the EVs could be mapped. However, due to the lack of replication, such perspectives remain to be validated in future larger GWAS analyses of these EV phenotypes.

We integrated GWAS summary-level data with EV protein measurements. Still, our current sample size is insufficient to conduct systematic genomic analyses of EV surface proteins and all types of protein-specific EVs. In future studies, we foresee that expansion in sample size would open a new area for extracellular vesicle omics, providing power to establish the genetic architecture of various kinds of EVs. Such resources would allow better causal inference of EVs on disease phenotypes. This is particularly possible for circulating EVs, for which only plasma samples would be required.

Although measured from the blood, with tissue- or cell-type-specific protein markers designed in the PBA panel, one can further consider PBA as a proxy single-cell technology. Future expansion of our panels may allow us to investigate specific tissues, either via tissue-specific markers or deconvolution algorithms, when tissue-specific EV proteomic profiles become available. This could expand this exciting research area to explore the genetic basis of tissue- and cell-type-specific EVs.

Although it is not novel to identify different obesity regulatory mechanisms in men and women, it is also noteworthy that we detected a more substantial sex-by-EV interaction effect in women than in men, meaning EVs have more impact on women in terms of their association with body fat. While sex differences in obesity and body fat distribution are well established, our results suggest that EVs might also play a role in these differences. Our results also reveal that EV-QTLs share heritability with LDL-c and total cholesterol, both of which increase the risk of cardiovascular and metabolic diseases. Though the biological mechanism underlying such associations remains unknown, our results show genetic connections between levels of protein-specific EVs and adiposity measurements.

A recent study showed that the liver could secret EVs to modulate adipocyte remodeling in response to excessive lipid (25). It is well-known that RNA transferred by EVs can regulate or serve as the template for protein synthesis in the recipient cells (26). Another study has shown that miRNAs in exosomes (a subpopulation of EVs) from obese visceral adipocytes could down-regulate the expression of ACVR2B in the TGF-β signaling pathway (27), which plays a crucial role in obesity and insulin resistance (28–30). Such discoveries are consistent with our finding that individuals with higher visceral fat tend to have fewer EVs with ITGB8 & VEGFA, where ITGB8 can activate TGF-β when combined with ITGAV (integrin α subunit V) (31,32).

The heterogeneity of the EV population has been problematic in studies using EVs as a diagnostic and therapeutic biomarker. Kaur *et al.* (33) reported that EVs captured by different surface protein markers had distinct RNA profiles. In accordance with their study, our results showed that EVs categorized by different surface proteins, such as integrins (ITGB8, ITGA9, and ITGA7), tetraspanins (CD63), heat shock protein (HSPA1A), etc., have similar but varied associations with human adiposity traits. Such results highlighted the heterogeneity of EV subpopulations, suggesting that single-EV measurement technologies such as PBA have the potential to study the origins and functions of different types of EVs.

We also provided a piece of new evidence on the potential EV-based liquid biopsy as a tool. The PBA technique only focuses on the membranous proteins of EVs at present, neglecting other functional molecules such as miRNAs, lipids, and endogenous proteins (34). It limits our insight into the underlying regulatory mechanism of EVs in physiological and pathological pathways. However, being able to quantify the particular molecular profiles of single EVs has already brought substantial power.

Materials and Methods

ORCADES samples

Our 96 individuals are a subset of 2080 volunteers from the Orkney Complex Disease Study (ORCADES) cohort. The Orkney Complex Disease Study (ORCADES) is a family-based, cross-sectional study that seeks to identify genetic factors influencing cardiovascular and other disease

risk in the isolated archipelago of the Orkney Isles in northern Scotland (35). Genetic diversity in this population is decreased compared to Mainland Scotland, consistent with the high levels of endogamy historically. 2,078 participants aged 16-100 years were recruited between 2005 and 2011, most having three or four grandparents from Orkney, the remainder with two Orcadian grandparents. Fasting blood samples were collected, and many health-related phenotypes and environmental exposures were measured in each individual. All participants gave written informed consent, and the study was approved by Research Ethics Committees in Orkney and Aberdeen (North of Scotland REC). Genome-wide genotyping was performed using Illumina HumanHap300 and OmniExpress arrays. The individuals were randomly selected with a balanced sex ratio from the subset of individuals with the most measured phenotypes.

Profiling EV surface proteins with PBA

All the samples were sent to Vesicode AB (Sweden) with dry ice, and proximity barcoding assay (PBA) (19) was carried out according to Vesicode AB's PBA standard operation procedure (SOP). The raw data BCL sequencing files were converted to a fastq file with bcl2fastq software (Illumina). Proteins were determined by mapping protein tags in the sequences to the designed panel of oligonucleotides conjugated to antibodies. The EV tags were extracted and used as the identity for single EVs. A file consisting of the EV tags and their protein expression profile was obtained for each sample. Different samples resulted in a different number of sequenced reads; each read count of the protein expression was normalized by the total number of reads in the corresponding sample. The readouts were comparable across samples as "the proportion of proteins per sample."

EV profiles of two testing samples via deep sequencing

Using deep sequence quantification (50 million per sample), we obtained approximately 1.5 million EVs in each of two testing plasma samples, of which ~ 62% were singleton (EVs that only carry

one type of protein), where ~ 50% of these only carry one protein. To reduce background noise, we first filtered EVs that have more than five proteins, which resulted in our EV matrix (D, $m \times n$). We calculated the expected values and c values of D, using the following equation: $E = AB^T$ where A is the matrix of sums of rows ($1 \times m$), and B is the matrix of sums of columns ($1 \times n$), $\chi = (D - E)/\sqrt{E}$ cells with top 95% quantile chi values were considered to be actual protein count, others were reset to be 0.

Gene-based statistics from GWAS summary statistics

Gene-based p-values were generated by PASCAL (21). In our analysis, the window size for computing sum and maximum of chi-squared statistics was 50kb up- and downstream. For better quality, the LD information was obtained from the UK10K data (https://www.ebi.ac.uk/ega/datasets/EGAD00001000776) instead of the default 1000 Genomes project.

Normalization of EV phenotypes and regression models

We used 96 individuals from the ORCADES cohort to test the associations between obesity-related traits and the abundance of EVs with specific surface proteins markers. For each obesity-related trait Y, we standardized the data as $Y^* = (Y - \mu)/\sigma$, where μ and σ represent the mean and standard error of the phenotypic data vector. For the abundance of each EV phenotype X, corresponding to a type of EVs that carry a particular protein profile, we inverse-Gaussian transformed the data vector into X^*, following a standard normal distribution. We conducted a multiple linear regression model to test specific associations between EV types and the obesity-related traits:

$$Y^* = \mu + X^* \beta_{EV} + Sex \beta_{Sex} + Age \beta_{Age} + X^* Sex \beta_{EV-by-sex} + \varepsilon$$

where μ is the overall mean, β's are the corresponding effect parameters, and ε is the residual. The regression analysis was performed using the lm() procedure in R.

Genome-wide association analysis of protein-specific EVs

Prior to GWAS, each EV phenotype was adjusted for fixed effects of sex, age, and the other experimental factors. The residuals were inverse-Gaussian-transformed to standard normal distributions. The residuals expressed as Z-scores were used for all genetic association analyses. In both the genotypes from the SNP array and 1000 Genomes-imputed data, markers with minor allele frequency < 0.05 were excluded. Population structure was corrected using the GRAMMAR+ transformation (36), implemented in the GenABEL package (37). The genomic relationship matrix used in the analyses was generated by the ibs() function (with weight = 'freq' option), which uses SNP array data to estimate the realized pairwise kinship coefficient. The polygenic() function was used to obtain the GRAMMAR+ transformed phenotypes (grresidualY) from linear mixed models. All univariate GWAS inflation factors (lambda values) were close to 1, showing that this method efficiently accounts for family structure. The processed phenotypes were tested in genome scans using REGSCAN (38).

Heritability enrichment analysis

For each obesity-related protein marker, we had 112 GWASed EV phenotypes for the EVs that carry different protein-protein combinations with this particular protein. To overcome the power loss due to the small sample size, for each of the 12 obesity-related proteins, we extracted all the SNPs with $P < 1\times10^{-6}$ in all 112 GWAS. In addition, we also extracted all the SNPs with $P < 1\times10^{-6}$ in 66 GWAS for the EVs that carry obesity-related protein-protein combinations. So, we get 13 sets of SNPs in total. Each set of SNPs was then annotated with a flanking window of 1 kilobase (kb). We then used stratified LD score regression (S-LDSC) (23,39) to test whether each set of such SNPs was enriched for the heritability of each obesity-related complex trait. The Z-scores of each complex trait GWAS were harmonized by the munge_sumstats.py procedure of the ldsc software. LD scores of HapMap3 SNPs (MHC region excluded) for the annotated SNPs were computed using a 1-cM window (default). The heritability enrichment was evaluated by an enrichment score of the proportion of explained heritability divided by the proportion of annotated SNPs.

Data availability

The summary statistics of 66 EV abundance phenotypes, corresponding to protein-protein combinations of 12 surface markers, will be made publicly available upon publication of this paper. Summary statistics of obesity-related GWAS are from the GIANT consortium: http://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files, URLs for GWAS we used in this paper are in **Supplementary Table 1**. The full results of the EV-obesity associations are available in **Supplementary Tables 2 and 3**, corresponding to **Figures 3 and 4**, respectively.

Code availability

MultiABEL: https://github.com/xiashen/MultiABEL. PASCAL: https://www2.unil.ch/cbg/index.php?title=Pascal. LDSC: https://github.com/bulik/ldsc.

Author contributions

X.S., J.F.W., and D.W. initiated the study; X.S. coordinated the study; J.F.W. contributed ORCADES plasma samples; D.W. performed measurements using PBA; R.Z. performed data analysis; L.P. contributed to EV data pre-processing. Z.Y., T.L., Z.N., and Y.P. contributed to data analysis and interpretation; R.Z. and X.S. drafted the manuscript; All authors approved the final manuscript.

Acknowledgments

X.S. was in receipt of a Swedish Research Council Starting Grant (No. 2017-02543) and a National Natural Science Foundation of China grant (No. 12171495). The Orkney Complex Disease Study (ORCADES) was supported by the Chief Scientist Office of the Scottish Government (CZB/4/276, CZB/4/710), a Royal Society URF to J.F.W., the MRC Human Genetics Unit quinquennial program "QTL in Health and Disease," Arthritis Research UK and the European Union framework program 6 EUROSPAN project (contract no. LSHG-CT-2006-018947). We acknowledge

348 support from the MRC Human Genetics Unit program grant, “Quantitative traits in health and disease”
349 (U. MC_UU_00007/10).
350 DNA extractions were performed at the Edinburgh Clinical Research Facility, University of
351 Edinburgh. We would like to acknowledge the invaluable contributions of the research nurses in
352 Orkney, the administrative team in Edinburgh, and the people of Orkney. We thank the GIANT
353 consortium for making their GWAS results publicly available.
354

355 **Conflict of Interest Statement**

356 D.W. has filed a patent application (PCT/SE2014/051133) describing the PBA technique. D.W.
357 holds shares in Vesicode AB commercializing the PBA technology. The remaining authors declare no
358 competing interests.
359

360 # References

361 1. WHO (2016) Obesity and overweight: Fact sheet. *WHO Media Centre*.
362 2. Kahn, S.E., Hull, R.L. and Utzschneider, K.M. (2006) Mechanisms linking obesity to insulin
363 resistance and type 2 diabetes. *Nature*, **444**, 840–846.
364 3. Hubert, H.B., Feinleib, M., McNamara, P.M. and Castelli, W.P. (1983) Obesity as an independent
365 risk factor for cardiovascular disease: a 26-year follow-up of participants in the Framingham Heart
366 Study. *Circulation*, **67**, 968–977.
367 4. Peeters, A., Barendregt, J.J., Willekens, F., Mackenbach, J.P., Mamun, A. al and Bonneux, L. (2003)
368 Obesity in Adulthood and Its Consequences for Life Expectancy: A Life-Table Analysis. *Annals of*
369 *Internal Medicine*, **138**, 24.
370 5. Bentham Science Publisher, B.S.P. (2006) Metabolic Obesity: The Paradox Between Visceral and
371 Subcutaneous Fat. *Current Diabetes Reviews*, **2**, 367–373.
372 6. Nguyen-Duy, T.-B., Nichaman, M.Z., Church, T.S., Blair, S.N. and Ross, R. (2003) Visceral fat and
373 liver fat are independent predictors of metabolic risk factors in men. *American Journal of*
374 *Physiology-Endocrinology and Metabolism*, **284**, E1065–E1071.
375 7. Snijder, M.B., Zimmet, P.Z., Visser, M., Dekker, J.M., Seidell, J.C. and Shaw, J.E. (2004)

Independent and opposite associations of waist and hip circumferences with diabetes, hypertension and dyslipidemia: the AusDiab Study. *International Journal of Obesity*, **28**, 402–409.

8. Heid, I.M., Jackson, A.U., Randall, J.C., Winkler, T.W., Qi, L., Steinthorsdottir, V., Thorleifsson, G., Zillikens, M.C., Speliotes, E.K., Mägi, R., *et al.* (2010) Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nature Genetics*, **42**, 949–960.

9. Shungin, D., Winkler, T., Croteau-Chonka, D.C., Ferreira, T., Locke, A.E., Mägi, R., Strawbridge, R.J., Pers, T.H., Fischer, K., Justice, A.E., *et al.* (2015) New genetic loci link adipose and insulin biology to body fat distribution. *Nature*.

10. Pulit, S.L., Stoneman, C., Morris, A.P., Wood, A.R., Glastonbury, C.A., Tyrrell, J., Yengo, L., Ferreira, T., Marouli, E., Ji, Y., *et al.* (2019) Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Human Molecular Genetics*, **28**, 166–174.

11. Locke, A.E., Kahali, B., Berndt, S.I., Justice, A.E., Pers, T.H., Day, F.R., Powell, C., Vedantam, S., Buchkovich, M.L., Yang, J., *et al.* (2015) Genetic studies of body mass index yield new insights for obesity biology. *Nature*, **518**, 197–206.

12. Pan, D.Z., Garske, K.M., Alvarez, M., Bhagat, Y. v., Boocock, J., Nikkola, E., Miao, Z., Raulerson, C.K., Cantor, R.M., Civelek, M., *et al.* (2018) Integration of human adipocyte chromosomal interactions with adipose gene expression prioritizes obesity-related genes from GWAS. *Nature Communications*, **9**, 1512.

13. Pavlides, J.M.W., Zhu, Z., Gratten, J., McRae, A.F., Wray, N.R. and Yang, J. (2016) Predicting gene targets from integrative analyses of summary data from GWAS and eQTL studies for 28 human complex traits. *Genome Medicine*, **8**, 84.

14. Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M.R., Powell, J.E., Montgomery, G.W., Goddard, M.E., Wray, N.R., Visscher, P.M., *et al.* (2016) Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nature Genetics*, **48**, 481–487.

15. Valadi, H., Ekström, K., Bossios, A., Sjöstrand, M., Lee, J.J. and Lötvall, J.O. (2007) Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nature Cell Biology*, **9**, 654–659.

16. Vallabhaneni, K.C., Penfornis, P., Dhule, S., Guillonneau, F., Adams, K. v., Mo, Y.Y., Xu, R., Liu, Y., Watabe, K., Vemuri, M.C., *et al.* (2015) Extracellular vesicles from bone marrow mesenchymal stem/stromal cells transport tumor regulatory microRNA, proteins, and metabolites. *Oncotarget*, **6**, 4953–4967.

17. Eguchi, A., Lazic, M., Armando, A.M., Phillips, S.A., Katebian, R., Maraka, S., Quehenberger, O., Sears, D.D. and Feldstein, A.E. (2016) Circulating adipocyte-derived extracellular vesicles are novel markers of metabolic stress. *Journal of Molecular Medicine*, **94**, 1241–1253.

18. Freeman, D.W., Noren Hooten, N., Eitan, E., Green, J., Mode, N.A., Bodogai, M., Zhang, Y., Lehrmann, E., Zonderman, A.B., Biragyn, A., *et al.* (2018) Altered Extracellular Vesicle Concentration, Cargo, and Function in Diabetes. *Diabetes*, **67**, 2377–2388.

19. Wu, D., Yan, J., Shen, X., Sun, Y., Thulin, M., Cai, Y., Wik, L., Shen, Q., Oelrich, J., Qian, X., *et al.* (2019) Profiling surface proteins on individual exosomes using a proximity barcoding assay. *Nature Communications*, **10**, 3854.

20. Shen, X., Klarić, L., Sharapov, S., Mangino, M., Ning, Z., Wu, D., Trbojević-Akmačić, I., Pučić-Baković, M., Rudan, I., Polašek, O., *et al.* (2017) Multivariate discovery and replication of five novel loci associated with Immunoglobulin G N-glycosylation. *Nature Communications*, **8**, 447.

21. Lamparter, D., Marbach, D., Rueedi, R., Kutalik, Z. and Bergmann, S. (2016) Fast and Rigorous Computation of Gene and Pathway Scores from SNP-Based Summary Statistics. *PLOS Computational Biology*, **12**, e1004714.

22. Loebig, M., Klement, J., Schmoller, A., Betz, S., Heuck, N., Schweiger, U., Peters, A., Schultes, B. and Oltmanns, K.M. (2010) Evidence for a Relationship between VEGF and BMI Independent of Insulin Sensitivity by Glucose Clamp Procedure in a Homogenous Group Healthy Young Men. *PLoS ONE*, **5**, e12610.

23. Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.-R., Anttila, V., Xu, H., Zang, C., Farh, K., *et al.* (2015) Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature Genetics*, **47**, 1228–1235.

24. Teslovich, T.M., Musunuru, K., Smith, A. v., Edmondson, A.C., Stylianou, I.M., Koseki, M., Pirruccello, J.P., Ripatti, S., Chasman, D.I., Willer, C.J., *et al.* (2010) Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*, **466**.

25. Zhao, Y., Zhao, M.-F., Jiang, S., Wu, J., Liu, J., Yuan, X.-W., Shen, D., Zhang, J.-Z., Zhou, N., He, J., *et al.* (2020) Liver governs adipose remodelling via extracellular vesicles in response to lipid overload. *Nature Communications*, **11**, 719.

26. Mittelbrunn, M., Gutiérrez-Vázquez, C., Villarroya-Beltri, C., González, S., Sánchez-Cabo, F., González, M.Á., Bernad, A. and Sánchez-Madrid, F. (2011) Unidirectional transfer of microRNA-loaded exosomes from T cells to antigen-presenting cells. *Nature Communications*, **2**, 282.

27. Ferrante, S.C., Nadler, E.P., Pillai, D.K., Hubal, M.J., Wang, Z., Wang, J.M., Gordish-Dressman, H., Koeck, E., Sevilla, S., Wiles, A.A., *et al.* (2015) Adipocyte-derived exosomal miRNAs: a novel mechanism for obesity-related disease. *Pediatric Research*, **77**, 447–454.

28. Samad, F., Yamamoto, K., Pandey, M. and Loskutoff, D.J. (1997) Elevated Expression of Transforming Growth Factor-β in Adipose Tissue from Obese Mice. *Molecular Medicine*, **3**, 37–48.

29. Tsurutani, Y., Fujimoto, M., Takemoto, M., Irisuna, H., Koshizaka, M., Onishi, S., Ishikawa, T., Mezawa, M., He, P., Honjo, S., *et al.* (2011) The roles of transforming growth factor-β and Smad3 signaling in adipocyte differentiation and obesity. *Biochemical and Biophysical Research Communications*, **407**, 68–73.

30. Yadav, H., Quijano, C., Kamaraju, A.K., Gavrilova, O., Malek, R., Chen, W., Zerfas, P., Zhigang, D., Wright, E.C., Stuelten, C., *et al.* (2011) Protection from Obesity and Diabetes by Blockade of TGF-β/Smad3 Signaling. *Cell Metabolism*, **14**, 67–79.

31. Edwards, J.P., Thornton, A.M. and Shevach, E.M. (2014) Release of Active TGF-β1 from the Latent TGF-β1/GARP Complex on T Regulatory Cells Is Mediated by Integrin β 8. *The Journal of Immunology*, **193**, 2843–2849.

32. Munger, J.S., Huang, X., Kawakatsu, H., Griffiths, M.J.D., Dalton, S.L., Wu, J., Pittet, J.-F., Kaminski, N., Garat, C., Matthay, M.A., *et al.* (1999) A Mechanism for Regulating Pulmonary Inflammation and Fibrosis: The Integrin αvβ6 Binds and Activates Latent TGF β1. *Cell*, **96**, 319–328.

33. Kaur, S., Elkahloun, A.G., Arakelyan, A., Young, L., Myers, T.G., Otaizo-Carrasquero, F., Wu, W., Margolis, L. and Roberts, D.D. (2018) CD63, MHC class 1, and CD47 identify subsets of extracellular vesicles containing distinct populations of noncoding RNAs. *Scientific Reports*, **8**,

2577.

34. Iraci, N., Leonardi, T., Gessler, F., Vega, B. and Pluchino, S. (2016) Focus on Extracellular Vesicles: Physiological Role and Signalling Properties of Extracellular Membrane Vesicles. *International Journal of Molecular Sciences*, **17**, 171.

35. McQuillan, R., Leutenegger, A.L., Abdel-Rahman, R., Franklin, C.S., Pericic, M., Barac-Lauc, L., Smolej-Narancic, N., Janicijevic, B., Polasek, O., Tenesa, A., *et al.* (2008) Runs of Homozygosity in European Populations. *American Journal of Human Genetics*, **83**.

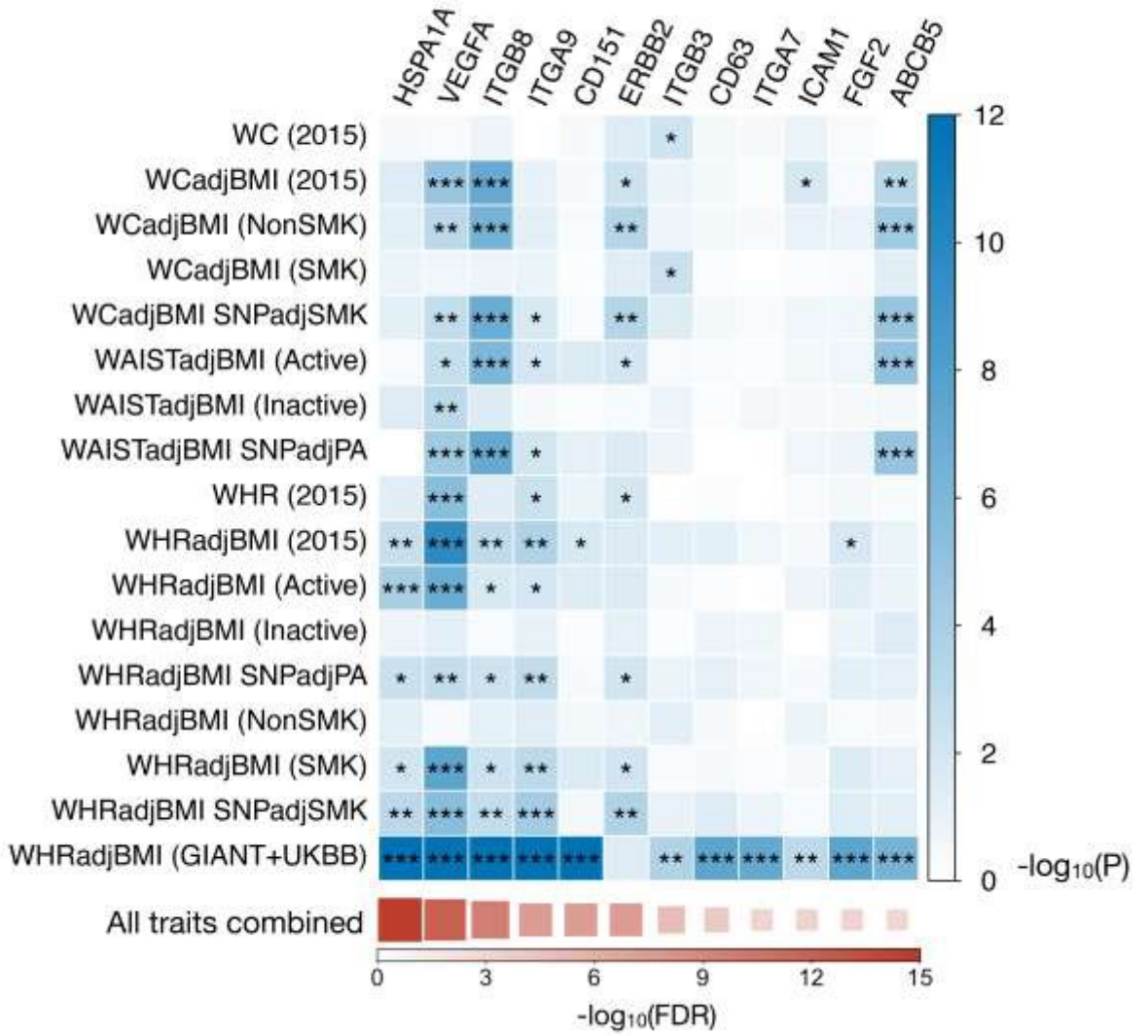
36. Belonogova, N.M., Svishcheva, G.R., van Duijn, C.M., Aulchenko, Y.S. and Axenovich, T.I. (2013) Region-Based Association Analysis of Human Quantitative Traits in Related Individuals. *PLoS ONE*, **8**, e65395.

37. Aulchenko, Y.S., Ripke, S., Isaacs, A. and van Duijn, C.M. (2007) GenABEL: an R library for genome-wide association analysis. *Bioinformatics*, **23**, 1294–1296.

38. Haller, T., Kals, M., Esko, T., Magi, R. and Fischer, K. (2015) RegScan: a GWAS tool for quick estimation of allele effects on continuous traits and their combinations. *Briefings in Bioinformatics*, **16**, 39–44.

39. Bulik-Sullivan, B.K., Loh, P.-R., Finucane, H.K., Ripke, S., Yang, J., Patterson, N., Daly, M.J., Price, A.L. and Neale, B.M. (2015) LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*, **47**, 291–295.

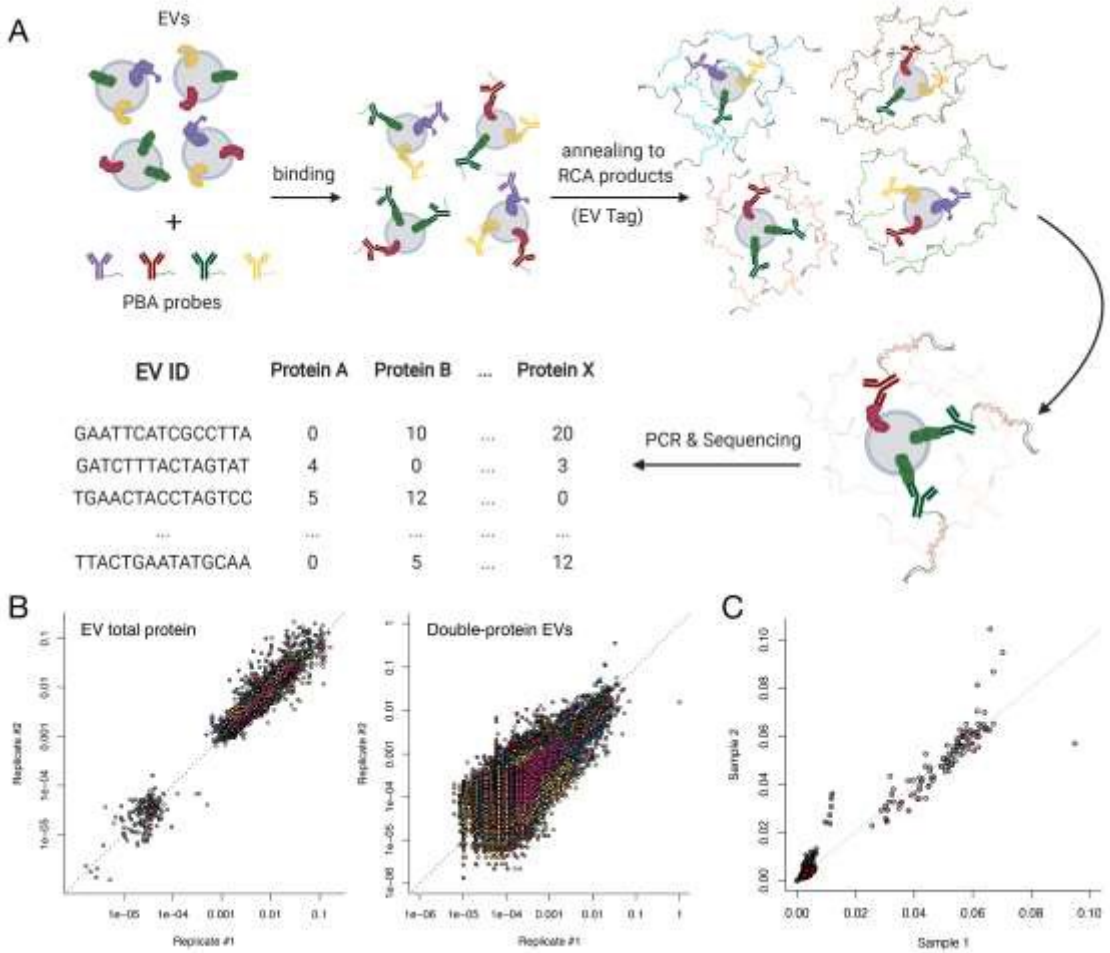
Legends to Figures



485

486 **Figure 1. Coding genes for the measured EV proteins that are associated with obesity-related**
487 **traits.** We performed multi-trait genome association analysis using summary association statistics of 17
488 traits from the GIANT consortium, which resulted in 18 (17 univariate and a multivariate analyses)
489 association p-values for each gene. Only the genes with $FDR_{Multivariate} < 0.01$ are shown and ordered by
490 -log10($FDR_{Multivariate}$) value in descending order. For the association with each trait, -log10(P) values are
491 colored in blue, and the associations with FDR < 0.05, 0.01, and 0.001 are marked by *, **, and ***,
492 respectively.

493



494

495 **Figure 2. Identifying surface proteomics profile of Extracellular Vesicles. (A)** Schematic
496 representation of Proximity Barcoding Assay (PBA). EVs are firstly incubated with PBA probes
497 (containing a protein tag), then each EV is allowed to hybridize with a unique RCA product (including
498 a unique EV tag), followed by enzymatic extension, protein tag on the same EV is then incorporated
499 with its EV tag, which can be sequenced after polymerase chain reaction. From the final pool of reads,
500 the surface proteins are quantified for millions of individual EVs. **(B)** EV proteomic profile of two
501 independent testing individuals. Percentages of EVs that carry each single protein and double-protein
502 combination are robustly measured and consistent within the two independent samples. **(C)** EV
503 proteomic profile of 96 ORCADES samples from two experiments (replicates 1 and 2). Each point
504 stands for a specific type of EV (left panel for the abundance of EVs with at least one kind of protein,
505 and the right panel for the abundance of EVs with at least two kinds of proteins). Different samples are
506 color-coded.

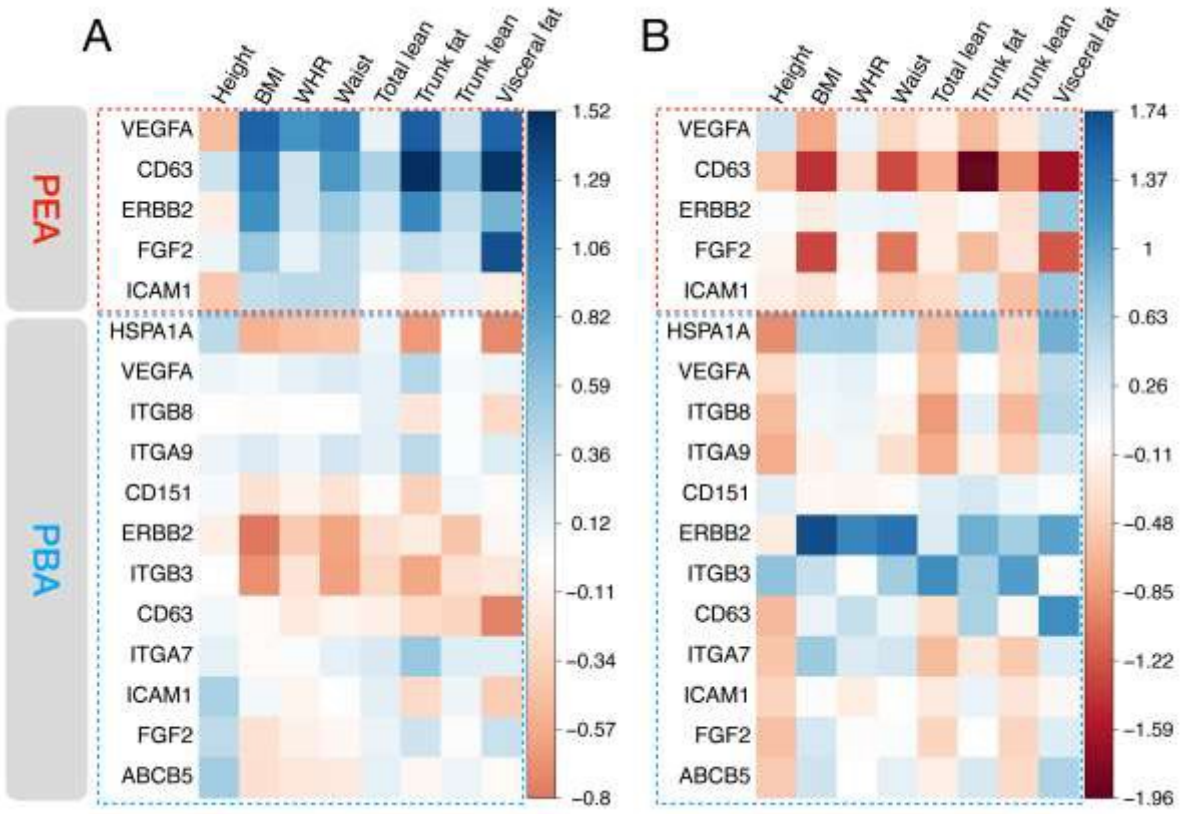
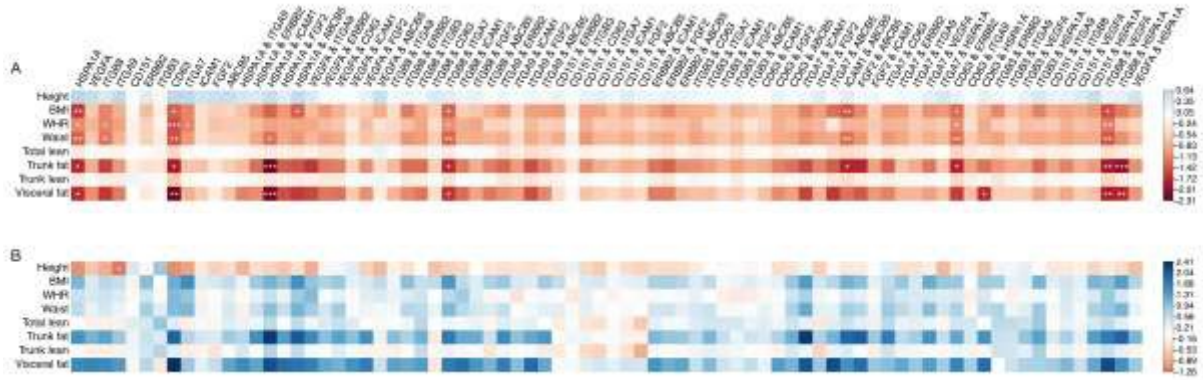


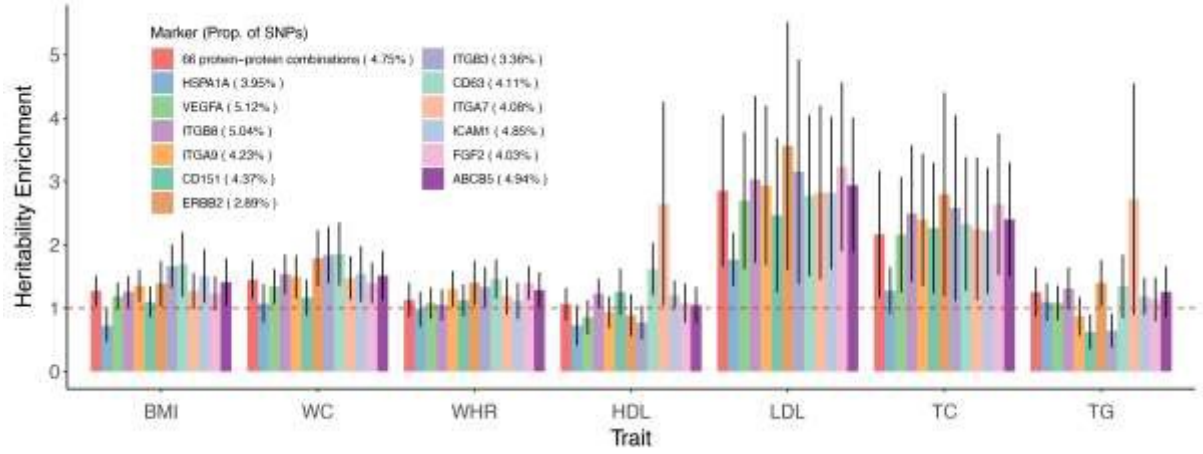
Figure 3: Associations between total protein level and anthropometric measurements in 96 ORCADES samples. In panels **A** and **B**, the upper part shows the effect of total protein level measured by PEA (Olink panels), and the lower part shows the effect of total protein quantified by PBA. Proteins in red dotted boxes are the markers of interest that overlap with Olink proteins. **(A)** The main effect of total protein count on seven body fat distribution-related traits, adjusted for sex and age. **(B)** Sex-by-protein interaction effect. Males are coded as 1 while females as 0, i.e., positive interaction (in blue) represents that the overall protein effects are weaker in men than in women. The effects (Beta) are color-coded.



518

519 **Figure 4: Association between the abundance of specific EVs and anthropometric**
520 **measurements in 96 ORCADES samples. (A)** The main effect of the abundance of specific EVs on
521 seven body fat distribution-related traits, adjusted for sex and age. **(B)** Sex-by-EV interaction effect.
522 Males are coded as 1 while females as 0, i.e., positive interaction (in blue) represents that the overall
523 EV effects are weaker in men than in women. The effects (Beta) are color-coded; Significance with
524 overall FDR thresholds of 0.05, 0.1, and 0.2 are marked by ***, **, and *.

525



526

527 **Figure 5. Heritability enrichment of EV-associated SNPs.** Heritability enrichment estimation of
528 EV-associated SNPs for each protein marker. Each bar represents heritability enrichment on an
529 obesity-related trait, and the error bars represent standard errors. The red dashed line indicates no
530 enrichment.