Extreme phenotypes define epigenetic and metabolic signatures in cardiometabolic syndrome.

Denis Seyres*[1,2,3], Alessandra Cabassi[4], John Lambourne[1,2,3], Frances Burden[1,2,3], Samantha Farrow[1,2,3], Harriet McKinney[2], Joana Batista[2], Carly Kempster[2], Maik Pietzner[5], Oliver Slingsby[6,7], Thong Huy Cao[6,7], Paulene Quinn[6,7], Luca Stefanucci[2,3,8], Matthew C Sims[2,3,9], Karola Rehnstrom[2], Claire Adams[10], Amy Frary[2], Bekir Erguener[11], Roman Kreuzhuber[2,12], Gabriele Mocciaro[13], Michael Allison[14], Simona D'Amore[14,15,16], Albert Koulman[5,17,18,19], Luigi Grassi[1,2,3], Julian L Griffin[13], Leong Loke Ng[6,7], Adrian Park[14], David B Savage[10], Claudia Langenberg[5], Christoph Bock[11,20,21], Kate Downes[1,2,22], Michele Vacca[10,13], Paul DW Kirk*[4,23], Mattia Frontini*[1,3,8]

1. National Institute for Health Research BioResource, Cambridge University Hospitals, Cambridge Biomedical Campus, Cambridge, United Kingdom
2. Department of Haematology, University of Cambridge, Cambridge, Cambridge Biomedical Campus, United Kingdom
3. NHS Blood and Transplant, Cambridge Biomedical Campus, Cambridge, United Kingdom
4. MRC Biostatistics Unit, University of Cambridge, Cambridge, Cambridge Biomedical Campus, United Kingdom
5. MRC Epidemiology Unit, University of Cambridge, Cambridge, UK.
6. Department of Cardiovascular Sciences, University of Leicester, Glenfield Hospital, Leicester, United Kingdom
7. National Institute for Health Research Leicester Biomedical Research Centre, Glenfield Hospital, Leicester, United Kingdom
8. British Heart Foundation Centre of Excellence,Cambridge Biomedical Campus, United Kingdom
9. Oxford Haemophilia and Thrombosis Centre, Oxford University Hospitals NHS Foundation Trust, NIHR Oxford Biomedical Research Centre, Oxford, UK
10. Metabolic Research Laboratories, Wellcome Trust-Medical Research Council Institute of Metabolic Science, University of Cambridge, Cambridge CB2 0QQ, UK
11. Ludwig Boltzmann Institute for Rare and Undiagnosed Diseases, Vienna, Austria. cbock@cemm.oeaw.ac.at.
12. European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, UK
13. Department of Biochemistry and the Cambridge Systems Biology Centre, University of Cambridge, The Sanger Building, 80 Tennis Court Road, Cambridge, CB2 1GA, UK.
14. Addenbrooke's Hospital, Cambridge University Hospitals NHS Foundation Trust, Cambridge, United Kingdom
15. Department of Medicine, Aldo Moro University of Bari, Piazza Giulio Cesare 11, 70124 Bari, Italy.
16. National Cancer Research Center, IRCCS Istituto Tumori 'Giovanni Paolo II', Viale Orazio Flacco, 65, 70124 Bari, Italy.

17. MRC Elsie Widdowson Laboratory, Cambridge, United Kingdom.
18. National Institute for Health Research Biomedical Research Centres Core Nutritional Biomarker Laboratory, University of Cambridge, Addenbrooke's Hospital, Cambridge, United Kingdom.
19. National Institute for Health Research Biomedical Research Centres Core Metabolomics and Lipidomics Laboratory, University of Cambridge, Addenbrooke's Hospital, Cambridge, United Kingdom.
20. CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria. cbock@cemm.oeaw.ac.at.
21. Department of Laboratory Medicine, Medical University of Vienna, Vienna, Austria. cbock@cemm.oeaw.ac.at.
22. East Midlands and East of England Genomic Laboratory Hub, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK
23. Cambridge Institute of Therapeutic Immunology & Infectious Disease (CITIID), Jeffrey Cheah Biomedical Centre, Cambridge Biomedical Campus, University of Cambridge, Puddicombe Way, Cambridge CB2 0AW, UK

*To whom correspondence should be addressed. E-mail: Denis Seyres: ds777@medschl.cam.ac.uk, Paul DW Kirk: paul.kirk@mrc-bsu.cam.ac.uk, Matia Frontini: mf471@cam.ac.uk

Abstract

Providing a molecular characterisation of cardiometabolic syndrome (CMS) could improve our understanding of its pathogenesis and pathophysiology, and provide a step toward the development of better treatments. To this end, we performed a deep phenotyping analysis of 185 blood donors, 10 obese, and 10 lipodystrophy patients. We analysed transcriptomes and epigenomes of monocytes, neutrophils, macrophages and platelets. Additionally, plasma metabolites including lipids and biochemistry measurements were quantified.

Multi-omics integration of this data allowed us to identify combinations of features related to patient status and to order the donor population according to their molecular similarity to patients. We also performed differential analyses on epigenomic, transcriptomic and plasma proteomic data collected from obese individuals before and six months after bariatric surgery. These analyses revealed a pattern of abnormal activation of immune cells in obese individuals and lipodystrophy patients, which was partially reverted six months after bariatric surgery.

Introduction

Cardiovascular disease (CVD) is the primary cause of death worldwide (17.9 million deaths in 2016, 31% of all deaths)[1] and the ever increasing number of overweight or obese individuals places a burden of hundreds of billions of dollars on healthcare systems each year[2,3]. Risk for CVD and type 2 diabetes (T2D) onset is increased by cardiometabolic syndrome (CMS)[4], a cluster of interrelated abnormalities (obesity, dyslipidemia, hyperglycemia, hypertension and non-alcoholic fatty liver disease[5]). These abnormalities have overlapping components which include abdominal obesity, high triglycerides, high LDL cholesterol, high fasting blood glucose, elevated blood pressure, decreased HDL cholesterol and low-grade inflammation[6–8]. Evidence of increased prevalence of peripheral vascular diseases, coronary artery disease[9] and myocardial infarctions as well as cerebro-vascular arterial diseases and stroke[10] in individuals developing CMS has also been shown[11]. The therapeutic approaches to mitigate their effect include weight loss strategies[12], lipid lowering drugs[13], antiplatelet therapies[14], glucose lowering medications[15,16] and anti-inflammatory therapies[17]. Several criteria to evaluate CVD risk have appeared based on different parameters and are recommended by several organisations: National Cholesterol Education Program and Adult Treatment Panel III (NCEP/ATPIII), world health organization (WHO), American Association of Clinical Endocrinologists (AACE) and International Diabetes Foundation (IDF)[18]. Additionally, indexes based on insulin and fasting glucose (Homeostatic Model Assessment for Insulin Resistance (HOMA-IR)[19]) or free fatty acid and insulin (Adipose Tissue Insulin Resistance (AT-IR)[20]) indexes are used to evaluate insulin resistance. Fatty liver index[21] and FIB-4 score[22] are used to evaluate liver health and several scores such as Framingham CVD score[23], JBS3 CVD score and QRisk2 score[24] have been proposed to evaluate lifetime cardiovascular risk.

The relationship between cardiometabolic health and body weight is complex[25]. Several studies have found that CVD risk varies between individuals of similar body mass index (BMI) and depends also on fat distribution and metabolic profile[26–30]. Adipose tissue (AT) plays a major role in metabolic disorders. AT acts as an active

endocrine organ[31,32]: it secretes lipids, adipokines and other metabolites and is central in whole-body homeostasis. Pathological AT remodelling (such as hypo or hypertrophia) causes hypoxia, fibrosis and low-grade inflammation[33]. The latter is accompanied by the production of cytokines and other pro-inflammatory signals[34,35], which triggers pro-inflammatory responses of immune cells[31]. Pathological AT is involved in cardiometabolic impairment[36–39] and is an important mechanism for the adverse effects of adiposity on the vessel wall[40]. Whilst the participation of platelets and neutrophils in thrombosis and of macrophages in plaque formation are well established[41–43], their role in the atherogenesis and CVD has begun to be appreciated only recently[44]. As a consequence, macrophage recruitment in the AT is observed[45], as well as changes in the ratio of pro-inflammatory and anti-inflammatory macrophages (M1 and M2, respectively)[46]. Furthermore, prolonged exposure to low-grade inflammation modifies cell functional phenotypes in monocytes, an effect named trained immunity[47], in platelets[48,49] and neutrophils[50,51].

Here, we present the results of applying multi-omic data integration approach to characterise the molecular hallmarks of CMS. We performed a deep phenotyping analysis of 185 blood donors, 10 obese, and 10 lipodystrophy patients. Lipodystrophy comprises heterogeneous disorders characterized by loss of fat tissue, mainly from subcutaneous compartment and occasionally affecting visceral fat which can be partial, localized, or generalized[52]. A generalized fat loss also leads to metabolic disorders, including insulin resistance, T2D, hyperlipemia and hepatic disease such as nonalcoholic fatty liver disease (NAFLD)[53]. Our group of lipodystrophy patients have all inherited familial partial lipodystrophy: 2 had a mutation in Peroxisome proliferator-activated receptor gamma (PPAR-γ) gene and 8 in Lamin A/C (LMNA) gene. The data sources included: (i) plasma biochemistry, metabolomic and lipidomic; (ii) transcriptomic data from platelets, neutrophils, monocytes and macrophages; (iii) active chromatin as measured by histone 3 lysine 27 acetylation (H3K27ac) in neutrophils, monocytes and macrophages); and (iv) DNA methylation by reduced representation bisulfite sequencing (RRBS) and Illumina HumanMethylation450 BeadChip in neutrophils, monocytes and macrophages. We used penalised logistic regression to define disease signatures comprising features from multiple layers of molecular data, which enabled individuals

with extreme phenotypes (obese referred for bariatric surgery and lipodystrophy patients) to be discriminated from lean, metabolically healthy individuals. Predictions made using the disease signatures were generally in good agreement with predictions made on the basis of plasma biochemistry markers only, although there was stark disagreement for some individuals. The approach was validated by training our classifier using data from one patient group, and making predictions for the other group. Moreover, we found that features identified in the lipidomic layer were associated with known risk factors in a large independent cohort.

As a result of our integrative approach, we also obtained the molecular signatures at transcriptional, chromatin (H3K27ac), and DNA methylation levels. These helped to shed light on the molecular events that determine the changes in the functional phenotypes of platelet, neutrophils, monocytes and macrophages.

Our approach was deployed to characterise the same individuals six months after bariatric surgery which is an effective therapy to reduce weight and improve wellbeing for morbidly obese individuals[54,55]. Major benefits are a substantial weight loss in the first year following the intervention, with maximum loss reached around 6-8 months after surgery[56], and improvement of several clinical parameters (HbA1c, glucose and cholesterol levels, insulin resistance, and modulations of gut hormones[57]). It has also been shown that bariatric surgery modifies metabolite abundance in the first year following surgery[58]. Moreover, bariatric surgery has been shown to affect DNA methylation patterns and gene expression[59]. Additionally, to reduce stomach size, bariatric surgery also reduces inflammatory markers[60,61]. We investigated and determined the changes that occur in the neutrophils, monocytes, macrophages and platelets. While the gene expression showed dramatic changes, especially in neutrophils and platelets more modest differences were observed in regulatory elements and almost none in methylation profiles. Plasma proteome analysis allowed us to have some insight on the changes in other tissues and organs whilst neutrophils and platelets cell function assays results indicated reduced ability to adhere, the key initial step for their activation.

Results

Metabolic signatures in obese individuals and lipodystrophy patients.

To determine the metabolic health of the different groups, we collected anthropometric characteristics: age and body weight (BW) and performed plasma biochemistry assays for the following: leptin, adiponectin, insulin, free fatty acids (FFA), glucose (GLC), triglycerides (TG), total cholesterol (TC), high density lipoprotein cholesterol (HDL-C), low-density lipoprotein (LDL-C)), activity of alanine and aspartate amino-transferases (ALT and AST, respectively) and high-sensitivity C-reactive Protein (hsCRP). Additionally, we computed leptin-adiponectin ratio (LAR), Homeostatic Model Assessment for Insulin Resistance (HOMA-IR) and adipose tissue insulin resistance (AT-IR) indexes (**Table 1; Table S1**). To investigate the combination of these parameters, we performed a principal component analysis (PCA) to reduce a large set of variables into a smaller set that still contains most of the information of the larger set. PCA showed that obese and lipodystrophy patients and WP10 individuals are distributed over distinct but partially spanning dimensions(**Fig. 1A**). The first two components (PC1 and PC2) distinguished well the different groups; in particular obese individuals were separated from WP10 participants along PC1 and lipodystrophy patients were separated from WP10 participants along PC2. Analysis of loading and contribution for each parameter analysed (**Fig. 1B**) indicated that weight contributed to both global variance (59.2%, **Fig. 1B** color scale) and to PC1 (**Fig. 1B,** arrow length and direction), as expected, followed by AGE (12.6%), LAR (7.4%) and HOMA-IR score (5.8%). ALT and AST explained most of the variance along PC2 (45.6% and 29.8% respectively), followed by HOMA-IR (8.6%) and AT-IR (4.2%) indexes, AGE (1.6%) and GLC (2.7%). Obesity has been shown to have a profound impact on plasma metabolites[62,63]. Previous studies have also shown that metabolite patterns were affected in patients suffering from lipodystrophy[64,65]. To determine which metabolites were present, all plasma samples were analysed on the Metabolon platform (Material and Methods). This led to the identification and quantification of 988 metabolites. To identify and characterise groups of metabolites whose levels were correlated across samples, we

	WP10 (n=185)	Lipodystrophy (n=10)	Obese (n=10)	Post surgery (n=10)
ALT (U/L)	34.5 (12)	56 (11.5)	35.7 (11.5)	36.1 (10.5)
AST (U/L)	25.5 (7)	39 (15.5)	22.6 (3.5)	18.9 (5)
hsCRP (mg/L)	1.9 (2)	2.3 (2.4)	7.3 (5.7)	2.9 (0.65)
Insulin (pmol/L)	118.4 (101)	261.7 (198.5)	190.6 (95.2)	178.7 (58.8)
GLC (mmol/L)	5.4 (1)	8.2 (6.3)	5.3 (0.6)	5.3 (0.95)
Leptin (ng/ml)	14.2 (13.4)	7.5 (7.9)	74.1 (21.2)	29.9 (23.4)
Adiponectin (µg/ml)	10.1 (5.6)	3.2 (1.6)	5.9 (2.1)	6.4 (2.3)
FFA (µmol/L)	189.3 (154.5)	259.6 (273.3)	293.2 (143.9)	232.2 (235.3)
TG (mmol/L)	1.5 (1)	5.6 (7.8)	1.8 (0.7)	1 (0.3)
HDL-C (mmol/L)	1.5 (0.6)	0.8 (1)	1.2 (0.2)	1.3 (0.3)
LDL-C (mmol/L)	2.9 (1.1)	1.7 (0.5)	2.4 (0.9)	2.5 (1)
TC (mmol/L)	5.2 (1.5)	4.2 (1.6)	4.1 (1.3)	4.5 (1.1)
BW (kg)	76 (18.8)	NA	138 (33.5)	NA
BMI (kg/m^2)	26.3 (5)	NA	45 (3.25)	NA
AGE(years)	57.2 (15)	43.2 (6)	46.3 (16)	44.4 (21)
LAR	1.7 (1.6)	2.2 (1.4)	13.7 (6.8)	5.5 (5.6)
HOMA-IR	4.2 (3.6)	13 (17)	7 (3.9)	8.6 (1.9)
AT-IR	2.5 (2.3)	8.4 (9.4)	7.1 (2.7)	4.7 (5.2)

Table 1 - Summary table of anthropometric and biochemical parameters.

performed a weighted co-expression network consensus analysis (WGCNA)[66]. Using the entire dataset, we identified 16 consensus modules (M1 to M16; **Table S2**). We did not identify a significant enrichment for a specific biological pathway for any of

these modules. Next, we determined if there were any associations between these modules and the results of the plasma biochemistry assays. Of the 208 tested associations, 11 showed significant correlation with the results of the plasma biochemistry assays in the combined patient group (FDR adjusted Fisher p values < 5%; 6 were positively correlated, 5 were negatively correlated; **Fig. 1C**) and 51 showed correlation with the results of the plasma biochemistry assays in the WP10 cohort (FDR adjusted Fisher p values < 5%; 27 were positively correlated, 24 were negatively correlated; **Fig. 1D**). Of these, 3 were shared: M6 was positively correlated with module TG both in patients and donors. whereas module M2 was negatively correlated with TG and positively correlated with HDL-C in patients and reversely in donors. This suggests that, depending on the individual status, M2 metabolites participate differently to TG and HDL-C concentrations.

In order to determine which of these modules could be associated with disease status, we analysed the eigen-metabolites adjacencies. Using patient samples, consensus modules form two clusters, C1 and C2 (**Fig. 1E**). To determine if these identify the two groups of patients, we represented the average eigen-metabolite value for each cluster together with the PC1 from **Fig. 1A**. This analysis showed that each group of patients could be identified by its distinct metabolic signature (**Fig. 1F**). The metabolites found in cluster C1 were enriched in the following pathways: alanine, aspartate and glutamate metabolism (p value=$4.3*10^{-5}$; FDR=$3*10^{-3}$; Fisher's Exact Test), phenylalanine metabolism (p value=$2*10^{-2}$; FDR=$3*10^{-2}$), nitrogen metabolism (p value=$9.5*10^{-5}$; FDR=$3*10^{-3}$), aminoacyl-tRNA biosynthesis (p value=$2.3*10^{-4}$; FDR=$6*10^{-3}$) and citrate cycle (TCA cycle; p value=$2*10^{-2}$; FDR=$3*10^{-2}$). The metabolites found in cluster C2 were enriched in the following pathways: aminoacyl-tRNA biosynthesis (p value=$1.5*10^{-4}$; FDR=$1*10^{-2}$), cysteine and methionine metabolism (p value=$1*10^{-3}$; FDR=$4*10^{-2}$) and glycine, serine and threonine metabolism (p value=$3*10^{-3}$; FDR=$8*10^{-2}$). No pathways were found enriched in metabolites forming cluster C3. Cluster C4 metabolites were enriched in aminoacyl-tRNA biosynthesis (p value=$6.8*10^{-11}$; FDR=$5.4*10^{-9}$), nitrogen metabolism (p value=$2.6*10^{-6}$; FDR=$1*10^{-4}$), alanine, aspartate and glutamate metabolism (p value=$8.6*10^{-5}$; FDR=$2*10^{-3}$), glycine, serine and threonine metabolism (p value=$7.8*10^{-4}$; FDR=$1.5*10^{-2}$), phenylalanine, tyrosine and

tryptophan biosynthesis (p value=$1*10^{-3}$; FDR=$2.4*10^{-2}$) and phenylalanine metabolism (p value=$3*10^{-3}$; FDR=$4*10^{-2}$). When the same pathway was found enriched, i.e. aminoacyl-tRNA biosynthesis or glycine, serine and threonine metabolism, it was due to different metabolites in the same pathway being found in different clusters (**Fig. S1C**). Moreover, clusters C1 and C2 were more patient specific compared to the two clusters that we observed within WP10 donors adjacencies heatmap, C3 and C4 (**Fig. S1A**). These two clusters were defined by different metabolites and were unable to discriminate between obese and lipodystrophy patients (**Fig. S1B**). Of the 37 metabolites associated previously associated with BMI[63], 15 were found to overlap with cluster C1 (p value=0.001; hypergeometric test). These belong to modules: M1 (Carnitine, Glycine, Tyrosine, Asparagine, N-acetylglycine, Kynurenine and Gamma-glutamyltyrosine), M7 (1-Eicosadienoylglycerophosphocholine), M9 (3-methyl-2-oxobutyrate, Lactate, 4-methyl-2-oxopentanoate, Hexanoylcarnitine, Propionylcarnitine and 3-methyl-2-oxovalerate) and M14 (Phenylalanine and Histidine). The only phenotypical association was found in module M1, the one having with the largest number of components shared between C1 and C4 was LDL-C (p value=0.02 and p value=0.01; Fisher's exact test; **Fig. 1C and Fig.1D**). However, M1 was negatively correlated with LDL-C in patients whereas, in WP10 donors, it was positively correlated with it. In summary, analysis of serum metabolites abundance revealed distinct clusters identifying patient groups, as well as, biological pathways known to be associated with AT dysfunction in individuals having high CMS risk.

Transcriptional and epigenetic signatures in obese and lipodystrophy patients for 3 innate immune cell types and platelets.

Next, we sought to determine if the metabolic changes observed in plasma were accompanied by changes in the transcriptional and epigenetic signatures of innate immune cells (neutrophils, monocytes macrophages) and platelets (**Fig. 2A**). We characterised the transcriptome by ribo depleted RNA-sequencing (RNA-seq), the genome-wide distribution of histone 3 lysine 27 acetylation (H3K27ac), a marker of active promoters and transcriptional enhancers, by chromatin immunoprecipitation (CHIP-seq) and DNA methylation by reduced representation bisulfite sequencing

(RRBS). First, we determined which, amongst the volunteers recruited as controls, were the metabolically healthy individuals (hereafter lean, selected using the following parameters: BMI < 25, glycaemia (GLC) <5.4 mmol/L, TG <1.7 mmol/L, LDL-C <2.59 mmol/L, HDL-C >1 mmol/L for men and >1.3 mmol/L for women, HOMA-IR score< 2.2) and used these in all the following comparisons (**Table S7**).

The comparison between obese and lean individuals (**Fig. 2B)** led to the identification of 42 genes differentially expressed (DEG) in macrophages (26 up and 16 down regulated in obese individuals, **Table S8**), 77 in monocytes (55 up and 22 down regulated, **Table S9**), 38 in neutrophils (13 up and 25 down regulated, **Table S10**) and 135 in platelets (92 up and 43 down regulated, **Table S11**); at a FDR threshold of 5%. We also identified limited changes in H3K27ac peaks (**Fig. 2B**): with 299 (298 up and 1 down regulated) differentially acetylated regions (DAcR) above 5% FDR in macrophages (**Table S12**), 16% of these were located on gene promoters or gene body, 31% intergenic and 53% in introns. Only 3 DAcR were identified in neutrophils (**Table S13**) and none in monocytes. DNA methylation analysis found 15 differentially methylated CpG islands (0.5% FDR) in macrophages (**Table S15**); 17 in monocytes (**Table S16**) and 18 in neutrophils (**Table S17**).

The comparison between lipodystrophy patients and lean individuals (**Fig. S2C)** identified 125 DEG in macrophages (115 up and 10 down regulated in lipodystrophy, **Table S8**), 8 in monocytes (2 up and 6 down regulated, **Table S9**) and 5 in neutrophils (2 up and 3 down regulated, **Table S10**). No DEG were identified in platelets. Few DAcR were observed in macrophages and monocytes: 17 DAcR (1 up and 16 down) in the former (**Fig. S2C and Table S12**), 1 of these was located on gene promoter, 12 (70)% were intergenic and 4 (23%) in introns and one in the latter (down regulated; intergenic). No DAcR were identified in neutrophils. DNA methylation analysis found 20 differentially methylated CpG islands in macrophages (**Table S15**); 60 in monocytes (**Table S16**) and 44 in neutrophils (**Table S17**).

The comparison between obese individual and lipodystrophy patients (**Fig. S2B)** identified 4 DEG in macrophages (2 up and 2 down regulated in lipodystrophy, **Table S8**), 40 in monocytes (22 up and 18 down regulated, **Table S9**), 1 upregulated gene each in neutrophils (**Table S10**) and in platelets (**Table S11**). We observed 1,764 DAcR in macrophages (all hyper-acetylated in lipodystrophy; **Fig. S2B and Table**

S12), 22% of these were located on gene promoters or gene body, 33% were intergenic and 45% in introns. We also observed 1,766 DAcR in monocytes (1,098 up and 668 down; **Fig. S2B and Table S13**), 38% of these were located on gene promoters or gene body, 13% were intergenic and 49% in introns. Of these, only 50 overlapped with those observed in macrophages.

In macrophages, DNA methylation analysis found 3 CpG islands differentially methylated (2 up and 1 down, **Table S15**), 90 in monocytes (49 up and 41 down, **Table S16**) and 3 in neutrophils (1 up and 2 down, **Table S17**).

To gain insight into the changes observed in gene expression, we performed functional annotation by gene ontology (GO) terms enrichment analysis. In the comparison between obese and lean individuals (**Fig. 2C**), we found an enrichment for GO terms related to phagocytic and degranulation activities, as well as markers of cardiometabolic risk in macrophages (**Table S18)**. In monocytes, down-regulated DEG were enriched in GO terms related to programmed cell death and ion homeostasis; whereas up-regulated DEG were enriched for GO terms related to inflammatory response (**Table S19**). In neutrophils, down-regulated DEG showed enrichment for GO terms related to vesicle trafficking, protein sulfation and regulation of lipid droplet morphology and activity (**Table S20**). In platelets, down-regulated DEG showed enrichment for GO terms related to cholesterol biosynthesis and C-type lectin receptor signaling pathway; whereas up-regulated DEG were enriched for GO terms related to bicellular tight junction mechanism and palmitoylation in platelet activation and thrombus formation (**Table S21**). Genes associated to DAcR with an increased acetylation in macrophages showed an enrichment in GO terms related to response to inflammation, infection and cell adhesion (**Table S22**). No functional enrichment was performed for monocytes and neutrophils due to the low numbers of genes associated with DAcR.

Multi omic signatures of obesity and partial lipodystrophy and their use in prediction of cardiometabolic risk

Multivariable regression approaches for variable selection have provided an effective means to integrate multiple omics layers and elucidate disease signatures[67,68]. We used one such approach to integrate RNAseq H3K27ac histone modification, DNA

methylation, metabolic and lipidomic datasets. We identified a training set comprising 6 donors from the WP10 group with values below this combination of thresholds: BMI <25, GLC <5.4 mmol/L, TG <1.7 mmol/L, LDL-C <2.59 mmol/L, HDL-C >1 mmol/L for men and >1.3 mmol/L for women, HOMA-IR index <2.2, 6 obese individuals and 10 lipodystrophy patients for which we had complete measurements on all omic data layers, in monocytes and in neutrophils. Using this training set, we used elastic net penalised logistic regression to identify putative signatures associated with an increased probability of belonging to the obesity and/or lipodystrophy groups (**Fig. 3A**). The values taken by the variables selected into each signature defined patterns characterising the different groups (WP10, all patients, obese and lipodystrophy) (**Fig. 3B**; **Table S33**). When comparing features selected for each comparison, i.e. obese versus lean and lipodystrophy versus lean, 27 genes in monocytes (pvalue=$1.5*10^{-26}$, hypergeometric test) and 6 genes (p value=$1*10^{-3}$, hypergeometric test) in neutrophils were also differentially expressed between obese and lean. Two and one selected genes were also differentially expressed between lipodystrophy and lean in monocytes and in neutrophils respectively. Genes within 10 kilobases of a H3K27ac peak or within 10 kilobases of a DNA methylation site identified using the variable selection did not return significant enrichment for gene ontology terms (above 5% FDR threshold). The variable selection analysis identified groups of features that were, together, predictive of patient status, whereas the differential analysis identified individual variables that were different between each of the patient groups and the lean individuals.

We used the variables selected for each signature, together with the biometric variables, to construct multivariable logistic regression models to predict whether an individual was a patient or donor (**Fig. 3C**). Although rigorous validation of the full predictive model was hampered by a paucity of other cohorts for which multiple omics datasets were available, we observed that training the model using just the obese and lean groups allowed us to correctly identify the lipodystrophy patients as having a CMS profile (**Fig. S3**); and vice versa (**Fig. S4**). Further validation was obtained using the lipidomic layer signature. We identified selected lipids that were also measured in two other studies: a subset of 1,507 participants of the Fenland study[69] which is a population-based cohort of 12,345 volunteers without diabetes

born between 1950 and 1975 and recruited within the Cambridgeshire region between 2005 and 2015, and a biopsy-confirmed nonalcoholic steatohepatitis (NASH) cohort comprising 42 individuals[70]. In the Fenland cohort (**Fig. 3D**), we found TG (52:2) and TG (50:1) to be positively associated with several risks factors, such as elevated glycaemia, increased fasting insulin level, HOMA-IR and liver indexes, HbA1c, leptin, LDL-C, hsCRP, TG, BMI, fat mass, ALT and ferritin. Conversely, they were inversely associated with adiponectin and HDL-C. Except for adiponectin and HDL-C, PC (40:7), PC (38:7), PC-O (36:2), PC (38:6) and PC (35:2) were inversely associated with all the remaining factors. To assess the specificity of the selected lipids, we repeated the analysis with 10 lipids that were not selected into the signature (**Fig. 3E**). We found far fewer associations were found to be significant. We performed the same analysis using data from the NASH cohorts (**Fig. S5**), as well as data from the present study (**Fig 3F**). Although these studies were insufficiently powered to allow us to identify significant associations after correcting for multiple testing, the effect estimates were similar to those obtained in the Fenland study. In summary, we managed to define molecular signatures representing CMS abnormalities overlapping components. We next sought if these signatures were reversible if obese individuals underwent important surgery which led stomach reduction.

Effect of bariatric surgery on transcriptional profile, epigenetic landscape and cell functions.

Bariatric surgery is an effective option for the management of extreme obesity and its comorbidities, including CMS risk[71], with well established long term benefits on weight loss, diabetes, hypertension and dyslipidemia[72]. Here, we investigated the effects of weight loss by bariatric surgery on the transcriptional and epigenetic profiles of innate immune cells and platelets, and on plasma proteins. To this end, a second blood sample was obtained six months after bariatric surgery and subjected to the same assays. Pairwise comparison of each biochemical parameter showed a decrease for LAR, TG, hsCRP, AT-IR and AST and an increase of HDL-C level (p values of: $7.22*10^{-6}$, $2.63*10^{-9}$, $4.98*10^{-4}$, $2.51*10^{-2}$, $1.48*10^{-3}$ and $1.86*10^{-3}$ respectively ; conditional multiple logistic regression ; **Table S1 for other**

comparisons). We next compared the transcriptional and epigenetic profiles in monocytes, neutrophils, macrophages and platelets before and after bariatric surgery. We identified, using paired analysis, 713 DEG in macrophages (403 down and 310 up regulated; **Table S8**); 2,081 in monocytes (1,204 down and 877 up regulated; **Table S9**); 3,564 in neutrophils (2,609 down and 955 up regulated; **Table S10**) and 2,741 in platelets (1,159 down and 1,582 up regulated; **Table S11**). No DAcR were found in macrophages (**Table S12**), 229 in monocytes (139 down and 90 up regulated; **Table S13**) and 788 in neutrophils (775 down and 13 up regulated; **Table S14**). RRBS analysis found 126 differentially methylated CpGs in macrophages (69 down and 57 up regulated; **Table S15**); 48 in monocytes (32 down and 16 up regulated; **Table S16**) and 55 in neutrophils (44 down and 11 up regulated; **Table S17**).

GO terms enrichment for DEG identified the following processes: ribosome formation and translation, platelet activation, fatty acids synthesis, lipids metabolism and transport, several immune related pathways (**Table S18**, **S19**, **S20** and **S21** for macrophages, monocytes, neutrophils and platelets, respectively). The results of the other possible comparisons (**Fig. S2**) are available in **Tables S8 to S17**. Next, we searched for those genes whose expression changed in obese individuals and reverted to the level observed in lean individuals after bariatric surgery. In macrophages we found 3 genes that after having been upregulated in obese individuals compared to lean individuals, returned to the same expression level after weight loss (MCEMP1, PHACTR1 and CTD-2135D7.2; p value=$1*10^{-2}$, hyper-geometric test). In monocytes, we found 7 genes with down-up-down profile (CLASP1, ATP11C, RALGAPA2, MTHFD2, GPRIN3, FAM129A and MCTP2; p value=$1*10^{-2}$, hyper-geometric test) and 4 genes with a up-down-up profile (HLA-DRB9, RMRP, LINC00899 and APEX1; p value=$5*10^{-2}$, hyper-geometric test). No genes with a down-up-down profile were found in neutrophils and only PCGF6 was found with a up-down-up profile (p value=$9*10^{-2}$; hyper-geometric test). In platelets we identified 39 genes with a down-up-down profile (C16orf72, G3BP2, PJA2, ZBED6, CBL, AKIRIN1, BSDC1, TC2N, FAM122B, SYTL4, ENDOD1, LYPLA1, F11R, SSFA2, UBN1, ATP9A, NDST1, DOK2, CLDN10, NUDT3, STIM1, GDI1, ST20-AS1, CDS2, LGMN, PROS1, ARF3, GTPBP1, TMPPE, KCMF1,

NBPF10, RP11-4O1.2, RP11-755F10.1, DANT2, DPP10, ITGB3, MLEC, RP11-667F9.1 and KIAA2012; p value=$7*10^{-19}$, hyper-geometric test) and 8 genes with a up-down-up profile (TM4SF19-AS1, MRAS, CXXC5, CAPG, A2M, CYB5A, CD52 and RP11-509J21.2; p value=$2*10^{-3}$, hyper-geometric test).

In addition, overlap between DEG and DAcR was found for neutrophils in obese versus post surgery (208 genes, p value=$4.9*10^{-30}$, hyper-geometric test) and lean versus post surgery (8 genes, p value=$7*10^{-3}$, hyper-geometric test). To obtain further insight into how bariatric surgery affects gene expression and signaling pathways in other tissues and organs, we investigated plasma protein levels before and after surgery. We quantified 3,098 plasma proteins; 604 of which were found to be differentially abundant (DAP, **Fig. 4C** and **Table S28**) above ordinal Q-value of $1*10^{-3}$. Proteins whose levels increased after bariatric surgery (72) were enriched in GO terms related to tight junction, WNT signalling, PI3K/AKT signalling, and sphingolipid signalling. Instead, proteins whose abundance decreased after surgery (532) were enriched in the following GO terms: cell cycle and DNA repair, ribosomal RNA metabolism and cell senescence, phagocytosis and T cell receptor signalling as well as FGF, IL2, VEGF and insulin signaling pathways (**Table S29**) in agreement with Albrechtsen and colleagues[73]. Plasma proteins can have different origins; to determine if any of the proteins identified could be linked to a specific tissue, we curated the GTEx project[74] database to extract tissue specific genes, these range from 286 in the heart left ventricle to 1286 in the spleen (**Table S30 and methods**). Tibial artery, heart atrial appendage and blood display an enrichment of tissue specific genes amongst DAP (p values: $2.2*10^{-2}$, $2.1*10^{-2}$ and $9*10^{-2}$, respectively; hyper-geometric test; **Table S30**). Of the 16 blood specific genes that encode for a DAP, 7 are also differentially expressed in at least one of the 4 studied cell types (**Fig. 4D**). SIRPB1, HBA2, ADGRE2 and RAC2 in monocytes, MYO1G in platelets, HIST1H3D and HIST1H2BC in neutrophils, monocytes, platelets and macrophages. The data generated in monocytes and macrophages also allowed us to explore the effect of bariatric surgery on trained innate immunity[75], as it has been shown that trained innate immunity could play a role in atherosclerosis[76,77]. We found two distinct overlaps: one (p value $5*10^{-2}$; t-test) between the genes associated with the top 500 regions with a gain in histone 3 lysine 4 trimethylation (H3K4me3) after β-glucan

treatment in Quintin and colleagues[75] and the DEG found comparing lean individuals and obese individuals. The second overlap (p value $2*10^{-4}$; t-test) between the genes associated with the top 500 regions with a gain in histone 3 lysine 4 trimethylation (H3K4me3) without treatment and the DEG found comparing obese individuals before and after bariatric surgery. All other overlaps were non-significant (**Table S31**).

We performed functional tests on neutrophils and platelets to determine if the changes observed at molecular levels resulted in changes in the functional phenotypes of these cells. After bariatric surgery, neutrophils showed a reduction in their ability to adhere when unstimulated and when subjected to a variety of stimuli (DTT, LBP, PAM3, PAF and fMLP) but not when treated with TNFalpha or PMA. These results were accompanied by a reduction in the cell surface levels of CD16 and CD32 but not CD66b, CD63, CD62L or CD11b (paired t-test, all result in **Table S32**). Alongside, we performed platelet functional tests which showed a reduction in P-selectin upon collagen stimulation, but not upon ADP or thrombin stimulations. These results were accompanied by a reduction in the cell surface levels of fibrinogen receptor (both CD61 and CD41a but not CD41b) and CD36, whereas no changes were observed for CD49b, CD42a, CD42b, CD29 and CD9 (paired t-test, all result in **Table S32**).

Discussion

Biochemical and Metabolic signatures.

We considered two groups of patients with extreme phenotypes associated with cardiometabolic syndrome. These two groups can be distinguished amongst the general population using anthropometric and plasma biochemistry parameters. PCA of these parameters showed that obese individuals were separateded along PC1, with weight explaining most of the variance. ALT and AST activities explained most of the variance along PC2 (45.6% and 29.8% respectively). These transaminases are known to be elevated in lipodystrophy patients[78]. Some WP10 donors overlapped either with obese or lipodystrophy patients (Fig. 1A) due to similarities either in weight or plasma biochemistry profile.

We used a network-based approach to determine the differences in metabolite abundances in the different groups. To increase statistical power, we merged the patient groups under the assumption that they share similar associations of metabolites and phenotypic traits. In the consensus analysis, we used a conservative approach, considering 988 metabolites Of these, 375 were assigned to 15 different modules and the remaining 613 were put in an ad hoc extra module because they did not show any correlation. Analysis of the correlation module matrix divided the modules into two different clusters for patients, C1 and C2. Together with the analysis performed using the results from PC1, in Fig.1E, we showed that these clusters represented each a patient group. However, these separations were based on CMS features rather than obesity and lipodystrophy specific features. C1 showed an enrichment for modules involved in alanine, aspartate and glutamate metabolism known to be associated with NAFLD[79,80], but they also showed enrichment for nitrogen and phenylalanine metabolisms, previously described in obesity[81,82]. The same functional annotations were retrieved in a large study on the effects of bariatric surgery on the metabolome[83]. C2 showed an enrichment for cysteine and methionine metabolism, and glycine, serine and threonine metabolism. These have been associated with NAFLD[84] and have also been shown to play a role in other CMS associated diseases such as obesity and T2D[85].

Transcriptional and epigenetic signatures.

The comparisons between lean/obese and lean/lipodytrophy (**Fig. 2A**) found a modest number of changes in terms of gene expression, active chromatin and DNA methylation (**Table S7**), however these were enough to clearly highlight inflammatory response and platelet activation related terms in GO enrichment analysis, thus confirming that these conditions modify the molecular phenotype of cell types involved in the development atherosclerosis and in thrombus formation. Similar results have already been reported for blood cells DNA methylation[86], while more extensive changes have been observed in adipose tissue[59].

The limited changes we observed can be, at least in part, explained by the absence of acute challenge when the samples were collected, as previously shown[87]. The largest number of changes observed in active chromatin, DAcR, was found, both in

monocytes and in macrophages, in the comparison between obese individuals and lipodystrophy patients, where the latter had in excess of 1,700 transcriptional enhancers (**Fig. S2B and Table S12**). However, this was not associated with changes of the same magnitude at transcriptional level, suggesting that the same transcriptional output can be achieved with different regulatory landscapes, as previously shown[88], but also that the same cell types could be primed differently in the two groups.

Multi omic signatures

We used variable selection approaches to identify signatures of disease, which we incorporated into a predictive mode. Within-study validation demonstrated the utility of the model for out of sample prediction (Fig. S3 and S4). We moreover showed that prioritized lipid species are associated with major cardiometabolic risk factors in the Fenland study. Among the selected lipids, previous studies have shown PC (38:6) to be reduced in models of liver damage, PC (36:2) to be reduced in obese mice livers[89], TG (50:1) and TG (52:2)[89] to be a product of de novo lipogenesis increased in NAFLD[69] and NASH[70]. Although we believe this represents the limits of the validation that can be performed using present datasets, we acknowledged that further multi-omic studies will be required to evaluate this model in external cohorts

Effect of bariatric surgery.

We found that bariatric surgery has remodelled plasma biochemistry results. In particular, the decrease of TG in obese individuals after bariatric surgery was in agreement with what was already shown by Szczuko and colleagues[90]. TG levels were different between lipodystrophy and post surgery groups, but not between lipodystrophy and obese groups. Although our results did not establish a direct effect of bariatric surgery on lipodystrophy patients, several previous studies have demonstrated the beneficial effects of bariatric surgery in lipodystrophy patients with BMI < 30[91–95]. Furthermore, our results showed that bariatric surgery had an profound effect on gene expression and epigenetic profiles of macrophages, monocytes, neutrophils, and platelets larger than those observed when comparing obese (or lipodystrophy) with lean individuals (**Tables S8** to **S17**). Although there

were genes whose expression levels, after surgery, were indistinguishable from the level observed in lean individuals. There were also many more genes whose expression levels changed after surgery to assume values not observed before in any of the conditions tested. These findings indicated that the reduction in inflammatory signatures observed after bariatric surgery in macrophages, monocytes, neutrophils, and platelets were due to novel gene expression landscapes. Of interest, in these cells we did not observe changes in DNA methylation levels of the same magnitude as those observed at gene expression level. The short life span of the cell types analysed and the overall small number of changes in DNA methylation observed in the different comparison, suggested that the changes from a pro-inflammatory to a healthy bone marrow environment had little effect on the hematopoietic stem cell epigenome. Plasma proteomic allowed us to obtain a whole body snapshot of the changes that occur after surgery. The majority of the changes were in proteins whose level decreased after surgery (532 out of 605; **Table S28**). These showed that the effect of the surgery inflammatory response (including NLRP3), insulin signalling, WNT signalling, VEGF signalling were reduced because of the reduction in fat mass but also vascular integrity was restored, as confirmed by the tissue specificity analysis that identified amongst others artery and blood as the sources of production of the reduced plasma proteins (**Table S29**).

Interestingly, we also observed that genes previously been implicated in trained immunity[75], a phenomenon associated with innate cells response to stimuli and also atherosclerosis[96], were found in the comparison between obese and lean individuals. However, post surgery we observed that many of the changes involved genes that belonged to the non-challenged set[75], suggesting that bariatric surgery had a positive impact on innate immune cells or that trained immunity acted downstream the hematopoietic stem cell pool and its effects were eventually diluted and lost. Lastly, the transcriptional events were accompanied by a decrease in adhesion observed in platelets and neutrophils. This could be due to a diminution of proinflammatory signals in the cell environment and would result in a diminished propensity to form a thrombus.

Conclusion

Our study provided a comprehensive overview of the transcriptional and epigenetics features associated with CMS in 3 immune cell types (monocytes, neutrophils and macrophages) and platelets. The integration of multiple -omics data layers allowed us to extract relevant features discriminating between patients and donors, and to construct a predictive model that we used to rank individuals by their likelihood of being a patient. Analysis of -omics data in obese individuals before and six months after bariatric surgery revealed deep rewiring of both transcriptional and epigenetic networks. Only a small proportion of the rewired features were already associated with obesity, suggesting that bariatric surgery does not revert the organism to the lean status but it creates a new status with low inflammation and low thrombotic propensity. Greater understanding of this impact will be required in order to optimize patient support after surgery.

Acknowledgments and funding.

L.S. is supported as PhD student by British Heart Foundation Centre of Excellence ; M.C.S is supported by a MRC Clinical Research Training Fellowships (MR/R002363/1) ; D.B.S is supported by the Wellcome Trust (WT 107064), the MRC Metabolic Disease Unit (MRC_MC_UU_12012.1), and The National Institute for Health Research (NIHR) Cambridge Biomedical Research Centre and NIHR Rare Disease Translational Research Collaboration ; K.D. is supported as a HSST trainee by NHS Health Education England ; P.D.W.K is supported by Medical Research Council (MC_UU_00002/13) ; M.F. is supported by the British Heart Foundation (FS/18/53/33863). D.S. work has been supported in part by an Isaac Newton fellowship to M.F.

Conflict of interest.

Authors have no CoI to declare.

Acknowledgements

We thank the Biomedical Sequencing Facility at CeMM for assistance with next-generation sequencing.

Legends.

Table 1 - Summary table of anthropometric and biochemical parameters. Average values are reported and interquartile ranges are given in brackets. BW: Body weight; FFA: free fatty acid; GLC: Glucose; TG: serum lipid profiles (triglycerides); TC: total cholesterol; HDL-C: high density lipoprotein; LDL-C: low-density lipoprotein; ALT: alanine amino-transferases; AST: aspartate amino-transferases; hsCRP: high-sensitivity C-reactive Protein ; LAR: leptin-adiponectin ratio; HOMA-IR: Homeostatic Model Assessment for Insulin Resistance; AT-IR: adipose tissue insulin resistance; BMI: body mass index

Figure legends.

Figure 1 - High cardiovascular risk groups characterisation using anthropometric, biochemical and metabolic profiles. A. Principal component analysis (PCA) of 3 groups: obese (BMI>40), green; lipodystrophy, blue; and blood donors (WP10), light red. PCA was performed using the parameters below. **B**. Representation of PCA loadings on: age, weight (BW), body mass index (BMI), leptin-adiponectin ratio (LAR), glucose (GLC), triglycerides (TG), total cholesterol (TC), high density lipoprotein (HDL-C), low-density lipoprotein (LDL-C), alanine amino-transferase (ALT), aspartate amino-transferase (AST), Homeostatic Model Assessment for Insulin Resistance (HOMA-IR) and adipose tissue insulin resistance (AT-IR) indexes and high-sensitivity C-reactive Protein (hsCRP). Colour and arrow length scale represent contribution to variance on the first two principal components. **C**. Metabolite module-trait associations using weighted Gene Correlation Network Analysis (WGCNA) consensus analysis and 988 metabolites (Metabolon). Each row

corresponds to a module eigen-metabolites (ME), and each column to a parameter. Number of metabolites in each module is indicated in brackets. Cell colour represents Pearson's correlation as shown by legend. Significance is annotated as follows: * P≤ 0.05, ** P ≤ 0.01, *** P ≤ 0.001, **** P ≤ 0.0001 (Fisher's test). Red stars indicate module-trait associations found to be significant in both groups. Left panel shows eigen-metabolites values in patients whereas the right panel shows the values in donors. **D.** Heatmap of patients' eigen-metabolites adjacencies in the consensus eigen-metabolites network. Each row and column corresponds to one eigen-metabolite (labeled by consensus module color). The heatmap is color-coded by adjacency, yellow indicating high adjacency (positive correlation) and blue low adjacency (negative correlation) as shown by the color legend. **E.** Beeswarm plot using average eigen-metabolites per cluster. Shapes indicate cohorts and color legend represents the first principal component value extracted from the plot in 1A.

Figure 2 - Transcriptional and epigenetic comparison of obese and lipodystrophy patients versus lean individuals in 3 immune cell types and platelets. **A.** Schematic overview of the comparisons made in the 4 different cell types (Monocytes: blue ; Neutrophils: green ; Macrophages: purple ; Platelets: yellow). **B**. Barplot showing number of features significantly different when comparing lean individuals and obese individuals in H3K27ac distribution (ChIP-seq), gene expression (RNA-seq) and DNA methylation (RRBS). Each bar is color coded to represent the different cell types as in A. **C**. Annotation of down-regulated genes in obese individuals compared to lean individuals, colour coded by cell types as above. Annotation source is indicated in brackets. The numbers near each dot indicate, from left to right: number of submitted genes, number of genes overlapping with the category and number of genes in the category. **D**. Annotation of up-regulated genes in obese individuals compared to lean individuals colored per cell types. Annotation source is indicated in brackets. The numbers near each dot indicate, from left to right: number of submitted genes, number of genes overlapping with the category and number of genes in the category.

Figure 3 - Multi omics integration strategy and signature identification . A. Presentation of the different layers used for multi omic integration, the strategy

leading to signature identification and schematic view of WP10 stratification. **B**. Heatmap showing the mean of the Z-score distribution for each group, for all features selected in each layer. **C.** Left: Heatmap representing Z-score values for 8 top ranked lipids (columns) and all individuals ranked by the probability of being a patient according to the multi omic model (rows). Gaps indicate individuals for whom lipid data were not collected. Right: Heatmap representing Z-score values for clinical parameters (columns) for all individuals (rows). **D.** Heatmap showing age and sex adjusted association values between 8 previously selected lipids matching quantified lipids in Fenland cohort and Fenland outcome parameters. Black frames indicate significant associations after correcting for multiple testing ($p<0.000125$). **E.** As left, but for 10 lipids not selected in our analyses **F.** As D, but using lipid and outcome data from the present study.

Figure 4 - Effect of bariatric surgery at molecular level in 3 immune cell types and platelets. **A**. Biochemical values distribution accross the four studied groups: obese (BMI>40) (dark green); lipodystrophy (blue); blood donors (WP10) (light red); and post bariatric surgery patients (light green). Asterisks indicate result of significance from multiple logistic regression models and conditional multiple logistic regression for obese and post surgery comparison. Significance is annotated as follows: * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$. **B**. Barplot showing number of features significantly different when comparing the same individuals before and after bariatric surgery colored by cell types. **C** .Volcano plot showing differentially abundant plasma proteins when comparing obese individuals before and after bariatric surgery. Whole blood specific genes associated with differentially abundant proteins have been annotated. **D**. RNAseq expression in the 4 different cell types of highlighted proteins in C). Asterisks indicate if the gene was differentially expressed genes in at least one cell type. **E**. Adhesion percentage of neutrophils measured in the presence of different pro-inflammatory molecules in obese (dark green) individuals and six months after bariatric surgery (light green). Asterisks indicate the result of significance from paired t-test. Significance is annotated as follows: * $P \leq 0.05$, ** $P \leq 0.01$.

Supplemental figure 1 - WGCNA analysis with WP10 donors metabolite values and cluster functional annotation. A. Heatmap of WP10 donors eigen-metabolites adjacencies in the consensus eigen-metabolites network. Each row and column corresponds to one eigen-metabolite (labeled by consensus module color). The heatmap is color-coded by adjacency, yellow indicating high adjacency (positive correlation) and blue low adjacency (negative correlation) as shown by the color legend. **B.** Beeswarm plot using average eigen-metabolites per cluster. Shapes indicate cohorts and color legend represents the first principal component value extracted from the plot in Figure 1A. **C**. Functional annotation of cluster C1, C2 and C4. Heatmap shows normalised abundance of metabolites belonging to each functional category.

Supplemental figure 2 - Summary plots of different feature numbers in all comparisons. Barplots showing the number of features significantly different for each comparison in H3K27ac distribution (ChIP-seq), gene expression (RNA-seq) and DNA methylation (RRBS). Each bar is color coded to represent the different cell types.

Supplemental figure 3 - Model built with obese individuals discriminated between lipodystrophy patients and donors. A. Heatmap showing the mean of the Z-score distribution for each group, for all features selected in each layer. **B.** Left: Heatmap representing Z-score values for 8 top ranked lipids (columns) and all individuals ranked according to the multi omic model probability of being an obese individual (rows). Gaps indicate individuals for whom lipid data were not collected. Right: Heatmap representing Z-score values for clinical parameters (columns) for all individuals (rows).

Supplemental figure 4 - Model built with lipodystrophy patients discriminated between obese individuals and donors. A. Heatmap showing the mean of the Z-score distribution for each group, for all features selected in each layer. **B.** Left: Heatmap representing Z-score values for 8 top ranked lipids (columns) and all individuals ranked according to the multi omic model probability of being a

lipodystrophy patient (rows). Gaps indicate individuals for whom lipid data were not collected. Right: Heatmap representing Z-score values for clinical parameters (columns) for all individuals (rows).

Supplemental figure 5 - Association of selected lipids with NASH cohort outcomes. Left: Heatmap showing age and sex corrected association values between 8 selected lipids matching quantified lipids in NASH cohort and NASH outcome parameters. Right: As left, but for 10 lipids not selected in our analyses.

Material and methods

Raw sequencing datasets are available under EGA study EGAS00001003780.
Code produced for this study is available at https://gitlab.com/dseyres/extremephenotype.

Patients recruitment and ethics

Obese individuals referred for obese surgery by the obesity clinic and lipodystrophy patient cared for by the National Severe Insulin Resistance Service respectively, both based at Addenbrooke's hospital, Cambridge University Hospitals were recruited to this study together with healthy individuals. Informed consent was obtained under the "Inherited Platelet Disorders" ethics (REC approval 10/H0304/66 for patients and 10/H0304/65 for healthy controls, NRES Committee East of England-Cambridge East).
BluePrint WP10 volunteers were recruited amongst NHS Blood and Transplant donors after informed consent under the "A Blueprint of Blood Cells" ethical approval (REC approval 12/EE/0040 NRES Committee East of England-Hertfordshire).

Cell types isolation

Whole blood (50ml) in citrate tubes was obtained after informed consent. Platelet rich plasma (PRP) was separated from the cellular fraction by centrifugation (20', 150g and very gentle break). Platelets then isolated from PRP after 2 more spins as above and leukodepleted using anti CD45 Dynabeads (Thermofisher) following the

manufacturer's instructions. Purified platelets were stored in TRIzol (Invitrogen) until RNA extraction. The remaining cells were resuspended in buffer 1 and separated on a Percoll gradient. Neutrophils were harvested from the pellet after red cell lysis (4.15 g NH4Cl, 0.5 g KHCO3 and 18.5 mg EDTA (triplex III, 0.01%) in 500 ml of water) and aliquots prepared for RNA extraction (TRIzol), DNA extraction for RRBS (snap frozen pellet) and ChIP-seq (formaldehyde fixation, see below). Monocytes were isolated from peripheral blood mononuclear cells (PBMC) layer by CD14 positive selection (Miltenyi) and aliquots prepared for RNA extraction (TRIzol), DNA extraction for RRBS (snap frozen pellet) and ChIP-seq (formaldehyde fixation, see below). Macrophages were cultured by plating 14*106 PBMC resuspended in 2 ml macrophage media (Macrophage-SFM [with L-Glutamine without Antibiotics], Fisher Scientific UK LTD). After 1h 30' non adherent cells were removed and 1 ml fresh macrophage media added together with 400 μl of autologous serum. Culture media was replaced after 3 or 4 days. On day 7 cells were harvested for RNA extraction (TRIzol), DNA extraction for RRBS (snap frozen pellet) and ChIP-seq (formaldehyde fixation). Cell purity was determined by flow cytometry as follows: neutrophils CD66b (BIRMA17c, FITC, 9453 https://ibgrl.blood.co.uk/), CD16 (VEP13, PE, 130-091-245 Miltenyi) and CD45 (HI30, PE-CY5.5, MHCD4518 Invitrogen); monocytes CD14 (MφP9, FITC, 345784 BD), CD16 (B73.1 / leu11c, PE, 332779 BD), CD64 (10.1, PerCP-Cy5.5, 561194 BD), CD45 (HI30, PE-CY7, MHCD4512 Invitrogen); macrophages panel (only done for WP10 donors) 1: CCR7/CD197 (150503, FITC 561271 BD), CD25-PE MACS 120-001-311 (10ul/test), CD14 (TuK4, PE-Cy5.5, MHCD1418 Invitrogen), CD40 (5C3, PE-Cy7, 561215 BD). Panel 2: CD206 (19.2, PE, 555954 BD), CD36 (SMΦ, FITC, 9605-02 Southern Biotech), CD45 (HI30, PE-Cy5.5, MHCD4518 Invitrogen). Samples whose purity was below 90% were discarded. WP10 samples isolation has been extensively described in Chen et al.{27863251}.

RNA sequencing

RNA extraction

RNA extraction from samples stored in TRIzol was performed following the manufacturer's instructions. Briefly, tubes were retrieved in small batches and

thawed on ice. Prior to extraction samples were vortexed for 30” to ensure complete lysis and let for 5’ at room temperature. Samples were then transferred to heavy phase lock tubes (5prime), chloroform was added and the tubes spun to separate RNA in the aqueous phase from the organic phase. RNA was precipitated from the former with isopropanol and glycogen. The RNA pellet was washed with 75% ethanol and resuspended in RNase free water. Purified RNA was stored in single use aliquots. Each sample was quality controlled by Bioanalayser (Agilent) and quantified via Qubit (Thermofisher).

Library preparation and sequencing

For cell types isolated from obese and lipodystrophy patients and day controls we used 100 ng total RNA for neutrophils, monocytes and macrophages; 200 ng total RNA for platelets to prepare libraries for sequencing using the Kapa stranded RNA-seq kit with riboerase (Roche) according to the manufacturer's instructions and sequenced 150bp paired end on Illumina HiSeq 2500 or Illumina HiSeq 4000. WP10 RNA-seq data (extensively described in Chen et al.[75,97]) were retrieved from European Genome-phenome Archive (EGA) - EMBL-EBI after application to the Data Access Committee.

Quantification

FastQ files were first checked for sequencing quality using FastQC (v.0.11.2) [https://www.bioinformatics.babraham.ac.uk/projects/fastqc/] and quality trimming were applied on reads with TrimGalore! (v.0.3.7) [https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/].

Transcript-level abundance was estimated using Kallisto (v0.42)[98] with 100 bootstrap iterations in single-end mode for patients samples in order to minimize technical batch effect with WP10 donors cohort. Transcript abundances were then summarized to gene-level with Tximport R package (v1.9)[99] by using tximport function and Ensembl reference transcriptome (Ensembl Genes 96)[99,100]. This step provides input counts matrix for DESeq2 (v.1.21.21)[101]. DESeq2 was used to normalize counts by library size and transformed by variance stabilisation (VST). We corrected for sequencing batch effect by using Combat (from sva R package

(v.3.29.1))[101,102] and individual status as covariate. Non-autosomal and genes with no or low variance (<0.05) across individuals were removed. The final gene sets (including coding and non-coding genes) were formed of 11,370 genes for monocytes and of 24,224 for neutrophils.

Differential analysis

For differential analysis, transcript-level abundance were estimated by Kallisto with 100 bootstrap iterations in paired-end mode for each group (obese, post surgery, lipodystrophy patients and lean individuals) using Ensembl reference transcriptome (Ensembl Genes 96). Transcript abundances were then summarized to gene-level with Tximport R package (v1.9) by using tximport function and DESeq2 object was created using DESeqDataSetFromTximport function from DESeq2 R package (v.1.21.21). Differential analysis was performed using Deseq function from DESeq2. Log fold changes were corrected with lfcShrink function from DESeq2. Genes with FDR < 5% were marked as differentially expressed. For obese versus post surgery comparison, we considered only paired samples ([S01RS6;S022QS][S01Y9G;S022UK][S01WCI;S0232Z][S01TEQ;S0234V][S01WXD;S023EB][S01WFC;S023F9][S01Y7K;S023H5][S022TM;S023PQ][S01XJ0;S023RM][S01SYR;S0240Z][S022GB;S0245P]) and therefore performed a paired analysis by adding relationship information as covariate in the design formula.

Functional annotation was performed with genes differentially expressed for each cell-types and comparisons, taking into account fold change direction. Lists of genes were submitted to EnrichR using R package EnrichR (v.1.0)[103,104] and the following databases: BioCarta_2016, DSigDB, GO_Biological_Process_2018, GO_Cellular_Component_2018, GO_Molecular_Function_2018, HMDB_Metabolites, KEGG_2019_Human, Reactome_2016 and WikiPathways_2015. To facilitate gene lists submission, we developed an R shiny interface to EnrichR (https://blueprint.haem.cam.ac.uk/EnrichR_Interface/).

Chromatin Immunoprecipitation sequencing

Sample preparation

Cells were fixed immediately after purification with 1% w/v formaldehyde for 10 min and quenched using 125 mM glycine before washing with PBS. Samples were sonicated using a Bioruptor (Diagenode), final SDS concentration of 0.1% w/v for 9 cycles of 30 s 'on' and 30 s 'off', and immunoprecipitated using an IP-Star Compact Automated System (Diagenode) using the histone H3K27ac antibody C15410196 (lot 1723-0041D) Diagenode. Immunoprecipitated and input DNA were reverse cross-linked (65 C for 4 h), treated with RNase and Proteinase K (65 C for 30 min).

Library preparation and sequencing

DNA was recovered with Concentrator 5 columns (Zymo) and prepared for sequencing using MicroPlex Library Preparation Kit v2 (C05010012, Diagenode). Libraries analysed using High Sensitivity Bioanalyzer chips (5,067–4,626, Agilent), quantified using qPCR Library Quantification Kit (KK4824, Kapa Biosystems), pooled and sequenced with a 50bp single end protocol on Illumina HiSeq 2500 or Illumina HiSeq 4000.

Peak calling and quantification

FastQ files were first checked for sequencing quality using FastQC (v.0.11.2) and quality trimming were applied on reads with TrimGalore! (v.0.3.7). Trimmed FASTQ files were aligned to the human genome (Ensembl GRCh38.80) with BWA (v.0.7.12)[105] aln and samse functions with default parameters. Low mapping quality reads (-q 15), multi-mapped and duplicate reads were marked and removed with respectively samtools (v.1.3.1)[106] and picard (http://broadinstitute.github.io/picard v.2.0.1).

A combination of quality metrics was used to assess samples quality: number of uniquely mapped reads, number of called peaks, NSC (Normalized strand cross-correlation) and RSC (relative strand cross-correlation) computed with Phantompeakqualtools (v.1.2)[107,108], area under the curve (AUC), X-intercept and Elbow Point computed with plotFingerPrint function from deepTools suite (v.3.0.2)[109] with --skipZeros --numberOfSamples 50000 options. Peaks were called with MACS2

(v.2.1.1) with --nomodel --shift -100 --extsize 200, a qvalue threshold of 1e-3 options and celltype matching input file scaled to sample read number. We used MACS2 randsample function to downscale inputs. We then computed a score by summing values obtained for each range of these metrics. We applied a threshold of -3 (total) to select the best quality data.

	-2	-1	0	1	2
Uniq reads (% raw reads)	<20	>=20 and <40	>=40 and <60	>=60 and <80	>=80
Encode - NSC	<0.9	>=0.9 and <1	>=1 and <1.1	>=1.1 and <1.2	>=1.2
Encode - RSC	<0.8	>=0.8 and <0.9	>=0.9 and <1	>=1 and <1.1	>=1.1
Deeptools - AUC	>=0.4	>=0.3 and <0.4	>=0.2 and <0.3	>=0.1 and <0.2	<0.1
Deeptools - X-intercept	>=0.3	>=0.2 and <0.3	>=0.15 and <0.2	>=0.1 and <0.15	<0.1
Deeptools - Elbow point	<0.65	>0.65 and <0.75	>0.75 and <0.85	>0.85 and <0.95	>0.95
Peak number	<(e-10 000)	>=(e-10000) and <(e-5000)	>=(e-5000) and <(e-2000)	>=(e-2000) and <e	>=e and <(e+25 000)

To build ChIP-seq layer for integrative analysis, we defined a master set of peaks and quantify H3K27ac ChIP-seq signal under these peaks. Peaks shared by at least 5 individuals were merged using R package DiffBind (v2.9)[110]. We obtained 67,763 and 49,188 peaks for monocytes and neutrophils, respectively. Minimum merged peak size was 244bp and 235bp, median peak size 1,392bp and 1,648bp and maximum peak size 75,534bp and 60,528bp for monocytes and neutrophils, respectively. We did not filter out very large merged peaks as they represent less than 3% of total peaks and indicate large acetylated regions. Read counts under merged peaks were TMM normalized using effective library size and logit transformed into count per million (CPM). Sequencing center batch effect was corrected with Combat (from sva R package (v.3.29.1)) using individual status

(Patient/Donor) as covariate. Non-autosomal and no or low variance (<0.1) peaks across individuals were removed. The final master set of peaks counted 25,600 regions in monocytes and 26,300 regions in neutrophils.

Differential analysis

For differential analysis, we used DiffBind with built-in DEseq2 method for statistical analysis. We merged peaks present in at least 50% of individuals and asked that all individuals have a FRiP value (Fraction of Reads in Peaks) over 5%. We then applied a FDR threshold of 5% to select H3K27ac peaks differentially acetylated peaks. For obese versus post surgery comparison, we considered only paired samples and therefore performed a paired analysis by using the block factor in DEseq2. Differentially acetylated regions (DAcR) were annotated with HOMER (v.4.10)[111], annotatePeak function and Hg38 RefSeq genome annotation (http://homer.ucsd.edu/homer/data/genomes/hg38.v6.0.zip).

Functional annotation was performed on genes within a window of 10kb around each DAcR, taking into account fold change direction. Similarly to RNA-seq, lists of genes were submitted to EnrichR interrogating the same databases.

Illumina 450K arrays and reduced representation bisulfite sequencing (RRBS)

Arrays and libraries preparation and sequencing

WP10 Infinium Human Methylation 450 arrays (Illumina) were retrieved from the European Genome-phenome Archive (EGA) - EMBL-EBI. DNA extraction and array generation have been described in detail in Chen et al.[97]. Briefly, cells were lysed using guanidine hydrochloride, sodium acetate and protease lysis buffer. DNA was extracted using chloroform and precipitated in ethanol prior to washing and resuspension in ultra-pure water. 500ng of DNA for each monocyte and neutrophil sample was randomly dispensed onto a 96-well plate to reduce batch effects. Samples were bisulfite-converted using an EZ-96 DNA Methylation MagPrep Kit (Zymo Research) following the manufacturer's instructions with optimized incubation conditions (i.e., 16 cycles of 95C for 30 s, 50C for 60 min; followed by 4C until further processing). Purified bisulfite-treated DNA was eluted in 15 mL of M-Elution Buffer (Zymo Research). DNA methylation levels were measured using Infinium Human Methylation 450 arrays (Illumina) according to the manufacturer's protocol.

For RRBS, 100 ng of genomic DNA were digested for 6h at 65°C with 20 U TaqI (New England Biolabs) and 6h hours at 37°C with 20 U of MspI (New England Biolabs) in 30 µl of 1x NEBuffer 2. To retain even the smallest fragments and to minimize the loss of material, end preparation and adaptor ligation were performed in a single-tube setup. End fill-in and A-tailing were performed by addition of Klenow Fragment 3' --> 5' exo- (Enzymatics) and dNTP mix (10 mM dATP, 1 mM dCTP, 1 mM dGTP New England Biolabs). After ligation to methylated Illumina TruSeq LT v2 adaptors using T4 DNA Ligase rapid (Enzymatics), the libraries were size selected by performing a 0.75x clean-up with AMPure XP beads (Beckman Coulter). The libraries were pooled based on qPCR data and subjected to bisulfite conversion using the EZ DNA Methylation Direct Kit (Zymo Research) with changes to the manufacturer's protocol: conversion reagent was used at 0.9x concentration, incubation performed for 20 cycles of 1 min at 95°C, 10 min at 60°C and the desulphonation time was extended to 30 min. These changes increase the number of CpG dinucleotides covered, by reducing double-strand break formation in larger library fragments. Bisulfite-converted libraries were enriched KAPA HiFi HS Uracil+ RM (Roche). The minimum number of enrichment cycles was estimated based on a qPCR experiment. After a 1x AMPure XP clean-up, library concentrations were quantified with the Qubit Fluorometric Quantitation system (Life Technologies) and the size distribution was assessed using the Bioanalyzer High Sensitivity DNA Kit (Agilent).

Processing and quantification

All Infinium Human Methylation 450 array data pre-processing steps were carried out using established analytical methods incorporated in the R package RnBeads (v.1.13.4)[112]. First, we performed background correction and dye-bias normalization using NOOB[113], followed by normalization between Infinium probe types with SWAN[114]. Next, we filtered out probes based on the following criteria: median detection p value 0.01 in one or more samples; bead count of less than three in at least 5% of samples; ambiguous genomic locations[115]; cross-reactive and SNP-overlapping probes[116].

The RRBS samples were sequenced on Illumina HiSeq3000 platform in 50bp single-end mode. Base calling was performed by Illumina Real Time Analysis

(v2.7.7) software and the base calls were converted to short reads using Illumina2bam (1.17.3 https://github.com/wtsi-npg/illumina2bam) tool before de-multiplexing (BamIndexDecoder) into individual, sample-specific BAM files. Trimmomatic (v0.32)[117] was used for trimming the adapter sequences. Trimmed short read sequences were aligned onto the GRCh38/hg38 human reference genome with BSMAP(v2.90)[118] aligner in RRBS mode which was optimized for aligning the RRBS data while being aware of the restriction sites and with the following options: -D C-CGG -D T-CGA -w 100 -v 0.08 -r 1 -p 4 -n 0 -s 12 -S 0 -f 5 -q 0 -u -V 2. R package RnBeads was used to filter out low confidence sites: sites overlapping any SNP, having a coverage lower than 5 and high coverage or missing in more than 5% or individuals were filtered out.

Integration analysis required to attenuate technology effect between 450K arrays and RRBS. To this goal, we generated RRBS data for 14 BluePrint donors for which we already have 450K array data in monocytes, and 9 in neutrophils. We first removed non reproducible sites between technologies as follows: for monocytes and neutrophils, 1) liftover 450K sites to Hg38 using UCSC liftover tool[119], 2) keep overlapping sites between array and RRBS, 3) filter out sites with high variation in methylation percentage observed in more than 70% of individuals. We excluded 14,301 and 14,972 sites for monocytes and neutrophils respectively. We have also excluded sites on sex chromosomes and imputed missing values using KNN networks (impute.knn function from impute R package (v.1.55.0)) [Hastie T, Tibshirani R, Narasimhan B, Chu G (2019). impute: impute: Imputation for microarray data.] with 10 nearest neighbors.

Finally, we adjusted for batch effects using an empirical Bayesian framework, as implemented in the ComBat function of the R package SVA (v.3.29.1) and individual status as covariate, transformed beta values to M values using beta2m function in R package lumi (v.2.33.0)[120,121], normalize by quantile using normalize.quantiles function from R package preprocessCore (v.1.43.0) [Bolstad B (2019). preprocessCore: A collection of pre-processing functions.] and remove zero or low variance sites (< 3). The final data matrix used for multi-omic integration, comprised DNA methylation M-values across 26,214 CpG sites and 193 samples in monocytes and 21,442 CpG sites and 187 samples in neutrophils.

Differential analysis

For differential analysis, we used methylKit R package (v.1.8.1)[122] and we compared only RRBS data. We first extracted methylation ratios from BSMAP mapping results with methratio.py python script provided with BSMAP. We then removed all sex chromosomes sites and filtered out non-retained sites from RnBeads RRBS processing. Finally, we used the methRead function from methylKit R package in CpGs context at base resolution to read in the input files and calculateDiffMeth function correcting for overdispersion (overdispersion="MN") and applying Chisq-test. Q Values are then computed using SLIM method[122,123]. We applied two thresholds: difference of methylation > 25 and qvalue < 0.05 and retrieved differentially methylated sites (DMS) with getMethylDiff function specifying type=”hypo” or type=”hyper” option to get down and up methylated CpGs respectively.

For obese (pre) versus post surgery comparison, we considered only paired samples and therefore performed a paired analysis. DMS were annotated with HOMER (v.4.10), annotatePeak function and Hg38 RefSeq genome annotation (http://homer.ucsd.edu/homer/data/genomes/hg38.v6.0.zip).

Functional annotation was performed on genes within a window of 10kb around each DMS, taking into account fold change direction. Similarly to RNA-seq and ChIP-seq, lists of genes were submitted to EnrichR interrogating the same databases.

Plasma biochemistry assays

Plasma biochemistry assays were performed in the Core Biochemical Assay Laboratory, Cambridge University Hospitals (https://www.cuh.nhs.uk/core-biochemical-assay-laboratory) as described in supplementary material and methods. Homeostatic Model Assessment for Insulin Resistance (HOMA-IR) index as follows: (glucose (mg/dL) x insulin (mIU/L)) / 405, and adipose tissue insulin resistance (AT-IR) index as follows: insulin (µU/mL) x free fatty acids (mmol/L).

Plasma metabolites measurement

Metabolites quantification

Metabolites profiling of obese and lipodhystrophy patients, day controls and blood donors (WP10 participants) was performed by Metabolon Inc. (https://www.metabolon.com/) using their standard protocol (see extended Methods). Briefly, Metabolon analytical platform incorporates two separate ultra-high performance liquid chromatography/tandem mass spectrometry (UHPLC/MS/MS2) injections and one gas chromatography GC/MS injection per sample. The UHPLC injections are optimized for basic species and acidic species. The numbers of compounds of known structural identity (named biochemicals) as well as compounds of unknown structural identity (unnamed biochemicals) detected by this integrated platform were respectively of 793 and 362 for the first batch and 947 and 433 for the second batch (with an overlap of 786 and 359 compounds respectively). All samples were rescaled to set the median to 1, missing values were imputed using KNN networks (impute.knn function from impute R package (v.1.55.0) with the following options: number of nearest neighbors=10, maximum missing values per metabolites < 50% and maximum missing values for individuals < 80%.) Finally, we adjusted for batch effects using the ComBat function of the R package SVA (v.3.29.1) and individual status as covariate.

Plasma lipids measurement

Plasma was frozen in dry ice immediately after collection and stored at -80C until analysis. Samples were prepared essentially as previously described[124]. Briefly, a 15 µL sample, controls and blanks were placed in a predefined random order across 96-well plates (Plate+, Esslab, Hadleigh, UK). To which, 750 µL methyl tert-butyl ether was added, along with 150 µl of internal standard mix, containing the following six internal standards (IS): 1,2-di-o-octadecyl-sn-glycero-3-phosphocholine (0.6 µM), 1,2-di-O-phytanyl-sn-glycero-3-phosphoethanolamine (1.2 µM), C8-ceramide (0.6 µM), N-heptadecanoyl-D-erythro-sphingosylphosphorylcholine (0.6µM), undecanoic acid (0.6µM), and trilaurin (0.6 µM), (Avanti Polar Lipids and SIgma Aldrich). Quality controls were derived from pooling all samples and serially diluting with chloroform. 25 µl of the sample/IS mixture was transferred to a glass coated 384 well plate and 90µl mass spectrometry (MS) mix [7.5mM NH4Ac IPA:MeOH (2:1)] added and then sealed. Lipidomics was performed using chip-based nanospray with an Advion

TriVersa Nanomate (Advion) interfaced to the Thermo Exactive Orbitrap (Thermo Scientific). Briefy, a mass acquisition window from 200 to 2000 m/z and acquisition in positive and negative modes were used with a voltage of 1.2kV in positive mode and −1.5 kV in negative mode and an acquisition time of 72 s. Raw spectral data were processed as previously described[125]. Raw data were then converted to.mzXML (usingMSconvert[126] with peakpick level 1), parsed with R and 50 spectra per sample (scan from 20 to 70) were averaged using XCMS42, with a signal cutoff at 2000. The files were aligned using the XCMS[127,128] grouping function using “mzClust” with a m/z-window of 22 ppm and a minimum coverage of 60%. Compound annotation was automated using both an exact mass search in compound libraries as well as applying the referenced Kendrick mass defect approach. Signal normalisation was performed by summing the intensities of all detected metabolites to a fixed value to produce a correction factor for the efficiency of ionisation. Exact masses were fitted to the lipid species library and subsequently annotated to the peak as described before[70].

Plasma proteomics

Sample preparation

Plasma was precleared by centrifugation at 3,000 g for 10 minutes and bound to 100 µL of calcium silicate matrix (CSM, 4 mg/mL) by rotation for 1 hour. The sample was centrifuged at 14,000 g for 1 minute and the supernatant was removed for further analysis. The pellet was washed in ammonium bicarbonate (50 mMoL, 1 mL) 3 times using the same centrifugation settings. The sample was then reduced for 30 minutes at 65°C using 200 µL of DL-dithiothreitol (DTT) premix (ADC 2%: ammonium bicarbonate 50 mMoL: DTT 1 MoL in the ratio of 50:49:1) and alkylated for 30 minutes in the dark with iodoacetamide (IAA) at 20 mMoL. Ammonium bicarbonate was added to dilute the ADC to 0.5%. Trypsin was added in the ratio of 1:25 trypsin to plasma and incubated overnight at 37°C. The ADC was precipitated with 1% formic acid (FA) and centrifuged at 14,000 g for 10 minutes. The peptides were isolated using solid phase EMPORE C18 discs which had been washed with 1 stem of methanol and 3 stem of 0.1% FA. The sample was left to bind to the column for 30 minutes before washing with 0.1% FA and eluting with 60% acetonitrile (ACN) with

0.1% FA and then 80% ACN with 0.1% FA. The ACN was removed by speed vacuum for 1 hour 15 minutes and freeze dried overnight. Peptide suspended in 30 µL of 0.1% FA and a peptide assay was performed to calculate the amount of peptides. 10 µL of peptides were removed from each sample and 0.1% FA added to equalise the volume and spiked with an internal standard protein (yeast alcohol dehydrogenase, ADH), with a known amount of 50 fmol injected for each run.

Waters NanoAcquity UPLC and Synapt G2S

Sample separation was performed using an Acquity UPLC Symmetry C18 trapping column (180 µm x 20mm, 5 µm) to remove salt and other impurities and a HSS T3 analytical column (75µm x 150mm, 1.8µm). Solvent A was compromised on 0.1% FA in HPLC grade water and solvent B contained 0.1% FA in ACN.

Time (minute)	Flow rate (µL/minute)	Solvent A (Water + 0.1% FA)	Solvent B (ACN + 0.1% FA)
3	0.3	97	3
20	0.3	86	14
30	0.3	80	20
40	0.3	75	25
51-52.2	0.3	69	31
53-53.1	0.3	65	35
54	0.3	63	37
55	0.3	58	42
63	0.3	31	69
65	0.3	97	3
80	0.3	50	50
80.5	0.3	10	90
82.2-87.5	0.3	97	3
99.5	0.3	50	50
101.5	0.3	10	90
103.5-110	0.3	97	3

Table above shows the gradient in 110 minutes of solvent A and B used in LC ESI-MS/MS analysis. The flow rate of solvents was 0.3 µL/minute. Coupled directly to the Nano Acquity UPLC was a Water Synapt G2S mass spectrometer (Waters Corporation, Manchester, UK). The Synapt G2S includes a nano electrospray ionisation (ESI), StepWave ion guide, Quadrupole, TriWave and TOF (Supplementary Figure 2).

Proteomic data processing and analysis

Progenesis QI for Proteomics (Nonlinear Dynamics, Waters Corporation, UK) was employed to identify and quantify proteins. The human database from UniProtKB was downloaded and used in FASTA format. The proteomic raw data was searched using strict trypsin cleavage rules with a maximum of two missed cleavages. Cysteine (Carbamidomethyl C) was set as a fixed modification. Deamidation N, Oxidation M and Phosphoryl STY were selected as variable modifications. Minimum of 2 fragments per peptide, minimum of 5 fragments per protein and minimum of 2 peptides per protein were set for parameters of identification. The maximum protein mass was set to 1000 kDa. The false rate discovery (FDR) for protein identification was set at a maximum rate of 1%. Then, proteomic data generated from using the Progenesis QI was exported to Microsoft Excel for further data analysis.

For differential analysis, we used LIMMA (v.3.37.4)[129]. Because we compared obese and post surgery patients, we performed a paired analysis. We then applied a threshold of 0.1% on ordinary qvalue.

To define whole blood specific genes, we exported GTEx project[130] expression table (in TPMs), converted it into SummarizedExperiment container using SummarizedExperiment R package ((v.1.11.6); Morgan M, Obenchain V, Hester J, Pagès H SummarizedExperiment: SummarizedExperiment container. (2019)) and used teGeneRetrieval function from the TissueEnrich R package (v.1.2.1)[131]. This package relies on Human Protein Atlas[132] to grouped genes as follows: Tissue Enriched (Genes with an expression level greater than 1 TPM that also have at least five-fold higher expression levels in a particular tissue compared to all other tissues), Group Enriched (Genes with an expression level greater than 1 TPM that also have at least five-fold higher expression levels in a group of 2-7 tissues compared to all other tissues, and that are not considered Tissue Enriched) and tissue Enhanced

(Genes with an expression level greater than 1 TPM that also have at least five-fold higher expression levels in a particular tissue compared to the average levels in all other tissues, and that are not considered Tissue Enriched or Group Enriched). With default parameters, we identified 693 whole blood specific genes. Finally we interesected genes coding for differentially abundant proteins and whole blood specific genes.

Weighted correlation network analysis (WGCNA)

WGCNA[66] is a correlation-based method that describes and visualizes networks of data points, whether they are gene expression estimates, metabolite concentrations or other phenotypic data. We followed the protocols of WGCNA to create metabolic networks. Metabolites are clustered into co-abundant "modules". Low correlations can be suppressed either in a continuous ("soft") manner or the discontinuous ("hard") thresholding used in constructing unweighted networks. To maintain scale-free topology, we estimated an applied power by computing soft-threshold with pickSoftThreshold function from WGCNA R package (v.1.64-1)[133]. To build network, we used blockwiseModules function with the following options: TOMType = "signed", minModuleSize = 20,reassignThreshold = 0, mergeCutHeight = 0.25 and corType="bicor". Each obtained module is notated by a unique color. Additionally, we assigned a name to each consensus module. Each module abundance profile can be summarised by one representative metabolite: the module eigen metabolite. Specifically, the module eigen metabolite was defined as the first right-singular vector of the standardized module expression data[134]. We performed 3 analysis: patients (obese and lipodystrophy patients were combined to get enough individual number for network analysis), donors (all WP10 individuals) and a consensus analysis. We identified 8, 22 and 16 modules with donors, patients and consensus data respectively. We computed eigenmodule and biochemical parameters correlations (leptin-adiponectin ratio (LAR), glucose (GLC), triglycerides (TG), total cholesterol (TC), high density lipoprotein (HDL-C), low-density lipoprotein (LDL-C), alanine amino-transferase (ALT), aspartate amino-transferase (AST), Homeostatic Model Assessment for Insulin Resistance (HOMAIR) and adipose tissue insulin resistance (AT-IR) indexes and high-sensitivity C-reactive Protein (hsCRP) and also

weight (WGT), BMI and age) using cor function from stats R base package (R version 3.5.0) and pearson method (default). P Value of each correlation was computed using corPvalueStudent function from WGCNA R package.

Pathways enrichment analysis were performed with MetaboAnalyst[135] and in particular Pathway analysis module by submitting combined list HMDB identifiers for clusters C1 And C2, hypergeometric test, relative-betweenness centrality topology analysis and KEGG database. In addition, we submitted these lists to the Reactome database.

Multi-omic integration

Training datasets

We identified 16 WP10 donors as controls, according to the following criteria: BMI < 25, glycaemia (GLUC) <5.4 mmol/L, TG <1.7 mmol/L, LDL <2.59 mmol/L, HDL >1 mmol/L for men and >1.3 mmol/L for women, HOMA score< 2.2. For training the multi-omics predictive model (see below), we used a reduced training dataset comprising the subset of individuals having measurements across all omics layers. This reduced set comprised 6 controls, 6 obese individuals and 10 lipodystrophy patients. For the clinical data, we first used multiple imputation by chained equations, as implemented in the mice R package (with default options) to impute missing values before construction of the training dataset.

Variable selection: multivariable regression approach

For each of the omics layers considered independently, we used elastic-net penalised logistic regression as implemented in the glmnet R package to identify putative signatures that discriminated between all patients (i.e. lipodystrophy + obese) versus controls. The elastic-net α parameter was fixed at $\alpha = 0.01$, while the λ parameter was determined using cross-validation. Since different cross-validation splits resulted in different choices for λ, we performed multiple rounds of cross-validation, and used the value of λ that resulted in the maximum number of selections.

Clinical predictive model

We trained a ridge-penalised logistic regression model predictive of the binary response (i.e. patient/control status) using the clinical training dataset.

Multi-omics predictive model

We used the omic variables selected by the multivariable approach described above, together with the clinical covariates, to train a ridge-penalised logistic regression model predictive of the binary response (i.e. patient/control status). We fitted this model using the reduced training dataset. We used this model to make predictions for the 96 individuals for which we had measurements across all omics layers. To allow us to make predictions for those individuals for which we only had measurements on a subset of the omics datasets, we additionally fitted models to each combination of subsets.

Validation of selected lipids

To further investigate the lipidomic signature, we identified selected lipid species that were also measured in two other studies: a subset of 1,507 participants of the Fenland study[69,70] which is a population-based cohort of 12,345 volunteers without diabetes born between 1950 and 1975 and recruited within the Cambridgeshire region between 2005 and 2015, and a biopsy-proven nonalcoholic steatohepatitis (NASH) cohort comprising 42 individuals[70]. We used linear regression analysis to test for association between plasma levels of 8 lipid speciess selected into the lipidomic signature and all relevant CMD parameters quantified in the Fenland cohort, adjusting for age and sex, and using the Bonferroni method to control for multiple testing. We repeated this analysis for a set of 10 lipids that were not selected by either our multivariable or univariate variable selection approaches.

Functional tests

Neutrophils Adhesion Method:

Polymorphonuclear granulocytes were isolated via density gradient (1.078g/mL) from 3.2% sodium citrated whole blood within 2hours of venipuncture. Neutrophil purity

was assessed by haematology analyser (Sysmex, XN-450) to ensure purity levels were satisfactory (≥90%) for subsequent functional assays. Isolated cells were incubated in a water bath at 37C for 30 minutes with fluorescently labelled Calcein-AM (4ug/mL, Molecular probes). Cells were washed twice with 1x PBS and resuspended at 2x106/ml in HEPES complete medium supplemented with calcium (1mM). 1.6x105 fluorescently labelled neutrophils were then added to relevant duplicate wells in a 96-well plate containing the following stimuli; fMLP, 10μM; DTT, 10mM; Pam3Cys, 20μg/ml; LBP+LPS, 50ng/mL and 20ng/mL; PAF, 1μM; PMA, 1μg/mL; TNF, 10ng/mL or HEPES only as a control in a final volume of in 100μl. Cells were incubated for 30 minutes at 37C in a 5% CO2 incubator, after which they were washed twice using 1x PBS before lysing in 100μl PBS with 0.5% triton. A 100% adhesion control was generated by lysing 1.6x105 fluorescently labelled neutrophils in 0.5% triton. Fluorescent intensity was measured using a Tecan Infinite® 200 PRO series plate reader (excitation of 485/20nm and emission of 535/25nm). The mean of duplicate values were calculated and the % adhesion over the hepes control calculated using the following formula: % adhesion = ((RFU stimuli – RFU HEPES)/ RFU 100% control) x 100.

CD63 Expression:

50ul of whole blood was incubated with antibodies:

CD16	PE	VEP13	Miltenyi
CD63	APC	H5C6	Miltenyi
CD11b	APC	ICRF44	BD Pharmingen™
CD62L	FITC	Dreg 56	BD Pharmingen™
CD32	FITC	FLI8.26	BD Pharmingen™
CD14	APC	MφP9	BD Pharmingen™

for 20 minutes, followed by a red cell lysis (BD FACS lyse) and resuspension in 0.2% formyl saline. Samples were analysed using flow cytometry (Beckman Coulter, FC500) within 4 hours. Neutrophils were identified using scatter properties and CD16 positivity. BD CompBeads were used to generate compensation controls. The

median fluorescence intensity (MFI) for each surface marker was calculated using Kaluza Analysis Software (Beckman Coulter).

References

1. Kelli, H. M. & Kassas, I. Cardio Metabolic Syndrome: A Global Epidemic. *J. Diabetes Metab.* **6**, (2016).
2. Go, A. S. *et al.* Heart disease and stroke statistics--2014 update: a report from the American Heart Association. *Circulation* **129**, e28–e292 (2014).
3. Leal, J., Luengo-Fernández, R., Gray, A., Petersen, S. & Rayner, M. Economic burden of cardiovascular diseases in the enlarged European Union. *Eur. Heart J.* **27**, 1610–1619 (2006).
4. Grundy, S. M. *et al.* Diagnosis and management of the metabolic syndrome: an American Heart Association/National Heart, Lung, and Blood Institute Scientific Statement. *Circulation* **112**, 2735–2752 (2005).
5. Azzu, V., Vacca, M., Virtue, S., Allison, M. & Vidal-Puig, A. Adipose tissue-liver cross talk in the control of whole-body metabolism: implications in non-alcoholic fatty liver disease. *Gastroenterology* (2020) doi:10.1053/j.gastro.2019.12.054.
6. 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 Atherosclerosis Society; and International Association for the Study of Obesity. *Circulation* **120**, 1640–1645 (2009).
7. Hotamisligil, G. S. Inflammation and metabolic disorders. *Nature* **444**, 860–867 (2006).
8. Stienstra, R. & Stefan, N. Tipping the inflammatory balance: inflammasome activation distinguishes metabolically unhealthy from healthy obesity. *Diabetologia* **56**, 2343–2346 (2013).

9. McNeill, A. M. *et al.* The metabolic syndrome and 11-year risk of incident cardiovascular disease in the atherosclerosis risk in communities study. *Diabetes Care* **28**, 385–390 (2005).
10. Ford, E. S. Risks for all-cause mortality, cardiovascular disease, and diabetes associated with the metabolic syndrome: a summary of the evidence. *Diabetes Care* **28**, 1769–1778 (2005).
11. Ninomiya, J. K. *et al.* Association of the metabolic syndrome with history of myocardial infarction and stroke in the Third National Health and Nutrition Examination Survey. *Circulation* **109**, 42–46 (2004).
12. Shimada, Y. J. *et al.* Bariatric surgery is associated with lower risk of acute care use for cardiovascular disease in obese adults. *Cardiovasc. Res.* **115**, 800–806 (2019).
13. Pahan, K. Lipid-lowering drugs. *Cell. Mol. Life Sci.* **63**, 1165–1178 (2006).
14. Majithia, A. & Bhatt, D. L. Novel Antiplatelet Therapies for Atherothrombotic Diseases. *Arterioscler. Thromb. Vasc. Biol.* **39**, 546–557 (2019).
15. Sanchez-Rangel, E. & Inzucchi, S. E. Metformin: clinical use in type 2 diabetes. *Diabetologia* **60**, 1586–1593 (2017).
16. Drucker, D. J. & Nauck, M. A. The incretin system: glucagon-like peptide-1 receptor agonists and dipeptidyl peptidase-4 inhibitors in type 2 diabetes. *Lancet* **368**, 1696–1705 (2006).
17. Kosmas, C. E. *et al.* Anti-inflammatory therapy for cardiovascular disease. *Ann Transl Med* **7**, 147 (2019).
18. Onesi, S. O. & Ignatius, U. E. Metabolic syndrome: Performance of five different diagnostic criterias. *Indian J. Endocrinol. Metab.* **18**, 496–501 (2014).
19. Matthews, D. R. *et al.* Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* **28**, 412–419 (1985).

20. Søndergaard, E., Espinosa De Ycaza, A. E., Morgan-Bathke, M. & Jensen, M. D. How to Measure Adipose Tissue Insulin Sensitivity. *J. Clin. Endocrinol. Metab.* **102**, 1193–1199 (2017).
21. Bedogni, G. *et al.* The Fatty Liver Index: a simple and accurate predictor of hepatic steatosis in the general population. *BMC Gastroenterol.* **6**, 33 (2006).
22. Sterling, R. K. *et al.* Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. *Hepatology* **43**, 1317–1325 (2006).
23. Artigao-Rodenas, L. M. *et al.* Framingham risk score for prediction of cardiovascular diseases: a population-based study from southern Europe. *PLoS One* **8**, e73529 (2013).
24. Hippisley-Cox, J. *et al.* Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. *BMJ* **336**, 1475–1482 (2008).
25. Stefan, N., Schick, F. & Häring, H.-U. Causes, Characteristics, and Consequences of Metabolically Unhealthy Normal Weight in Humans. *Cell Metab.* **26**, 292–300 (2017).
26. Kip, K. E. *et al.* Clinical importance of obesity versus the metabolic syndrome in cardiovascular risk in women: a report from the Women's Ischemia Syndrome Evaluation (WISE) study. *Circulation* **109**, 706–713 (2004).
27. St-Pierre, A. C. *et al.* Insulin resistance syndrome, body mass index and the risk of ischemic heart disease. *CMAJ* **172**, 1301–1305 (2005).
28. Katzmarzyk, P. T., Janssen, I., Ross, R., Church, T. S. & Blair, S. N. The importance of waist circumference in the definition of metabolic syndrome: prospective analyses of mortality in men. *Diabetes Care* **29**, 404–409 (2006).
29. Nichols, G. A. *et al.* Cardiometabolic Risk Factors Among 1.3 Million Adults With Overweight or Obesity, but Not Diabetes, in 10 Geographically Diverse Regions of the United States, 2012-2013. *Prev. Chronic Dis.* **14**, E22 (2017).
30. Virtue, S. & Vidal-Puig, A. Adipose tissue expandability, lipotoxicity and the Metabolic Syndrome--an allostatic perspective. *Biochim. Biophys. Acta* **1801**, 338–349 (2010).

31. Choe, S. S., Huh, J. Y., Hwang, I. J., Kim, J. I. & Kim, J. B. Adipose Tissue Remodeling: Its Role in Energy Metabolism and Metabolic Disorders. *Front. Endocrinol.* **7**, 30 (2016).
32. Hotamisligil, G. S., Shargill, N. S. & Spiegelman, B. M. Adipose expression of tumor necrosis factor-alpha: direct role in obesity-linked insulin resistance. *Science* **259**, 87–91 (1993).
33. Vishvanath, L. & Gupta, R. K. Contribution of adipogenesis to healthy adipose tissue expansion in obesity. *J. Clin. Invest.* **129**, 4022–4031 (2019).
34. Visser, M., Bouter, L. M., McQuillan, G. M., Wener, M. H. & Harris, T. B. Elevated C-reactive protein levels in overweight and obese adults. *JAMA* **282**, 2131–2135 (1999).
35. Park, H. S., Park, J. Y. & Yu, R. Relationship of obesity and visceral adiposity with serum concentrations of CRP, TNF-alpha and IL-6. *Diabetes Res. Clin. Pract.* **69**, 29–35 (2005).
36. Lancha, A., Frühbeck, G. & Gómez-Ambrosi, J. Peripheral signalling involved in energy homeostasis control. *Nutr. Res. Rev.* **25**, 223–248 (2012).
37. Frühbeck, G. *et al.* Normalization of adiponectin concentrations by leptin replacement in ob/ob mice is accompanied by reductions in systemic oxidative stress and inflammation. *Sci. Rep.* **7**, 2752 (2017).
38. Gómez-Ambrosi, J. *et al.* Increased cardiometabolic risk factors and inflammation in adipose tissue in obese subjects classified as metabolically healthy. *Diabetes Care* **37**, 2813–2821 (2014).
39. Lawler, H. M. *et al.* Adipose Tissue Hypoxia, Inflammation, and Fibrosis in Obese Insulin-Sensitive and Obese Insulin-Resistant Subjects. *J. Clin. Endocrinol. Metab.* **101**, 1422–1428 (2016).
40. Berg, A. H. & Scherer, P. E. Adipose tissue, inflammation, and cardiovascular disease. *Circ. Res.* **96**, 939–949 (2005).
41. Ramirez, G. A., Manfredi, A. A. & Maugeri, N. Misunderstandings Between Platelets and

Neutrophils Build in Chronic Inflammation. *Front. Immunol.* **10**, 2491 (2019).

42. Puhr-Westerheide, D. *et al.* Neutrophils promote venular thrombosis by shaping the rheological environment for platelet aggregation. *Sci. Rep.* **9**, 15932 (2019).
43. Bobryshev, Y. V., Ivanova, E. A., Chistiakov, D. A., Nikiforov, N. G. & Orekhov, A. N. Macrophages and Their Role in Atherosclerosis: Pathophysiology and Transcriptome Analysis. *Biomed Res. Int.* **2016**, 9582430 (2016).
44. Nording, H. M., Seizer, P. & Langer, H. F. Platelets in inflammation and atherogenesis. *Front. Immunol.* **6**, 98 (2015).
45. Weisberg, S. P. *et al.* Obesity is associated with macrophage accumulation in adipose tissue. *J. Clin. Invest.* **112**, 1796–1808 (2003).
46. Aron-Wisnewsky, J. *et al.* Human adipose tissue macrophages: m1 and m2 cell surface markers in subcutaneous and omental depots and after weight loss. *J. Clin. Endocrinol. Metab.* **94**, 4619–4623 (2009).
47. van Tuijl, J., Joosten, L. A. B., Netea, M. G., Bekkering, S. & Riksen, N. P. Immunometabolism orchestrates training of innate immunity in atherosclerosis. *Cardiovasc. Res.* **115**, 1416–1424 (2019).
48. Gros, A., Ollivier, V. & Ho-Tin-Noé, B. Platelets in inflammation: regulation of leukocyte activities and vascular repair. *Front. Immunol.* **5**, 678 (2014).
49. Koupenova, M., Clancy, L., Corkrey, H. A. & Freedman, J. E. Circulating Platelets as Mediators of Immunity, Inflammation, and Thrombosis. *Circ. Res.* **122**, 337–351 (2018).
50. Caielli, S., Banchereau, J. & Pascual, V. Neutrophils come of age in chronic inflammation. *Curr. Opin. Immunol.* **24**, 671–677 (2012).
51. Wright, H. L., Moots, R. J., Bucknall, R. C. & Edwards, S. W. Neutrophil function in inflammation and inflammatory diseases. *Rheumatology* **49**, 1618–1631 (2010).
52. Herranz, P., de Lucas, R., Pérez-España, L. & Mayor, M. Lipodystrophy syndromes. *Dermatol. Clin.* **26**, 569–78, ix (2008).

53. Polyzos, S. A., Perakakis, N. & Mantzoros, C. S. Fatty liver in lipodystrophy: A review with a focus on therapeutic perspectives of adiponectin and/or leptin replacement. *Metabolism* **96**, 66–82 (2019).

54. Karlsson, J., Taft, C., Rydén, A., Sjöström, L. & Sullivan, M. Ten-year trends in health-related quality of life after surgical and conventional treatment for severe obesity: the SOS intervention study. *Int. J. Obes.* **31**, 1248–1261 (2007).

55. Maggard, M. A. *et al.* Meta-analysis: surgical treatment of obesity. *Ann. Intern. Med.* **142**, 547–559 (2005).

56. Sjöström, L. *et al.* Effects of bariatric surgery on mortality in Swedish obese subjects. *N. Engl. J. Med.* **357**, 741–752 (2007).

57. Modesitt, S. C. *et al.* Women at extreme risk for obesity-related carcinogenesis: Baseline endometrial pathology and impact of bariatric surgery on weight, metabolic profiles and quality of life. *Gynecol. Oncol.* **138**, 238–245 (2015).

58. Gralka, E. *et al.* Metabolomic fingerprint of severe obesity is dynamically affected by bariatric surgery in a procedure-dependent manner. *Am. J. Clin. Nutr.* **102**, 1313–1322 (2015).

59. Ling, C. & Rönn, T. Epigenetics in Human Obesity and Type 2 Diabetes. *Cell Metab.* **29**, 1028–1044 (2019).

60. Miller, G. D., Nicklas, B. J. & Fernandez, A. Serial changes in inflammatory biomarkers after Roux-en-Y gastric bypass surgery. *Surg. Obes. Relat. Dis.* **7**, 618–624 (2011).

61. Hafida, S., Mirshahi, T. & Nikolajczyk, B. S. The impact of bariatric surgery on inflammation: quenching the fire of obesity? *Curr. Opin. Endocrinol. Diabetes Obes.* **23**, 373–378 (2016).

62. Cirulli, E. T. *et al.* Profound Perturbation of the Metabolome in Obesity Is Associated with Health Risk. *Cell Metab.* **29**, 488–500.e2 (2019).

63. Moore, S. C. *et al.* Human metabolic correlates of body mass index. *Metabolomics* **10**,

259–269 (2014).

64. Fiorenza, C. G., Chou, S. H. & Mantzoros, C. S. Lipodystrophy: pathophysiology and advances in treatment. *Nat. Rev. Endocrinol.* **7**, 137–150 (2011).
65. Huang-Doran, I., Sleigh, A., Rochford, J. J., O'Rahilly, S. & Savage, D. B. Lipodystrophy: metabolic insights from a rare disorder. *J. Endocrinol.* **207**, 245–255 (2010).
66. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **4**, Article17 (2005).
67. Liu, J., Liang, G., Siegmund, K. D. & Lewinger, J. P. Data integration by multi-tuning parameter elastic net regression. *BMC Bioinformatics* **19**, 369 (2018).
68. Wu, C. *et al.* A Selective Review of Multi-Level Omics Data Integration Using Variable Selection. *High Throughput* **8**, (2019).
69. Lindsay, T. *et al.* Descriptive epidemiology of physical activity energy expenditure in UK adults (The Fenland study). *Int. J. Behav. Nutr. Phys. Act.* **16**, 126 (2019).
70. Sanders, F. W. B. *et al.* Hepatic steatosis risk is partly driven by increased de novo lipogenesis following carbohydrate consumption. *Genome Biol.* **19**, 79 (2018).
71. Busetto, L. *et al.* Practical Recommendations of the Obesity Management Task Force of the European Association for the Study of Obesity for the Post-Bariatric Surgery Medical Management. *Obes. Facts* **10**, 597–632 (2017).
72. Adams, T. D. *et al.* Weight and Metabolic Outcomes 12 Years after Gastric Bypass. *N. Engl. J. Med.* **377**, 1143–1155 (2017).
73. Wewer Albrechtsen, N. J. *et al.* Plasma Proteome Profiling Reveals Dynamics of Inflammatory and Lipid Homeostasis Markers after Roux-En-Y Gastric Bypass Surgery. *Cell Syst* **7**, 601–612.e3 (2018).
74. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).

75. Quintin, J. *et al.* Candida albicans infection affords protection against reinfection via functional reprogramming of monocytes. *Cell Host Microbe* **12**, 223–232 (2012).

76. Leentjens, J. *et al.* Trained Innate Immunity as a Novel Mechanism Linking Infection and the Development of Atherosclerosis. *Circ. Res.* **122**, 664–669 (2018).

77. Bekkering, S. *et al.* Treatment with Statins Does Not Revert Trained Immunity in Patients with Familial Hypercholesterolemia. *Cell Metab.* **30**, 1–2 (2019).

78. Sattar, N., Forrest, E. & Preiss, D. Non-alcoholic fatty liver disease. *BMJ* **349**, g4596 (2014).

79. Hyötyläinen, T. *et al.* Genome-scale study reveals reduced metabolic adaptability in patients with non-alcoholic fatty liver disease. *Nat. Commun.* **7**, 8994 (2016).

80. Sookoian, S. & Pirola, C. J. Alanine and aspartate aminotransferase and glutamine-cycling pathway: their roles in pathogenesis of metabolic syndrome. *World J. Gastroenterol.* **18**, 3775–3781 (2012).

81. Rath, R. & Masek, J. Changes in the nitrogen metabolism in obese women after fasting and refeeding. *Metabolism* **15**, 1–8 (1966).

82. Libert, D. M., Nowacki, A. S. & Natowicz, M. R. Metabolomic analysis of obesity, metabolic syndrome, and type 2 diabetes: amino acid and acylcarnitine levels change along a spectrum of metabolic wellness. *PeerJ* **6**, e5410 (2018).

83. Samczuk, P., Ciborowski, M. & Kretowski, A. Application of Metabolomics to Study Effects of Bariatric Surgery. *J Diabetes Res* **2018**, 6270875 (2018).

84. Pacana, T. *et al.* Dysregulated Hepatic Methionine Metabolism Drives Homocysteine Elevation in Diet-Induced Nonalcoholic Fatty Liver Disease. *PLoS One* **10**, e0136822 (2015).

85. Alves, A., Bassot, A., Bulteau, A.-L., Pirola, L. & Morio, B. Glycine Metabolism and Its Alterations in Obesity and Metabolic Diseases. *Nutrients* **11**, (2019).

86. Kvaløy, K., Page, C. M. & Holmen, T. L. Epigenome-wide methylation differences in a

group of lean and obese women - A HUNT Study. *Sci. Rep.* **8**, 16330 (2018).

87. Tarkin, J. M. *et al.* Detection of Atherosclerotic Inflammation by Ga-DOTATATE PET Compared to [F]FDG PET Imaging. *J. Am. Coll. Cardiol.* **69**, 1774–1791 (2017).
88. Kieffer-Kwon, K.-R. *et al.* Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. *Cell* **155**, 1507–1520 (2013).
89. Hall, Z. *et al.* Lipid zonation and phospholipid remodeling in nonalcoholic fatty liver disease. *Hepatology* **65**, 1165–1180 (2017).
90. Szczuko, M. *et al.* Body Weight Reduction and Biochemical Parameters of the Patients After RYGB and SG Bariatric Procedures in 12-Month Observation. *Obes. Surg.* **27**, 940–947 (2017).
91. Melvin, A. *et al.* Roux-en-Y Gastric Bypass Surgery in the Management of Familial Partial Lipodystrophy Type 1. *J. Clin. Endocrinol. Metab.* **102**, 3616–3620 (2017).
92. Ciudin, A. *et al.* Successful treatment for the Dunnigan-type familial partial lipodystrophy with Roux-en-Y gastric bypass. *Clin. Endocrinol.* **75**, 403–404 (2011).
93. Grundfest-Broniatowski, S., Yan, J., Kroh, M., Kilim, H. & Stephenson, A. Successful Treatment of an Unusual Case of FPLD2: The Role of Roux-en-Y Gastric Bypass-Case Report and Literature Review. *J. Gastrointest. Surg.* **21**, 739–743 (2017).
94. McGrath, N. M. & Krishna, G. Gastric bypass for insulin resistance due to lipodystrophy. *Obes. Surg.* **16**, 1542–1544 (2006).
95. Utzschneider, K. M. & Trence, D. L. Effectiveness of gastric bypass surgery in a patient with familial partial lipodystrophy. *Diabetes Care* **29**, 1380–1382 (2006).
96. Riksen, N. P. Trained immunity and atherosclerotic cardiovascular disease. *Curr. Opin. Lipidol.* **30**, 395–400 (2019).
97. Chen, L. *et al.* Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. *Cell* **167**, 1398–1414.e24 (2016).
98. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq

quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).

99. Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res.* **4**, 1521 (2015).

100. Zerbino, D. R. *et al.* Ensembl 2018. *Nucleic Acids Res.* **46**, D754–D761 (2018).

101. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

102. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).

103. Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).

104. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–7 (2016).

105. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).

106. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

107. Landt, S. G. *et al.* ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* **22**, 1813–1831 (2012).

108. Kharchenko, P. V., Tolstorukov, M. Y. & Park, P. J. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat. Biotechnol.* **26**, 1351–1359 (2008).

109. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–5 (2016).

110. Ross-Innes, C. S. *et al.* Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389–393 (2012).

111. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime

cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

112. Müller, F. *et al.* RnBeads 2.0: comprehensive analysis of DNA methylation data. *Genome Biol.* **20**, 55 (2019).

113. Triche, T. J., Jr, Weisenberger, D. J., Van Den Berg, D., Laird, P. W. & Siegmund, K. D. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Res.* **41**, e90 (2013).

114. Maksimovic, J., Gordon, L. & Oshlack, A. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. *Genome Biol.* **13**, R44 (2012).

115. Nordlund, J. *et al.* Genome-wide signatures of differential DNA methylation in pediatric acute lymphoblastic leukemia. *Genome Biol.* **14**, r105 (2013).

116. Chen, Y.-A. *et al.* Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics* **8**, 203–209 (2013).

117. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

118. Xi, Y. & Li, W. BSMAP: whole genome bisulfite sequence MAPping program. *BMC Bioinformatics* **10**, 232 (2009).

119. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).

120. Du, P. *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* **11**, 587 (2010).

121. Du, P., Kibbe, W. A. & Lin, S. M. lumi: a pipeline for processing Illumina microarray. *Bioinformatics* **24**, 1547–1548 (2008).

122. Akalin, A. *et al.* methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* **13**, R87 (2012).

123. Wang, H.-Q., Tuominen, L. K. & Tsai, C.-J. SLIM: a sliding linear model for estimating the proportion of true null hypotheses in datasets with dependence structures. *Bioinformatics* **27**, 225–231 (2011).

124. O'Brien, K. A. *et al.* Metabolomic and lipidomic plasma profile changes in human participants ascending to Everest Base Camp. *Sci. Rep.* **9**, 2297 (2019).

125. Eiden, M. *et al.* Mechanistic insights revealed by lipid profiling in monogenic insulin resistance syndromes. *Genome Med.* **7**, 63 (2015).

126. Race, A. M., Styles, I. B. & Bunch, J. Inclusive sharing of mass spectrometry imaging data requires a converter for all. *J. Proteomics* **75**, 5111–5112 (2012).

127. Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R. & Siuzdak, G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* **78**, 779–787 (2006).

128. Tautenhahn, R., Böttcher, C. & Neumann, S. Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics* **9**, 504 (2008).

129. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

130. Carithers, L. J. *et al.* A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. *Biopreserv. Biobank.* **13**, 311–319 (2015).

131. Jain, A. & Tuteja, G. TissueEnrich: Tissue-specific gene enrichment analysis. *Bioinformatics* **35**, 1966–1967 (2019).

132. Uhlén, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).

133. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).

134. Langfelder, P. & Horvath, S. Eigengene networks for studying the relationships between co-expression modules. *BMC Syst. Biol.* **1**, 54 (2007).

135. Chong, J., Yamamoto, M. & Xia, J. MetaboAnalystR 2.0: From Raw Spectra to Biological Insights. *Metabolites* **9**, (2019).

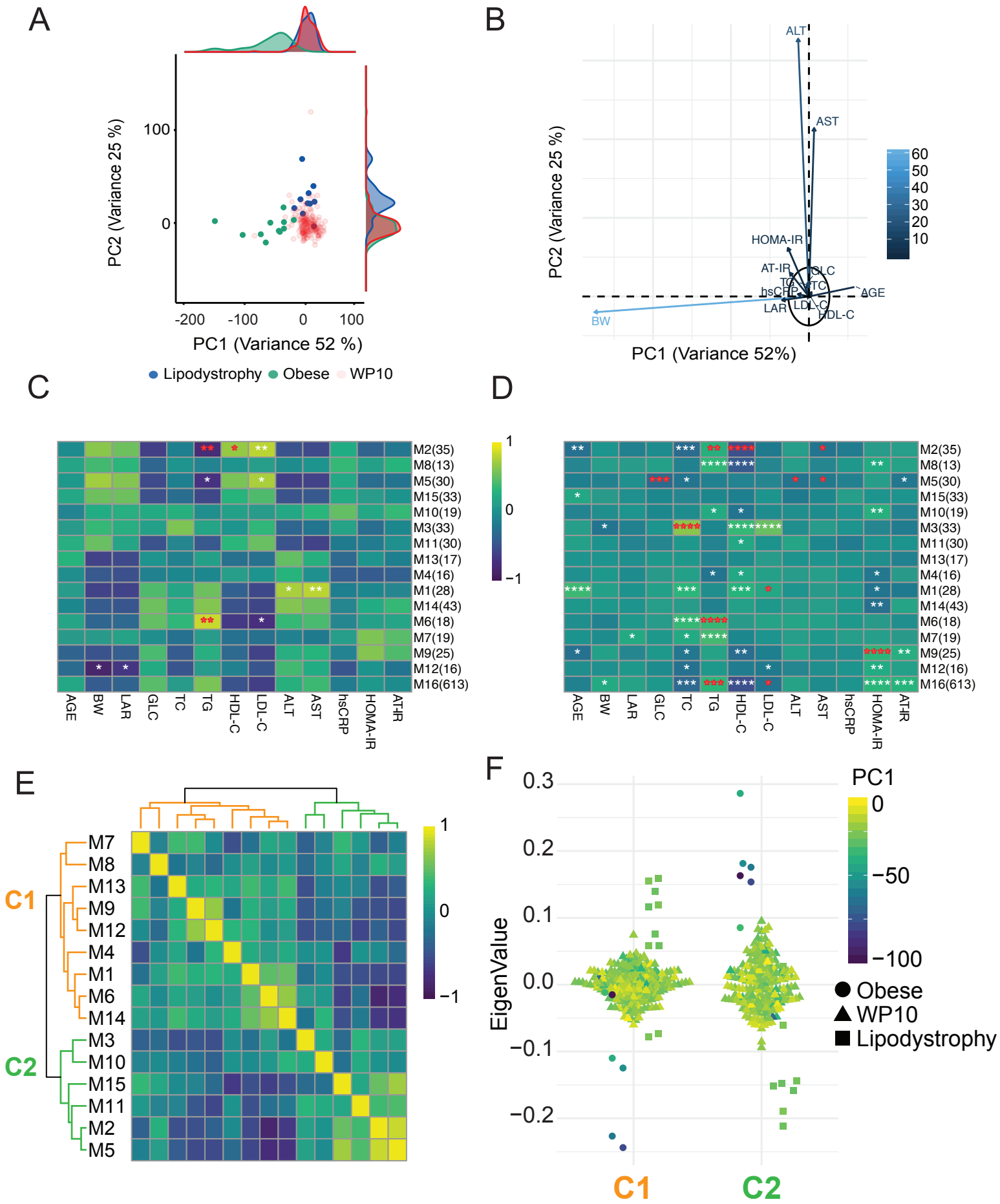


Figure 1 - High cardiovascular risk groups characterisation using anthropometric, biochemical and metabolic profiles.

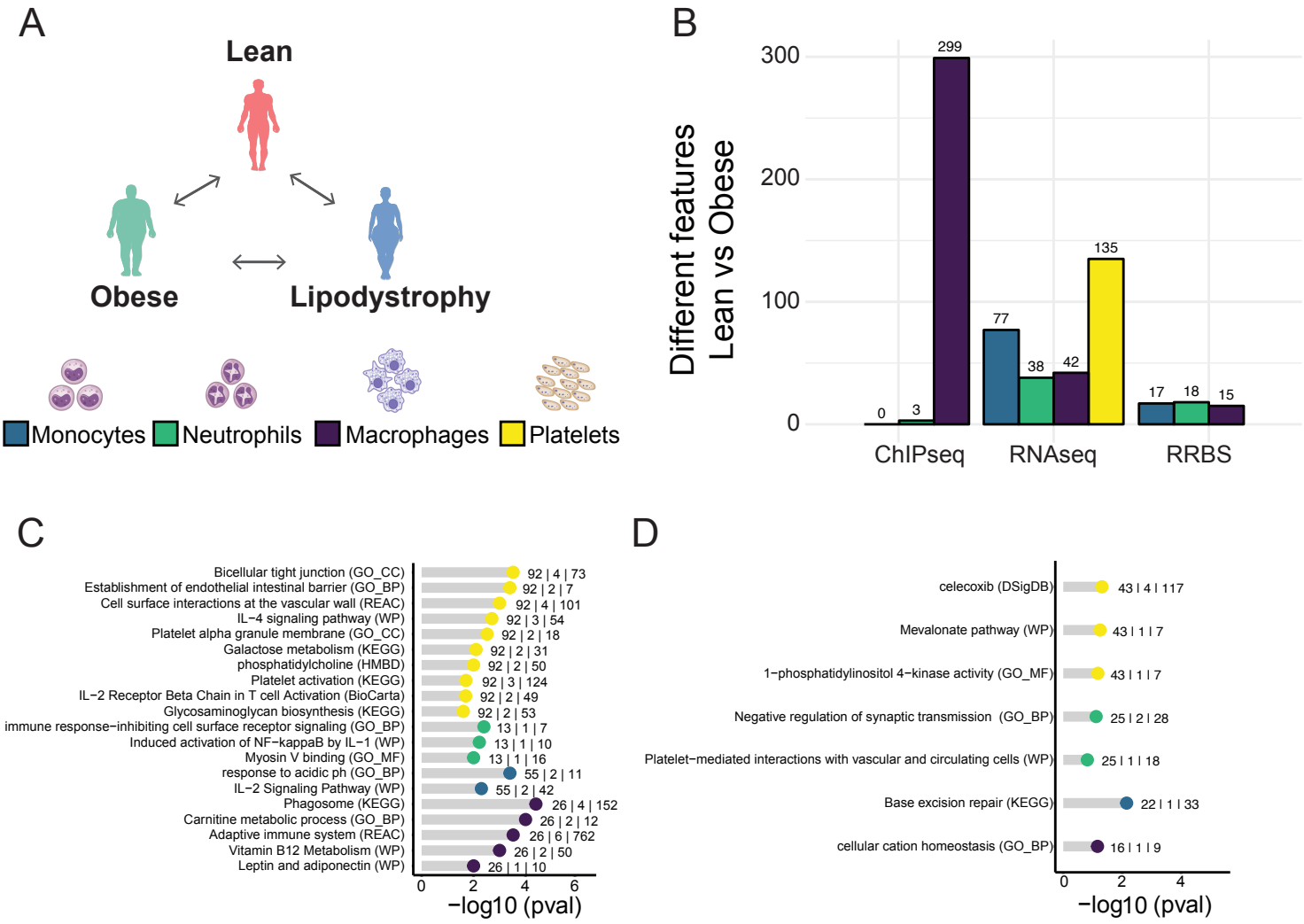


Figure 2 - Transcriptional and epigenetic comparison of obese and lipodystrophy patients versus lean individuals in 3 immune cell types and platelets.

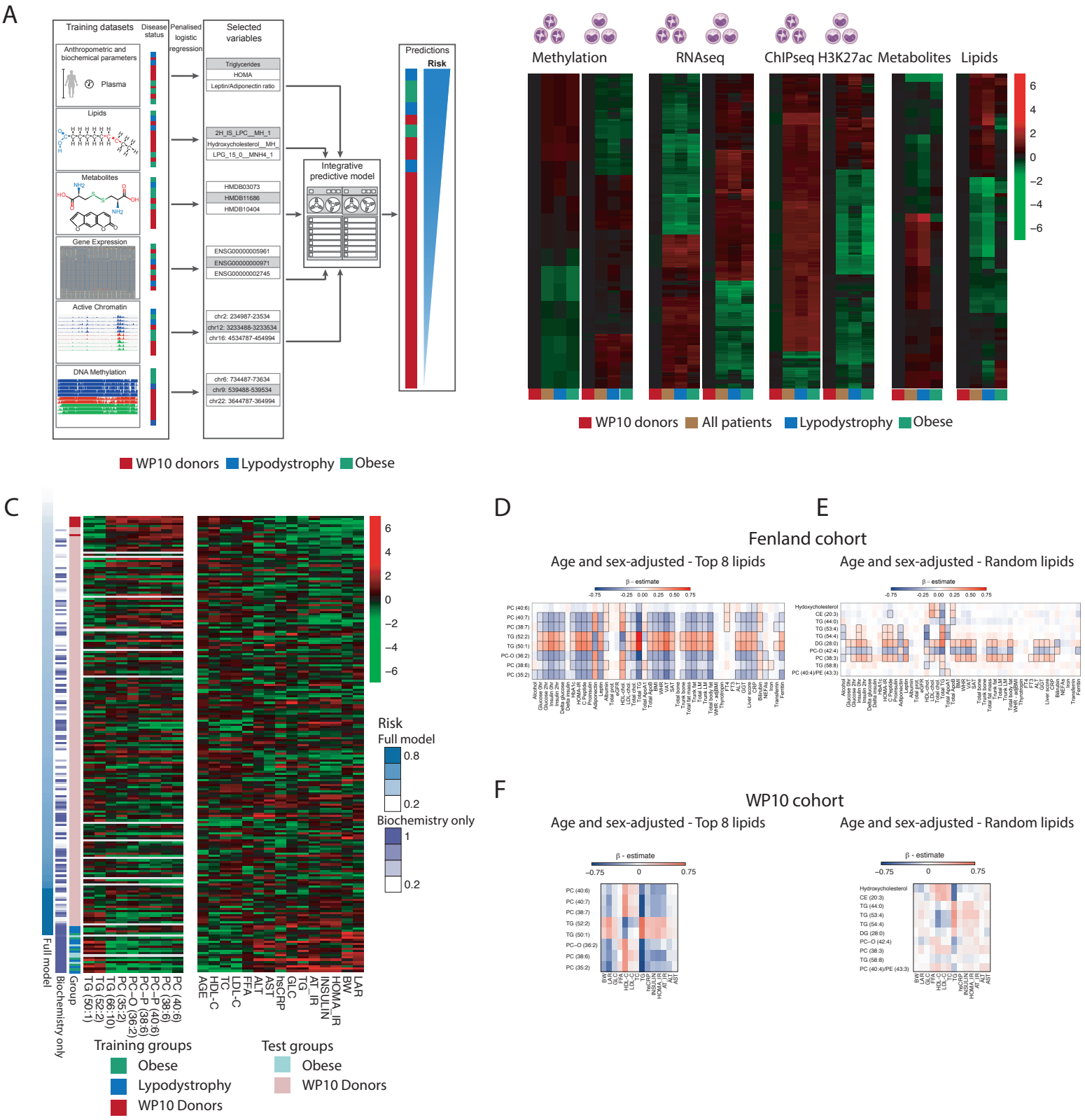


Figure 3 - Multi omics integration strategy and signature identification.

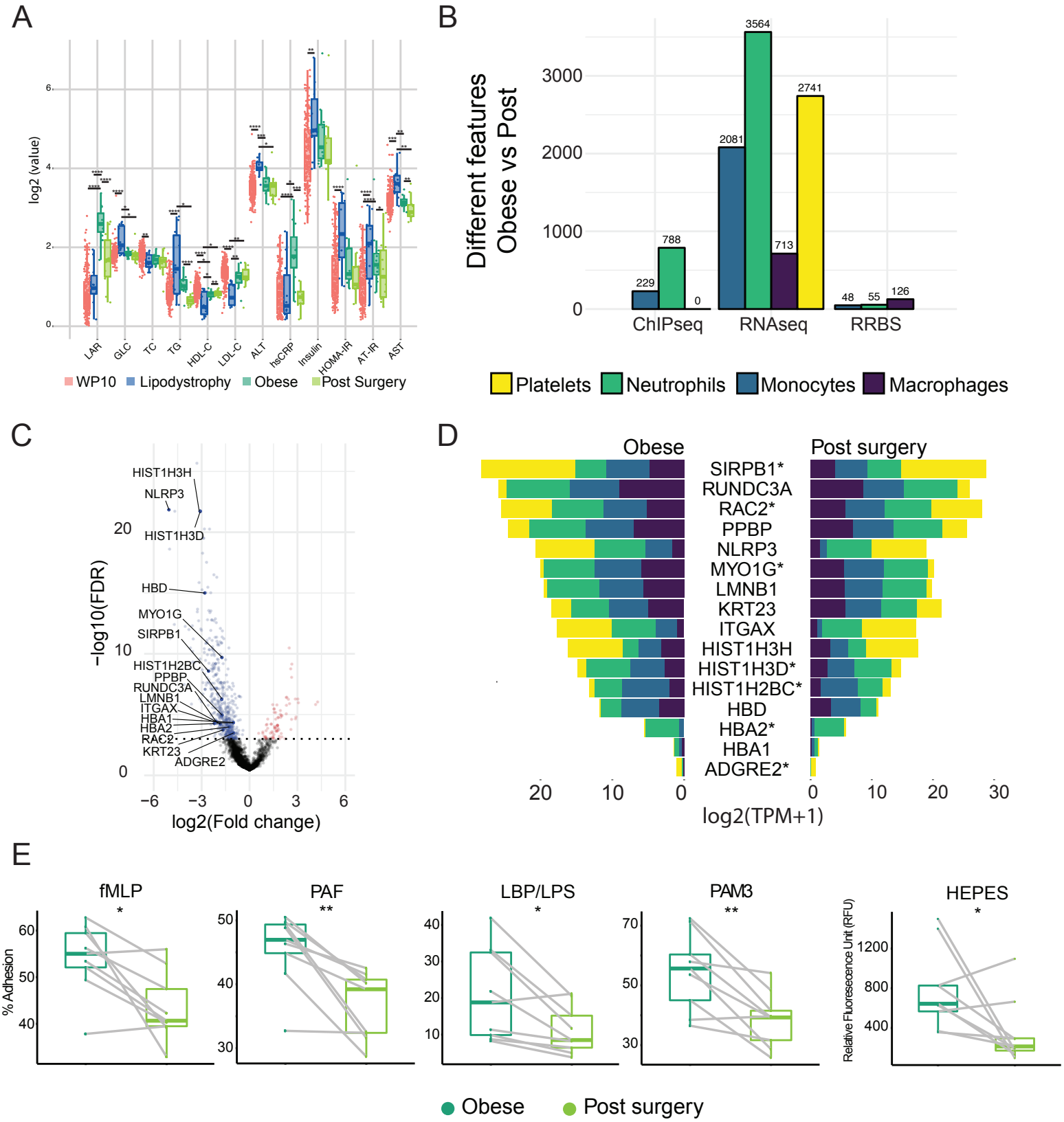
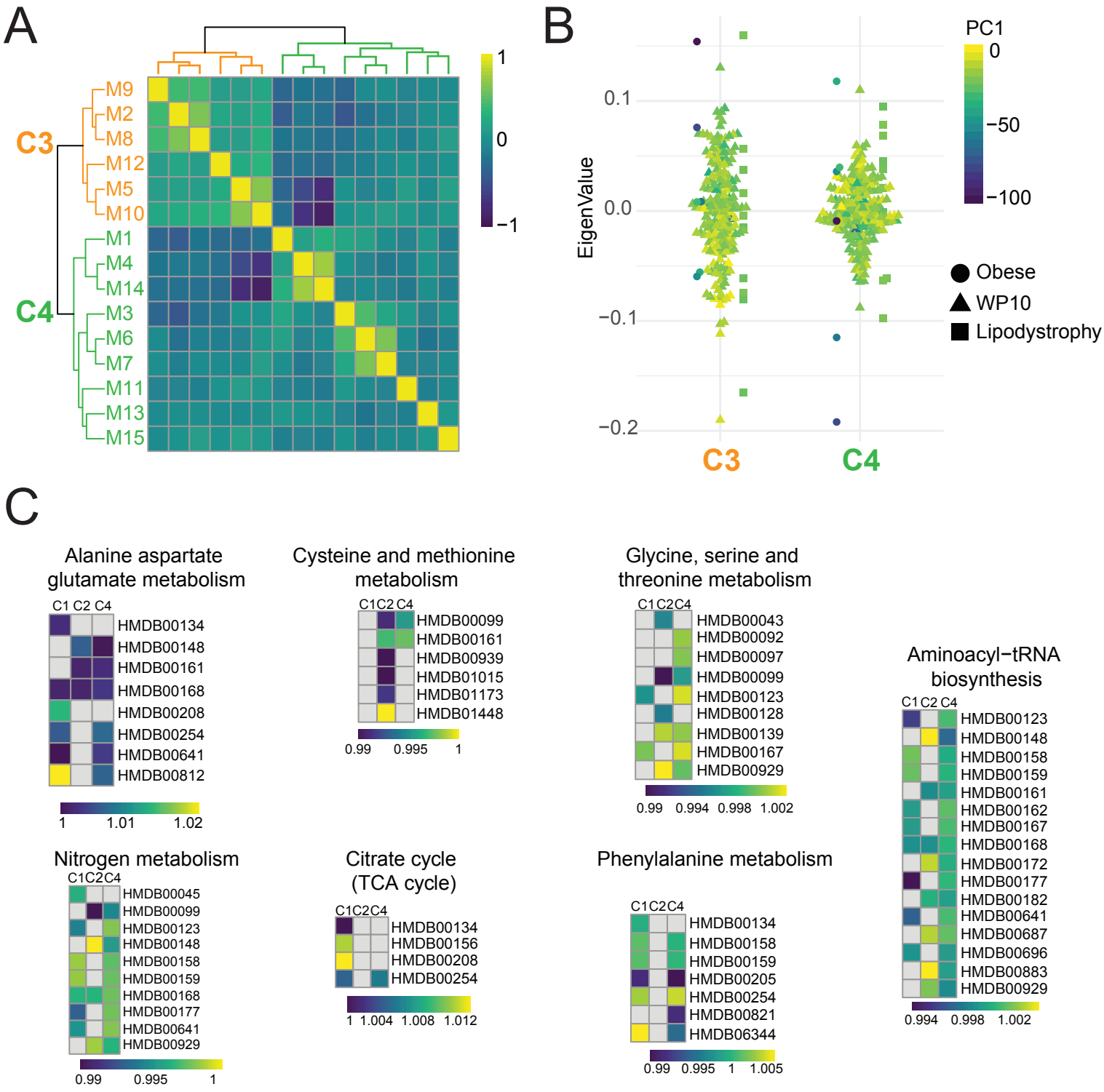
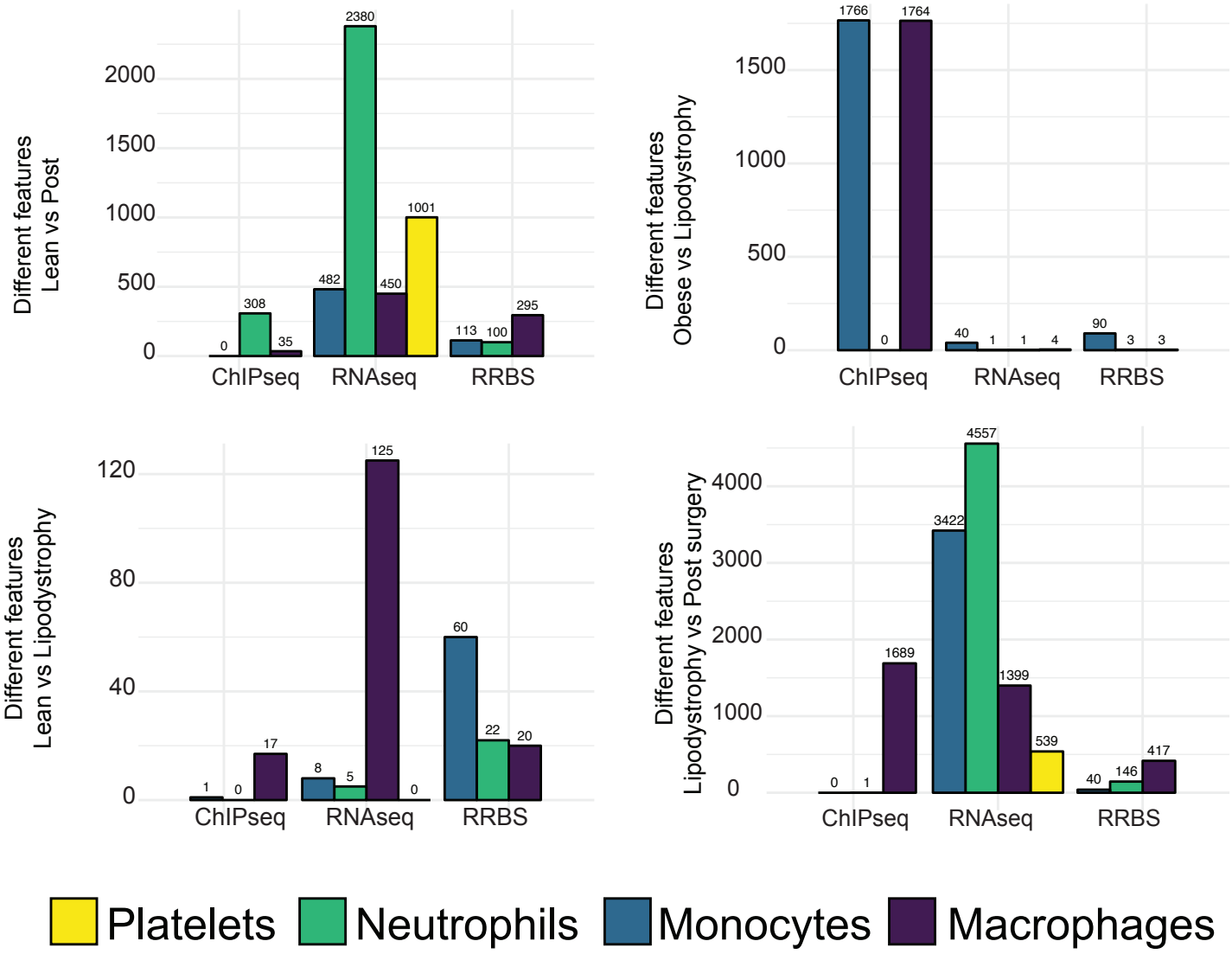


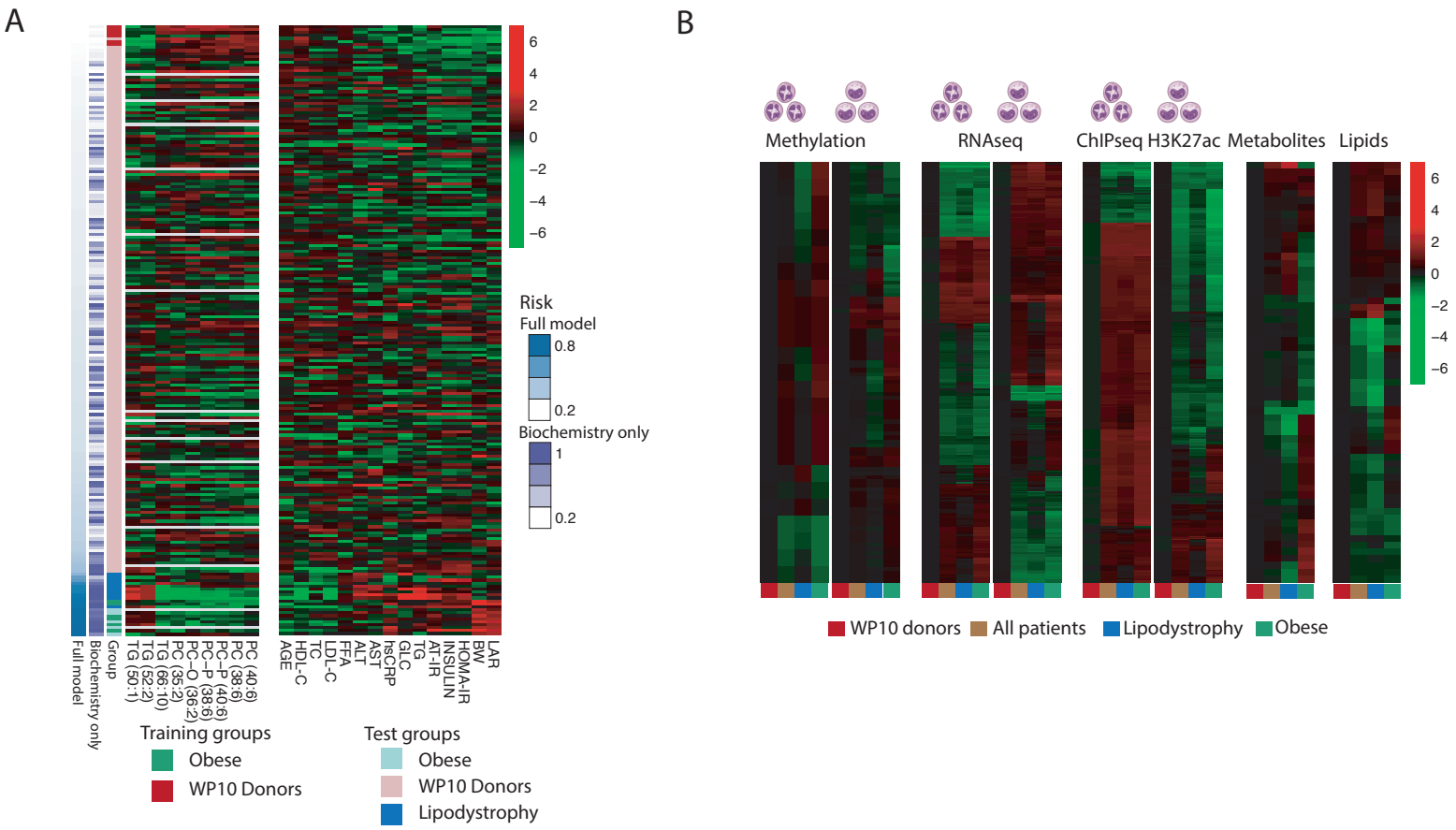
Figure 4 - Effect of bariatric surgery at molecular level in 3 immune cell types and platelets.



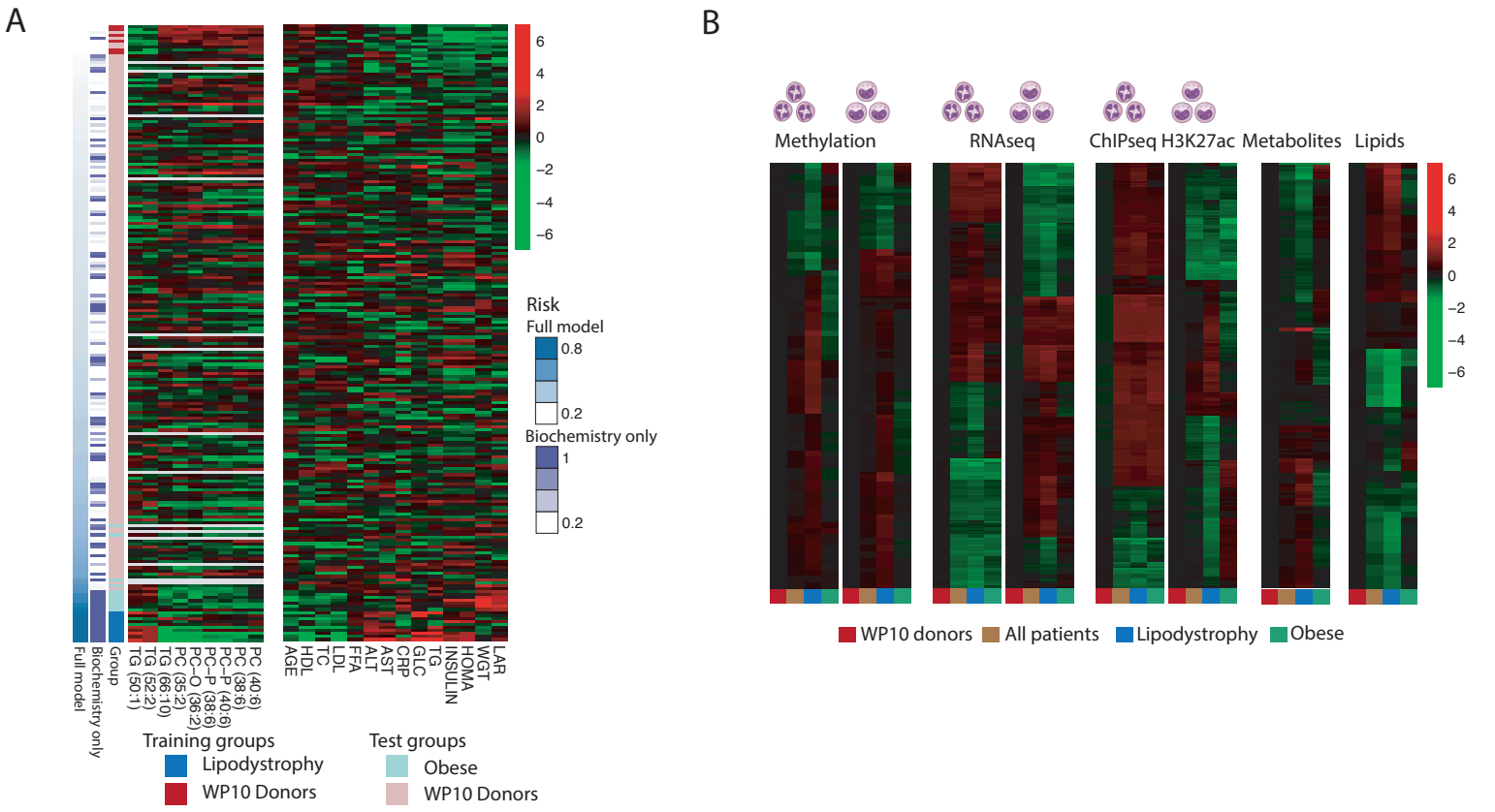
Supplemental Figure 1 - WGCNA analysis with WP10 donors metabolite values and cluster functional annotation.



Supplemental Figure 2 - Summary plots of different feature numbers in all comparisons.



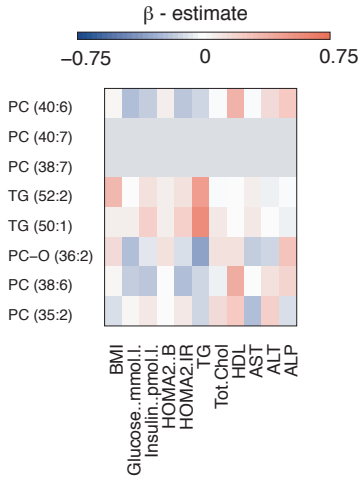
Supplemental figure 3 - Models trained using obese patients predicted lipdoystropy patients as being at high risk of disease.



Supplemental figure 4 - Models trained using lipodystrophy patients predicted obese patients as being at high risk of disease.

NASH cohort

Age and sex-adjusted - Top 8 lipids



Age and sex-adjusted - Random lipids

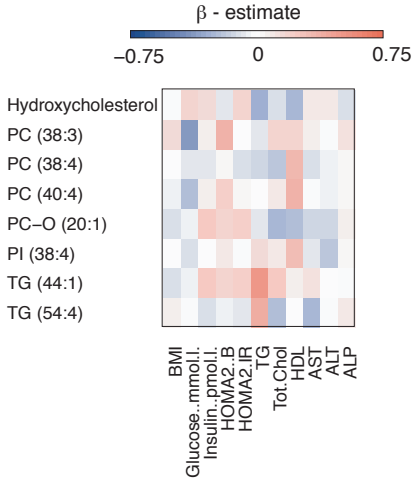


Figure S5 - Associations between lipids and outcomes in the NASH cohort.