Transcriptional, epigenetic and metabolic signatures in cardiometabolic syndrome defined by extreme phenotypes.

Denis Seyres*[1,2,3], Alessandra Cabassi[4], John J 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 A Quinn[6,7], Luca Stefanucci[2,3,8], Matthew C Sims[2,3,9], Karola Rehnstrom[2], Claire L Adams[10], Amy Frary[2], Bekir Ergüener[11], Roman Kreuzhuber[2,12], Gabriele Mocciaro[13], 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], Michael Allison[14], 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. CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria
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, NIHR Cambridge Biomedical Research Centre, 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. Ludwig Boltzmann Institute for Rare and Undiagnosed Diseases, Vienna, Austria
21. Department of Laboratory Medicine, Medical University of Vienna, Vienna, Austria
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, Mattia Frontini: mf471@cam.ac.uk, m.frontini@exeter.ac.uk

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Denis Seyres(ds777@medschl.cam.ac.uk).

Keywords

Epigenetics, Metabolites, Lipids, Multi-omics, Obesity, Lipodystrophy, Bariatric Surgery, Classification, Innate Immune Cells

Abstract

The molecular characterisation of the cardiometabolic syndrome could improve our understanding of its pathogenesis and pathophysiology and guide the identification of new treatment strategies. We identified the combinations of features that help to define the cardiometabolic syndrome using a multi-omic penalised logistic regression approach, from in-depth phenotyping, including transcriptome and epigenome profiling of innate immune cells, platelets, plasma metabolomics and extensive biochemistry, of 202 blood donors and two groups with extreme phenotypes (obese and lipodystrophy). This allowed us to determine the likelihood of the individuals in the donor group of having an increased cardio-metabolic risk and to determine the molecular mechanisms at play. To investigate if the observed effects were reversible, we repeated the in-depth phenotyping six months after bariatric surgery. These analyses revealed patterns of abnormal activation in innate immunity cells in the extreme phenotype groups, which were abrogated after surgery with the establishment of new gene expression landscapes.

Introduction

Cardiovascular disease (CVD) is the primary cause of death worldwide (17.9 million deaths in 2016, 31% of all deaths)(Kelli and Kassas, 2016) and the ever increasing number of overweight and obese individuals places a burden of hundreds of billions of dollars on healthcare systems each year(Leal *et al.*, 2006; Go *et al.*, 2014). CVD and type 2 diabetes (T2D) risk are increased in cardiometabolic syndrome (CMS)(Grundy *et al.*, 2005), a cluster of interrelated features including: obesity, dyslipidemia, hyperglycemia, hypertension and non-alcoholic fatty liver disease(Azzu *et al.*, 2020). These features have overlapping components, which include visceral fat deposition, high triglycerides, high low-density lipoprotein (LDL)-cholesterol, high fasting blood glucose, hypertension, decreased high-density lipoprotein (HDL)-cholesterol and low-grade inflammation(Hotamisligil, 2006; Alberti *et al.*, 2009; Stienstra and Stefan, 2013). Furthermore, increased prevalence of peripheral vascular disease, coronary artery disease and myocardial infarctions(McNeill *et al.*, 2005), as well as cerebro-vascular arterial diseases and stroke(Ford, 2005) in individuals developing CMS highlight the role of thrombosis in this pathology(Ninomiya *et al.*, 2004). The therapeutic

approaches to mitigate their presentation include weight loss strategies(Shimada *et al.*, 2019), lipid lowering drugs(Pahan, 2006), antiplatelet therapies(Majithia and Bhatt, 2019), glucose lowering(Drucker and Nauck, 2006; Sanchez-Rangel and Inzucchi, 2017) and anti-inflammatory drugs(Kosmas *et al.*, 2019). The relationship between cardiometabolic health and body weight is complex(Stefan, Schick and Häring, 2017). CVD risk varies between individuals of similar body mass index (BMI) depending on adipose tissue (AT) distribution and functionality(Kip *et al.*, 2004; St-Pierre *et al.*, 2005; Katzmarzyk *et al.*, 2006; Virtue and Vidal-Puig, 2010; Nichols *et al.*, 2017). AT acts as an active endocrine organ(Hotamisligil, Shargill and Spiegelman, 1993; Choe *et al.*, 2016) and when dysfunctional, plays a major role in metabolic disorders inducing peripheral insulin resistance, and contributing to chronic-low grade inflammation(Vishvanath and Gupta, 2019). When dysfunctional, AT is accompanied by macrophage recruitment (Weisberg *et al.*, 2003), altered ratio of pro- and anti-inflammatory macrophages(Aron-Wisnewsky *et al.*, 2009), enhanced production of cytokines and other pro-inflammatory signals(Visser *et al.*, 1999; Park, Park and Yu, 2005), which trigger inflammatory responses in immune cells(Choe *et al.*, 2016). AT dysfunction also has been involved in cardiometabolic disease(Lancha, Frühbeck and Gómez-Ambrosi, 2012; Lawler *et al.*, 2016) and atherosclerosis(Berg and Scherer, 2005).

Several risk score algorithms have been developed to predict the risk of complications associated with obesity (Onesi and Ignatius, 2014; Matthews *et al.*, 1985; Søndergaard *et al.*, 2017; Bedogni *et al.*, 2006; Sterling *et al.*, 2006; Artigao-Rodenas *et al.*, 2013; Hippisley-Cox *et al.*, 2008; Onesi and Ignatius, 2014), however a number of questions still remain. CVD may also occur in the absence of other comorbidities and certain type of events have a better clinical outcome in overweight and obese patients compared with their leaner counterparts (the so-called "obesity paradox")(Elagizi *et al.*, 2018).

Blood cells omics studies have helped identify pathways involved in CVD and obesity, and confirmed their utility as a source of surrogate biomarkers able to delineate the metabolic status(Ghosh *et al.*, 2010). Whilst the participation of platelets and neutrophils in thrombosis and that of macrophages in atherosclerotic plaque formation are well established(Bobryshev *et al.*, 2016; Puhr-Westerheide *et al.*, 2019; Ramirez, Manfredi and Maugeri, 2019), the role of these cell types in atherogenesis and CVD onset has been appreciated only recently(Nording, Seizer and Langer, 2015). Furthermore, prolonged exposure to low-grade inflammation modifies the functional phenotype of monocytes(an effect named trained

immunity(van Tuijl *et al.*, 2019)), platelets(Gros, Ollivier and Ho-Tin-Noé, 2014; Koupenova *et al.*, 2018) and neutrophils(Wright *et al.*, 2010; Caielli, Banchereau and Pascual, 2012). The molecular characterisation of these phenotypic changes remains incomplete, motivating the in-depth molecular phenotyping of these cells performed in the present study, including transcriptome and epigenome profiling.

Here, we present the molecular characterization of the transcriptional (RNA sequencing, RNAseq) and epigenetic (histone 3 lysine 27 acetylation, H3K27ac; reduced representation bisulfite sequencing, RRBS, and Illumina HumanMethylation450 BeadChip) changes of morbidly obese (BMI>40kg/m2) and lipodystrophy individuals in neutrophils, monocytes, macrophages and platelets.

We also developed a multi-omic data integration approach to determine the CMS associated molecular signatures, using penalised logistic regression, across multiple data layers. The resulting molecular signatures were highly predictive and, generally, in good agreement with the predictions made using plasma biochemistry markers only, although there was stark disagreement for some individuals. We used these signatures to stratify leaner individuals based on their CMS risk. The features identified in the plasma lipidomic layer were also associated with known risk factors in a large independent cohort.

Finally, to investigate the reversibility of these molecular profiles, we deployed the enhanced phenotype approach to characterise the obese group six months after bariatric surgery. We found changes in gene expression, especially in neutrophils and platelets, accompanied by more modest differences in regulatory elements usage and almost no differences in DNA methylation profiles. Plasma proteome analysis allowed us to determine changes in other tissues and organs, whilst neutrophil and platelet cell function assay results indicated reduced ability to adhere, the key initial step during their activation.

Results

Metabolic signatures in the obese and lipodystrophy groups.

Four different groups of participants were recruited: healthy day controls (N=20; from which metabolically healthy where then selected), familiar partial lipodistrophy type 2 (N=10; hereafter lipodystrophy; carrying mutations in PPARG or LMNA genes, as verified by whole genome sequence(Turro *et al.*, 2020)), obese referred for bariatric surgery (N=11; without

clear genetic causes(Turro *et al.*, 2020)) and blood donors (hereafter "BD"; N=202)(Chen *et al.*, 2016). To determine the metabolic health of the individuals in these groups, we collected data on age and body weight (BW), and performed plasma biochemistry assays for the following (all values and differences reported in **Table S1**): leptin, adiponectin, insulin, free fatty acid (FFA), glucose (GLC), serum lipid (triglycerides (TG), total cholesterol (TC), high density lipoprotein (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) indices. The lipodystrophy group had elevated GLC, TC, TG, ALT, AST, insulin (and consequently HOMA-IR and AT-IR; HDL-C and LDL-C were decreased. The obese group had elevated LAR, LDL-C and hsCRP. To illustrate the influence of these parameters, we performed a principal component analysis (PCA). PCA is a statistical method which allows the visualization of the information content in large data tables by means of a smaller set of factors or principal components (PCs). This showed that obese, lipodystrophy and BD groups were distributed over distinct, albeit partially overlapping dimensions (**Fig.1A**). The first two components (PC1 and PC2) were sufficient to distinguish the different groups (Obese/Lipodystrophy: p value=0.002; Obese/BD: p value< 2.2e-16; Lipodystrophy/BD: p value< 2.2e-16; Hotelling's T-squarred test with F distribution). In particular the extreme phenotype groups (obese and lipodystrophy) were separated from BD participants along PC1, whilst the obese and the lipodystrophy groups were separated from each other along PC2.

Loading and contribution analysis (**Fig.1B**, arrow length and direction) showed that BW, LAR, hsCRP, AST, ALT, GLC, AT-IR, HOMA-IR and TG were the main contributors to the separation along PC1. Additionally, BW, LAR, hsCRP separated the obese from the lipodystrophy groups in one direction, whilst AST, ALT, GLC, AT-IR, HOMA-IR and TG separated them in the opposite direction along PC2. Obesity(Moore *et al.*, 2014; Cirulli *et al.*, 2019) and lipodystrophy(Huang-Doran *et al.*, 2010; Fiorenza, Chou and Mantzoros, 2011) have been shown to be associated with plasma metabolite levels. To determine which metabolites were present and their abundance, plasma samples were analysed on the Metabolon® platform (see methods), which identified and quantified 988 species. To cluster and then characterise groups of metabolites, whose levels were correlated across samples, we performed a weighted gene co-expression network consensus analysis (WGCNA)(Zhang and Horvath, 2005). We identified

16 consensus modules (M1 to M16; **Table S2**) and we determined if there were any associations between these modules with age, BW and the plasma biochemistry assays results. Of the 208 tested associations, 11 showed significant correlation with the results of the plasma biochemistry assays in the extreme phenotypes group (FDR adjusted Fisher p values < 0.05; six were positively correlated, 5 were negatively correlated; **Fig.1C**); where none of the modules showed significant correlation with the results of the plasma biochemistry assays in the BD cohort (not shown). These modules also showed no enrichment for biological pathways, which could be partly explained by the presence of cryptic metabolites not yet functionally annotated and that cannot be mapped to existing databases. Nevertheless, taken individually, some of their metabolites were known to be associated with their respective traits. For example, in the module M6, which was associated with both TG and LDL-C, we found: eicosapentaenoic and docosahexaenoic acids already known to be associated with cardiometabolic risk factors(Innes and Calder, 2018), 3-Carboxy-4-methyl-5-propyl-2-furanpropanoic acid known to be associated with non-alcoholic fatty liver disease (NAFLD) (Dai *et al.*, 2019) and L-cysteine known to be associated with T2D(Jain *et al.*, 2014) all being negatively associated with TG and positively associated with LDL levels. By extension, this suggested new similar roles for the phosphatidylcholine (PC) and lysophospholipid (LysoP), molecules present in this module and not yet annotated, and also for homo-L-arginine which was known to be associated with cirrhosis and hyperargininemia (full list of modules and metabolites in **Table S2**).

To determine which of these modules could be associated with disease status, we analysed the eigen-metabolite adjacencies. Consensus modules formed two clusters, C1 and C2 (**Fig.1D**) using extreme phenotype group samples whereas, two distinct clusters, C3 and C4, were found using BD samples only (**Fig.S1A**). To determine if C1 and C2 represented the obese and lipodystrophy groups, we plotted the average eigen-metabolite value for each cluster. This showed that the extreme phenotypes could be identified by their distinct metabolic signature (**Fig.1E**), whereas clusters C3 and C4, defined by different metabolites, were unable to discriminate between them (**Fig.S1B**). When the same pathway was found enriched in multiple clusters (**Table S2**; i.e. aminoacyl-tRNA biosynthesis or glycine, serine and threonine metabolism), it was because different metabolites in the same pathway were found in different clusters (**Fig.S1C**).

In summary, analysis of serum metabolites abundance, in individuals at high CMS risk, revealed patterns correlating with anthropometric and plasma biochemistry parameters that are disease specific.

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

Next, we sought to determine if the metabolic changes observed in plasma were accompanied by changes at transcriptional and epigenetic levels in innate immune cells (neutrophils, monocytes, macrophages) and platelets, as these are some of the key players in atherogenesis and thrombus formation(Swystun and Liaw, 2016) (**Fig.2A**). We characterised: transcriptome by ribo depleted total RNAsequencing (RNAseq), genome-wide distribution of histone 3 lysine 27 acetylation (H3K27ac), a marker of active promoters and transcriptional enhancers, by chromatin immunoprecipitation (ChIPseq) and DNA methylation by reduced representation bisulfite sequencing (RRBS). First, we determined which, amongst the participants recruited as day controls, were the metabolically healthy individuals (hereafter "lean", selected using the following parameters: 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) and used these in all the following comparisons (**Table S7**).

The comparison between the obese and lean groups (**Fig.2B)** led to the identification of 191 differentially expressed genes (DEG) in macrophages (142 up and 49 down regulated in obese individuals, **Table S8**), 80 in monocytes (45 up and 35 down regulated, **Table S9**), 79 in neutrophils (30 up and 49 down regulated, **Table S10**) and 120 in platelets (71 up and 49 down regulated, **Table S11**); at a FDR 5%. We identified only one differentially acetylated region (DAcR; **Fig.2B**) below 5% FDR in each of the cell types (**Tables S12-S14**). DNA methylation analysis found 17 differentially methylated CpG islands (0.5% FDR) in macrophages (**Table S15**); 14 in monocytes (**Table S16**) and 13 in neutrophils (**Table S17**).

The comparison between the lipodystrophy and lean groups (**Fig.2C**) identified 194 DEG in macrophages (139 up and 55 down regulated in lipodystrophy, **Table S8**), 90 in monocytes (51 up and 39 down regulated, **Table S9**), 284 in neutrophils (160 up and 124 down regulated, **Table S10**) and 60 in platelets (29 up and 31 down regulated, **Table S11**). DAcR were observed in macrophages and monocytes: 402 DAcR (392 up and 10 down) in the former (**Table S12**), 9 of these were located on gene promoter, 176 (43%) were intergenic and 195 (48%) in introns

and one in the latter (up regulated; intergenic). No DAcR were identified in neutrophils. DNA methylation analysis found 11 differentially methylated CpG islands in macrophages (**Table S15**); 46 in monocytes (**Table S16**) and 16 in neutrophils (**Table S17**).

The comparison between the obese and lipodystrophy groups (**Fig.S2B)** identified 26 DEG in macrophages (17 up and 9 down regulated in lipodystrophy, **Table S8**), 24 in monocytes (15 up and 9 down regulated, **Table S9**), 59 in neutrophils (34 up and 25 down regulated, **Table S10**) and 79 in platelets (33 up and 46 down regulated, **Table S11**). We observed 17 DAcR in macrophages (all hyper-acetylated in lipodystrophy; **Fig.S2B and Table S12**), 7 (41%) were intergenic and 9 (52%) in introns. We also observed 3,616 DAcR in monocytes (2,234 up and 1,382 down; **Fig.S2B and Table S13**), 1,259 (35%) of these were located in UTRs, TSSs, TTSs or exons, 301 (8%) were intergenic and 1,798 (49%) in introns. No DAcR were found in neutrophils.

In macrophages, DNA methylation analysis found 6 CpG islands differentially methylated (2 up and 4 down, **Table S15**), 91 in monocytes (50 up and 41 down, **Table S16**) and 35 in neutrophils (14 up and 21 down, **Table S17**).

To gain insight into the gene expression changes, we performed functional annotation by gene ontology (GO) terms enrichment analysis. In the comparison between the obese and lean groups (**Fig.2D**), we found an enrichment for GO terms related to interferon alpha/beta signalling pathway, as well as focal adhesion in DEG up-regulated 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 genes responding to antithrombotic drugs (**Table S20**). In the comparison between the lipodystrophy and lean groups (**Fig.2D**), macrophages up-regulated DEG were enriched for GO terms related to cholesterol biosynthesis and immune response activation. In monocytes and neutrophils, up-regulated DEG were enriched in GO terms related to interferon signalling and immune response. The above pairwise comparisons provided insight on the differences existing in the different omics layers. It provided reliable multi-omic layers to integrate in order to identify combinations of features (i.e. molecular signatures) from across all layers that characterise the extreme phenotype groups.

Multi-omic signature classification of extreme phenotypes

Multivariable selection approaches have provided an effective means to integrate multiple omics layers and elucidate disease signatures(Liu *et al.*, 2018; Wu *et al.*, 2019). We used one such approach to integrate RNAseq, H3K27ac histone modification, DNA methylation, metabolic and lipidomic datasets. We identified a training set comprising: (i) a control group of 6 BD (selected using the criteria used to define the lean group above); and (ii) a group of 6 obese individuals 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 (Zou and Hastie, 2005) to identify signatures associated with an increased probability of belonging to the extreme phenotype obese group and therefore having some or all of the features associated with CMS (**Fig.3A**). We chose to use this approach, since we have a large number of variables, but small sample size (see, for example, (Waldmann *et al.*, 2013)). The values taken by the variables selected into each signature defined patterns characterising the groups (**Fig.3B**; **Table S24-26**).

We used the variables selected for each signature, together with the biometric variables, to construct multivariable logistic regression models to predict if an individual would be better classified as belonging to the control group or the obese group. We considered predictions made by models trained on each layer individually, as well as, a multi-layer predictive model. All models allowed us to rank individuals according to their probability of belonging to the obese group (**Fig.3C**, **Fig.S3A**). To quantify each model's ability to discriminate individuals in the extreme phenotype groups from all other individuals, we calculated the log loss (also known as the cross-entropy loss; **Fig.3C** and **Fig.S3A**)(Murphy, 2012). This demonstrated that the multi-layer model provided the best separation, followed by the models trained on the RNAseq layers, then the models trained on the metabolite and methylation layers. We also performed this analysis using the lipodystrophy and lean groups to train the multi-layer predictive model. In this case we found that obese individuals were often predicted to belong to the lipodystrophy group (**Fig.S3B**).

The extent of the possible validation of the model we proposed, using existing datasets, was rather limited and we acknowledge that further multi-omic studies will be required for a complete evaluation in external cohorts, as well as to establish the most cost-effective combination of data layers in respect of their predictive power. However, since lipidomic data from external cohorts were already available, we focused on investigating the lipidomic

signature. We prioritised a reduced set of 9 lipid species from the signature to test for univariate association with known CMS risk factors (see Materials and Methods for details, and **Fig.3D** for results). Of these nine species, we matched eight with lipid species that had also been measured in a subset of 1,507 participants of the Fenland study(Lindsay *et al.*, 2019), 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. In the Fenland cohort, the null hypothesis of no association was rejected in 225 out of 368 tests (≈ 61%) after correcting for multiple testing. We found TG (52:2) and TG (50:1) to be positively associated with several risks factors, such as fasting plasma glucose, fasting insulin level, HOMA-IR, a fatty liver index, HbA1c, leptin, LDL-C, hsCRP, TG, BMI, fat mass, ALT, and ferritin (**Table S27**). 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 lipid species, we repeated the analysis with a set of five lipid species randomly selected from lipid species not included in the predictive signature. Far fewer associations were found to be significant (49 out of 230 tests, ≈ 21%). We observed the same pattern of associations between the prioritised lipid species and known CMS risk factors using data from the present study (**Fig.3D; Table S27**), as well as in a biopsy-confirmed non-alcoholic steatohepatitis (NASH) cohort comprising 73 individuals (Sanders *et al.*, 2018) (**Fig.S4**; **Table S27**).

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(Busetto *et al.*, 2017), with well-established long-term benefits on weight loss, diabetes, hypertension and dyslipidemia(Adams *et al.*, 2017). 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 expand our ability to detect changes happening elsewhere in the body. To this aim, a second blood sample from obese individuals 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, to allow us to adjust for age and sex; **Fig.4A; Table S1** for other comparisons). We next compared the transcriptional and epigenetic profiles in monocytes, neutrophils, macrophages and platelets before and after bariatric surgery (**Fig.4B**). We identified, using paired analysis, 599 DEG in macrophages (340 up and 259 down regulated; **Table S8**); 1,931 in monocytes (1,118 up and 813 down regulated; **Table S9**); 2,571 in neutrophils (556 down and 2015 up regulated; **Table S10**) and 2,883 in platelets (1,223 up and 1,660 down regulated; **Table S11**). No DAcR were found in macrophages (**Table S12**), 229 in monocytes (139 up and 90 down regulated; **Table S13**) and 788 in neutrophils (13 up and 775 down regulated; **Table S14**). RRBS analysis found 201 differentially methylated CpGs in macrophages (97 up and 104 down regulated; **Table S15**); 48 in monocytes (32 up and 16 down regulated; **Table S16**) and 198 in neutrophils (139 up and 58 down regulated; **Table S17**).

Overall, GO terms enrichment for DEG identified the following processes: ribosome formation, metabolism of amino acid and proteins, several immune related pathways and cytoplasm translation were up-regulated whereas regulation of cholesterol metabolic process (through SREBF and miR33(Horie *et al.*, 2013)) and mRNA processing pathways were down-regulated (**Table S18**, **S19**, **S20** and **S21** for macrophages, monocytes, neutrophils and platelets, respectively). The results of the other 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. We found 9 (RHPN1, DGKQ, TCTEX1D2, MVD, LDLR, BCAR1, ANKRD33B, FASN, COL5A3; overlap p value = $3.6*10^{-8}$, hyper-geometric test), 7 (EPB41L3, LRRC8B, STARD4, ZNF331, SEMA6B, DSC2, RGPD8; overlap p value = $5*10^{-6}$), 5 (NAIP, RP11-1319K7.1, LINC01271, LINC01270, DNAH17; overlap p value = $1.3*10^{-5}$) and 10 (CTC-429P9.4, XXbac-BPG300A18.13, RP11-386G11.10, MT-TG, TVP23C-CDRT4, SHE, MPZL3, CLIP1, RGPD1, RPL23AP7; overlap p value=$6.5*10^{-5}$) genes, in macrophages, monocytes, neutrophils and platelets respectively, that after having been upregulated in obese individuals compared to lean individuals, returned to similar expression level after weight loss. Intriguingly, key genes involved in lipoprotein metabolism (LDL receptor, LDLR), fatty acid synthesis (Fatty acid synthase, FASN) in macrophages and cholesterol transport (StAR-related lipid transfer protein 4, STARD4) in monocytes were modulated. We found 2 (SNHG5, EVI2A; overlap p value = 0.03) and 3 (XXbac-BPG32J3.22,

MEIS2, MS4A14; overlap p value = 0.03) genes with the opposite progression in macrophages and monocytes respectively.

To obtain further insight into the effects of bariatric surgery on gene expression and signaling pathways in other tissues and organs, we investigated plasma proteins level 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 signalling pathways (**Table S29**), of note, no significant changes in full blood count except mean platelet volume (p value = 0.03; paired t-test) and eosinophils count (p value = 0.03) were observed (**Table S1**). 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 database(GTEx Consortium, 2015) to extract tissue specific genes, these ranged from 286 in the heart left ventricle to 1,286 in the spleen (**Table S30 and methods**). Tibial, coronary and aorta artery, heart atrial appendage and left ventricle and blood displayed an enrichment of tissue specific genes amongst DAP (p values: $1.6*10^{-2}$, $8*10^{-3}$, $2*10^{-2}$, $1.8*10^{-2}$, $1.6*10^{-2}$ and $5*10^{-2}$, respectively; hyper-geometric test; **Table S30**). Of the 13 blood specific genes that encoded for a DAP, six were also differentially expressed in at least one of the four studied cell types (**Fig.4D**). The data generated in monocytes and macrophages allowed us to explore the effect of bariatric surgery on trained innate immunity(Quintin *et al.*, 2012), as it has been shown that it plays a role in atherosclerosis(Leentjens *et al.*, 2018; Bekkering *et al.*, 2019). We found overlaps between the genes associated with the top 500 regions that gain histone 3 lysine 4 trimethylation (H3K4me3) both after β-glucan treatment (BG) or untreated (NT) in Quintin and colleagues(Quintin *et al.*, 2012) and the DEG found comparing lean individuals and obese individuals post surgery (overlap p value = 0.015 and $4.49*10^{-6}$ respectively; hypergeometric test). These conditions had also significant overlap with DEG before and after bariatric surgery (overlap p value = 0.018 and $8.2*10^{-13}$ respectively; hyper-geometric test). The overlap was more significant if we considered only genes downregulated after bariatric surgery and NT

top 500 H3K4me3 genes (overlap p value = 0.005) and the DEG found comparing obese individuals before and after bariatric surgery (**Table S31**).

Lastly, 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 (**Fig.4E**). 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 surface exposure 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 CD41b) and CD36, whereas no changes were observed for CD49b, CD42a, CD42b, CD29 and CD9 (paired t-test, all result in **Table S32**).

Discussion

We successfully deployed an enhanced phenotyping strategy to obtain CMS signatures from extreme phenotype groups (morbid obesity and lipodystrophy) and we developed an integrative multi-omic strategy to use these signatures to determine the cardiometabolic status of a cohort of blood donors, aged between 40 and 70 (Table S1), which represent a group at high risk of this disease.

Biochemical and Metabolic group characterisation.

Our data showed that some individuals in the BD group were difficult to distinguish from those in the extreme phenotype groups on the basis of plasma biochemistry and, not surprisingly, BW (**Fig.1A**). The main contributors to the separation of the extreme phenotype groups from the bulk of BD and between each other were, in agreement with the existing literature, BW, LAR, hsCRP singling out the obese group. Whereas, ALT, AST, GLC, TG, HOMA-IR and AT-IR had a larger contribution towards the separation of the lipodystrophy group (**Fig.1B**). This indicates that AT dysfunction is a shared component among the two groups, but with a different influence on each of them.

We then investigated plasma metabolite levels to refine the separation. Metabolites can be separated in 16 distinct modules, which contain different species in the extreme phenotypes

and in BD groups (**Table S2**). Using a network-based approach, we determined the relationships between modules and plasma biochemistry parameters, this revealed the association between some of the 16 modules and BW, TG, HDL-C, LDL-C, AST and ALT (**Fig.1C**), whereas no associations were found for the modules in the BD group (**not shown**). For each module associated with a trait, several metabolites had already been shown to be linked to it. By extrapolation, we could infer that other metabolites belonging to that module were also associated with the trait. The analysis of module correlations defined two different clusters, C1 and C2 (**Fig.1D**), each representing one of the extreme phenotype groups (**Fig.1E**), thus demonstrating that the separation of the two groups was observed also at this level. Not unexpectedly, the two clusters were enriched for modules previously associated with NAFLD(Sookoian and Pirola, 2012; Hyötyläinen *et al.*, 2016; Pacana *et al.*, 2015; Alves *et al.*, 2019) and with obesity(Rath and Masek, 1966; Libert, Nowacki and Natowicz, 2018; Alves *et al.*, 2019).

Transcriptional and epigenetic signatures.

The comparison between extreme phenotypes and lean groups showed a modest impact on H3K27ac levels and its distribution, on DNA methylation profiles and on gene expression, in the absence of acute challenge, in neutrophils, monocytes macrophages and platelets (**Fig.2B and 2C**). The largest number of changes observed in active chromatin (3,616 DAcR) was found in macrophages, between the obese and lipodystrophy groups (**Fig.S2B**). These differences were not accompanied by as many changes at transcriptional level, indicating that the same transcriptional output can be achieved with different regulatory landscapes, as previously shown(Kieffer-Kwon *et al.*, 2013) or that these cells have been primed differently in the two groups to have different responses upon stimulation.

This observation was also in agreement with the absence of a significant overlap between DEG and genes previously associated with trained immunity(Quintin *et al.*, 2012) when comparing lean versus obese and lean versus lipodystrophy groups. However limited, the observed changes were specific enough to highlight, in functional enrichment analysis, GO terms related to inflammatory response thus confirming that the underlying chronic inflammation associated with these conditions modifies the molecular phenotype of the cell types involved in the development of atherosclerosis and in thrombus formation (**Fig.2D**). Similar results have previously been reported for whole blood cell DNA methylation(Kvaløy,

Page and Holmen, 2018), while more extensive changes have been observed in adipose tissue(Ling and Rönn, 2019).

Overall, across all comparisons, with the exception of H3K27ac in macrophages, we found that the two extreme phenotype groups are more similar to each other than to BDs.

Multi-omic signatures

Given the limited number of changes identified using layer-by-layer univariate comparisons, we sought to identify multivariable signatures characterising the extreme phenotypes and lean groups to better discriminate between the different groups by combining information from across omic layers (**Fig.3A**). Using the obese and lean groups as training data, we identified signatures for each layer (**Fig.3B**) and trained classification models on each layer separately, as well as training a multi-layer integrative model. As quantified by the log loss, we found the multi-layer model to provide the clearest separation between the extreme phenotype and donor groups, followed by models trained on the individual RNAseq, metabolite and methylation layers (**Fig.3C** and **S3**). This indicates that, while each individual's rank differs in the single layer models, those individuals that are predicted by the integrative model to belong (with a high probability) to the extreme phenotype groups, are the ones with the most similarity, across all layers, of the lipodystrophy and obese groups. The differences observed when comparing the models trained using lean vs obese and lean vs lipodystrophy were likely reflecting the higher heterogeneity in the obese group which was lacking a high penetrance genetic cause.

It should be noted that among the lipids prioritised by our analysis, PC (38:6) and PC (36:2) have already been identified in obesity studies(Hall *et al.*, 2017); and, TG (50:1) and TG (52:2) have previously been linked to NAFLD(Lindsay *et al.*, 2019) and NASH(Sanders *et al.*, 2018). Moreover, we showed the diagnostic value of the prioritised lipid species through their association with major cardiometabolic risk factors in the Fenland study and in the present study (**Fig.3D**); as well as, albeit not significantly due to the small sample size, in the NASH cohort (**Fig.S4**).

Effect of bariatric surgery.

Bariatric surgery is an effective therapeutic approach to reduce weight and improve wellbeing for morbidly obese individuals(Maggard *et al.*, 2005; Karlsson *et al.*, 2007) and it has been

shown to have beneficial effects in lipodystrophy patients with BMI < 30(McGrath and Krishna, 2006; Utzschneider and Trence, 2006; Ciudin *et al.*, 2011; Grundfest-Broniatowski *et al.*, 2017; Melvin *et al.*, 2017).

We found that bariatric surgery not only induced significant changes in plasma biochemistry parameters, as already established(Wewer Albrechtsen *et al.*, 2018), but it had a profound effect on transcriptional and active chromatin landscapes. We found genes that, after bariatric surgery, revert to the expression level observed in lean individuals. However, the majority of genes whose expression levels changed after surgery either did not revert to the values observed in lean individuals or was not differentially expressed between the obese and lean group. This suggests that the reduction in inflammatory signatures observed in macrophages, monocytes, neutrophils, and platelets after bariatric surgery is mostly due to the establishment, at least in the short term, of novel gene expression landscapes.

The overall small number of changes in DNA methylation observed, together with the short life span of the cell types analysed, indicates that reverting from a pro-inflammatory to a healthy environment has little effect on the hematopoietic stem cell epigenome and effect observed in rodents(Singer *et al.*, 2014) are either species specific or are dilute and then lost with the turnover of the hematopoietic progenitor pool. Other organs and tissues are impacted more severely by this change of environment. Plasma proteomic analysis gave us a snapshot of the changes that occur after surgery through the entire body. The majority of the changes were due to proteins whose level decreased after surgery (532 out of 605; **Table S28**). These were associated with inflammatory response, insulin signalling, WNT signalling, VEGF signalling and their decrease was likely due to the reduction in fat mass. The tissue specificity analysis identified amongst the sources of the DAP heart, artery and blood (**Table S29**). The former two showed that vascular integrity, compromised by obesity(Shah *et al.*, 2015), was restored, as also observed by Albrechtsen and colleagues(Wewer Albrechtsen *et al.*, 2018). Almost half (6/13) of the blood specific DAP were also DEG in at least one cell type; SIRPB1 is known to be involved in impaired immune response and leptin resistance(Rendo-Urteaga *et al.*, 2015), RAC2 plays a role in immune pathways(Das, Ma and Sharma, 2015), PPBP (also known as CXCL7) is a chemokine involved in neutrophils recruitment during thrombosis(Brown, Sepuru and Rajarathnam, 2017) while ADGRE2 mediates macrophage differentiation and inflammatory response(Kuan-Yu *et al.*, 2017). This showed that some of

the changes that occurred at transcriptional level are directly involved in the reduction of the proinflammatory environment but also that other levels of regulation are involved.

The comparison of pre and post surgery of gene expression data allowed us also to determine the behaviour of genes involved in trained immunity(Quintin *et al.*, 2012). Post surgery, we observed that many of the changes involved expression of genes that were upregulated in the absence of training(Novakovic *et al.*, 2016), suggesting that bariatric surgery had a positive impact on innate immune cells, as well as, that trained immunity acted downstream of the hematopoietic stem cell pool and its effects were diluted and eventually lost with the renewal of the hematopoietic progenitors pool.

Lastly, we observed that the transcriptional events were accompanied by a decrease in adhesion in platelets and neutrophils. The decrease in adhesion was also accompanied by a reduction in the level of surface markers, such as, CD16 and CD32 in neutrophils and CD61, CD41b and CD36 in platelets.

These changes suggested a decreased capacity to activate and a decreased capacity to interact with other immune cells, such as leukocytes(Spangenberg *et al.*, 1993). Together these findings indicate that many key players in thrombus formation have reduced ability to respond to stimuli after bariatric surgery.

Conclusion

This study provides a comprehensive view of the transcriptional and epigenetic changes associated with CMS in three innate immune cell types (monocytes, neutrophils and macrophages) and platelets. The integration of multiple -omics data layers allowed us to extract the features discriminating between extreme phenotype groups and between these and blood donors. Moreover, it allowed us to construct a predictive model that we used to rank individuals by their probability of belonging to one of these groups. We also identified an extensive rewiring of the transcriptional programs in neutrophils, monocytes, macrophages and platelets after bariatric surgery. Only a small proportion of the rewired features were a return to the values observed in the lean group, suggesting that bariatric surgery does not revert the metabolism to the lean status but rather it defines a new status with low inflammation and low thrombotic propensity.

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). The Human Research Tissue Bank is supported by the NIHR Cambridge Biomedical Research Centre.

Conflict of interest.

Authors have no CoI to declare.

Author contributions

Conceptualization, D.S., J.J.L, L.G., L.L.N., M.V., P.D.W.K and M.F.; Methodology, D.S., A.C., T.H.C., L.L.N., P.D.W.K and M.F.; Software, D.S., A.C., B.E. and P.D.W.K; Validation, D.S., M.P, C.L. and P.D.W.K; Formal analysis, D.S., A.C., T.H.C., B.E., L.L.N., K.D. and P.D.W.K; Investigation, D.S., A.C., T.H.C., P.A.Q., L.L.N., A.P., P.D.W.K and M.F.; Data curation, D.S., A.C., C.L.A., L.L.N. and P.D.W.K; Writing – Original Draft, D.S., A.C., T.H.C., L.L.N., P.D.W.K and M.F.; Writing – Review & Editing, D.S., A.C., M.P, T.H.C., L.S., M.C.S., G.M, S.D., L.L.N., A.P., M.V., P.D.W.K and M.F.; Visualization, D.S., A.C., M.P, P.D.W.K and M.F.; Funding Acquisition, M.F.; Resources, J.J.L, F.B., S.F, H.M, J.B., C.K., A.K., L.L.N., D.B.S., C.L., C.B., K.D., M.A., M.V. and M.F.; Supervision, L.L.N. P.D.W.K. and M.F, Project administration, M.F.

Legends.

Figure 1 - Metabolic signatures in obese individuals and lipodystrophy patients. A. Principal component analysis (PCA) of three groups: obese, green; lipodystrophy, blue; and blood donors (BD), 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 WGCNA consensus analysis and 988 metabolites. 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: * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$ (Fisher's test p value corrected for multi testing). **D.** Heatmap of patients' MEs adjacencies in the consensus MEs network. The heatmap is color-coded by adjacency, yellow indicating high adjacency (positive correlation) and blue low adjacency (negative correlation). **E.** Beeswarm plot using average MEs per cluster presented in **D**.

Figure 2 - Transcriptional and epigenetic signatures in obese and lipodystrophy patients for three innate 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 and C**. Barplot showing the number of features significantly different: H3K27ac distribution (ChIPseq), gene expression (RNAseq) and DNA methylation (RRBS). Each bar is color coded to represent the different cell types as in A. **B** represents results when comparing lean and obese individuals. **C** represents results when comparing lean individuals and lipodystrophy patients. **D**. Functional GO term annotation of up-regulated genes when comparing lean versus obese individuals (top) and lean individuals versus lipodystrophy patients (bottom), colour coded by cell types as above. 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-omic signatures of obesity and their use in prediction. **A.** Presentation of the different layers used for multi-omic integration, the strategy leading to signature identification and schematic view of BD ranking. **B**. Heatmaps showing the mean of the Z-score distribution for each group, for all features selected in each layer. **C.** Plots showing individuals ranked by their predicted probability of belonging to the obese group, using models trained using data from individual layers, as well as a multi-layer predictive model (as indicated by the plot titles). Plots are ordered by decreasing log loss, with smaller values corresponding to better discrimination of individuals in the extreme phenotype group from all other individuals. **D.** Heatmap showing age and sex adjusted association values between (left) 8 prioritised lipid species and risk factors measured in the Fenland and present cohorts; and (right) a negative control set of 5 unselected lipid species and the same risk factors. Black asterisks indicate significant associations after correcting for multiple testing.

Figure 4 - Effect of bariatric surgery on transcriptional profile, epigenetic landscape and cell functions. **A**. Biochemical values distribution across the four studied groups: obese (dark green); lipodystrophy (blue); blood donors (BD) (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 versus 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**. Bar Plot shows number of features significantly different when comparing obese 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 in at least one cell type. **E**. Adhesion percentage of neutrophils measured in the presence of different pro-inflammatory molecules in obese (dark green) and post surgery (light green) individuals. 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 - Related to Figure 1 - WGCNA analysis with BD individuals metabolite values and cluster functional annotation. **A.** Heatmap of BD individuals eigen-metabolites adjacencies in the consensus eigen-metabolites network. Each row and column correspond

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. Colors indicate cohorts. **C**. Functional annotation of cluster C1, C2 and C4. Heatmap shows normalised abundance of metabolites belonging to each functional category.

Supplemental figure 2 - Related to 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 (ChIPseq), gene expression (RNAseq) and DNA methylation (RRBS). Each bar is color coded to represent the different cell types.

Supplemental figure 3 - Related to Figure 3 - Multi-omic signatures of extreme phenotype groups and their use in prediction. A. Plots showing individuals ranked by their predicted probability of belonging to the obese group. As in **Figure 3C**, but for the Methylation (monocytes), RNAseq (monocytes), Metabolites, and ChIPseq (monocytes) data layers. **B.** Multi-omic model trained using lipodystrophy patients often predicts obese individuals to belong to the lipodystrophy group. As in **Figure 3C** (final plot), but training the multi-layer model using the Lipodystrophy and Lean groups (rather than the Obese and Lean groups). Using this model, Obese individuals are often predicted as belonging to the Lipodystrophy group.

Supplemental figure 4 - Related to Figure 3 - A common pattern of associations between the prioritised lipid species and known CMS risk factors. The pattern of association between the prioritised lipids and known CMS risk factors in the NASH cohort (NASH cohort; left) agrees with the results from the present study (BD cohort; right).

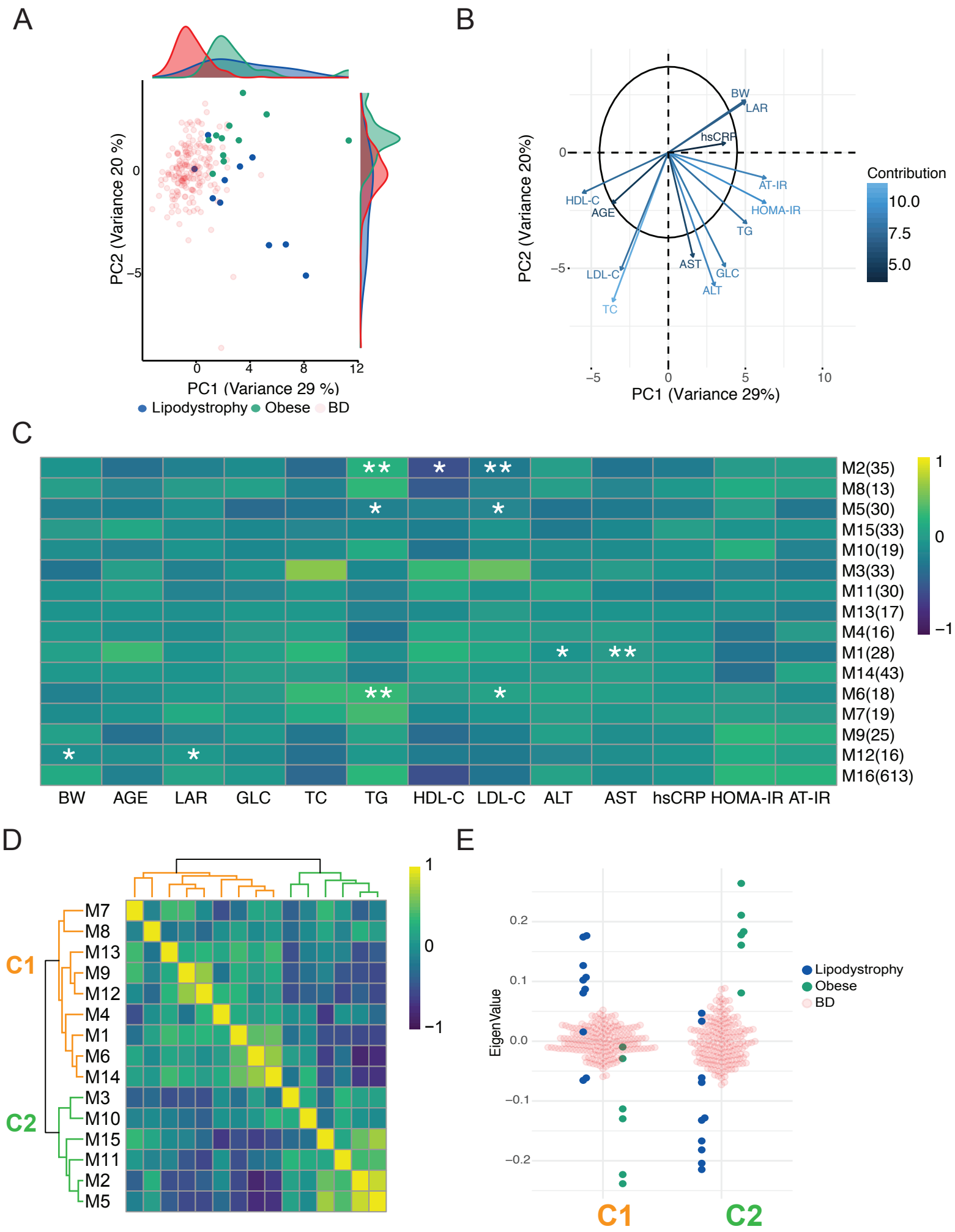


Figure 1 - Metabolic signatures in obese individuals and lipodystrophy patients.

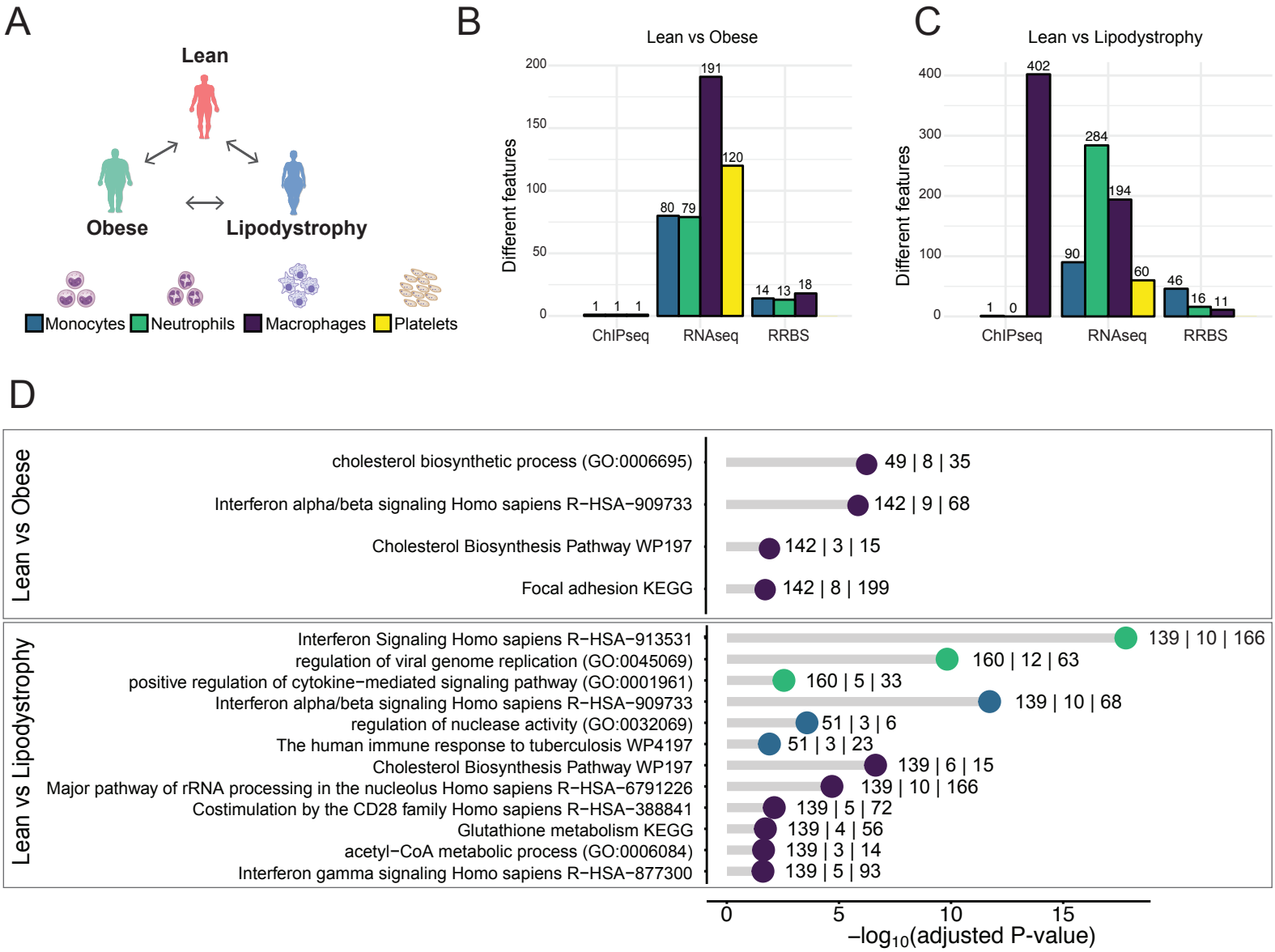


Figure 2 - Transcriptional and epigenetic signatures in obese and lipodystrophy patients for 3 innate immune cell types and platelets.

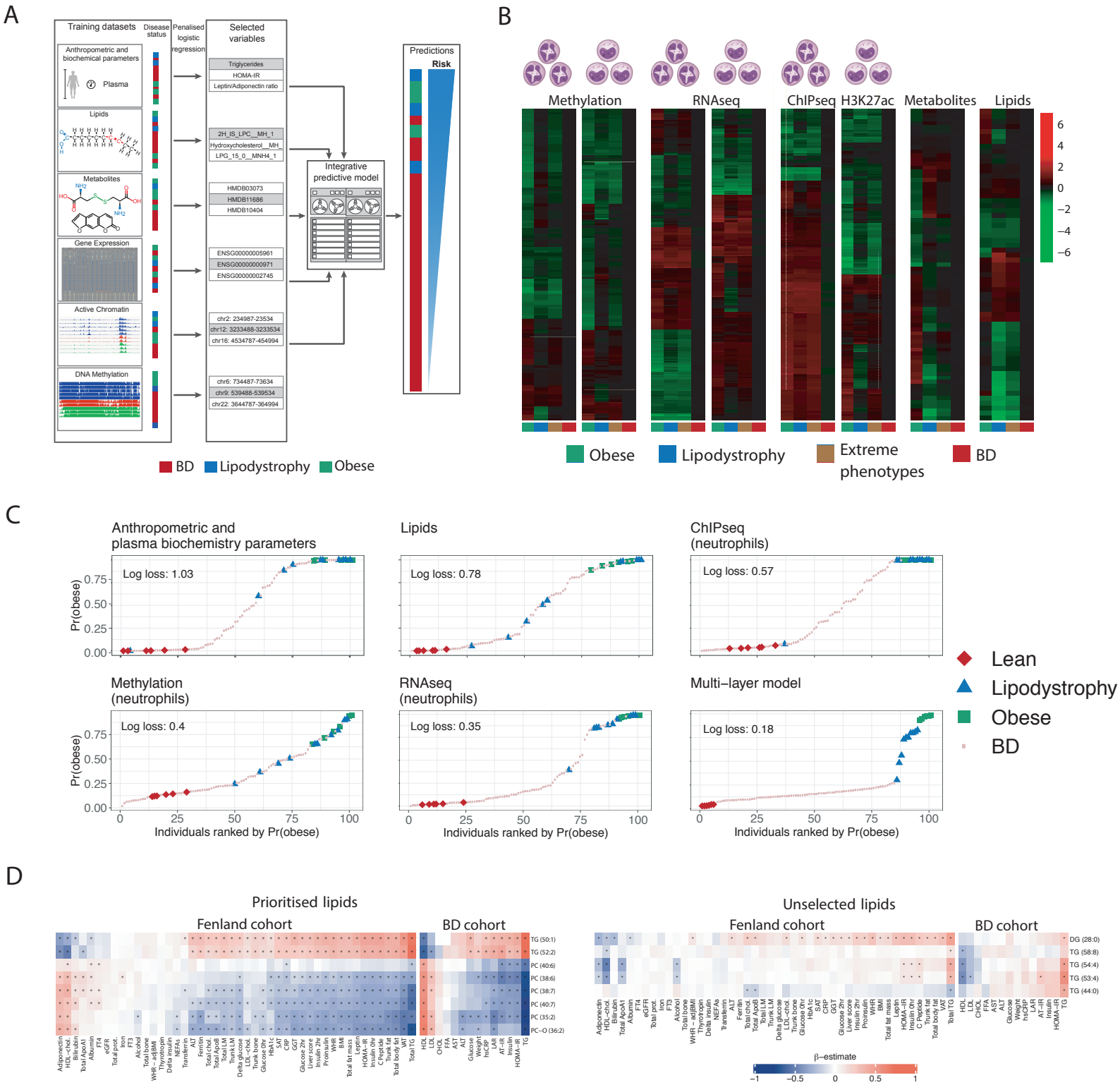


Figure 3 - Multi-omic signatures of obesity and their use in prediction.

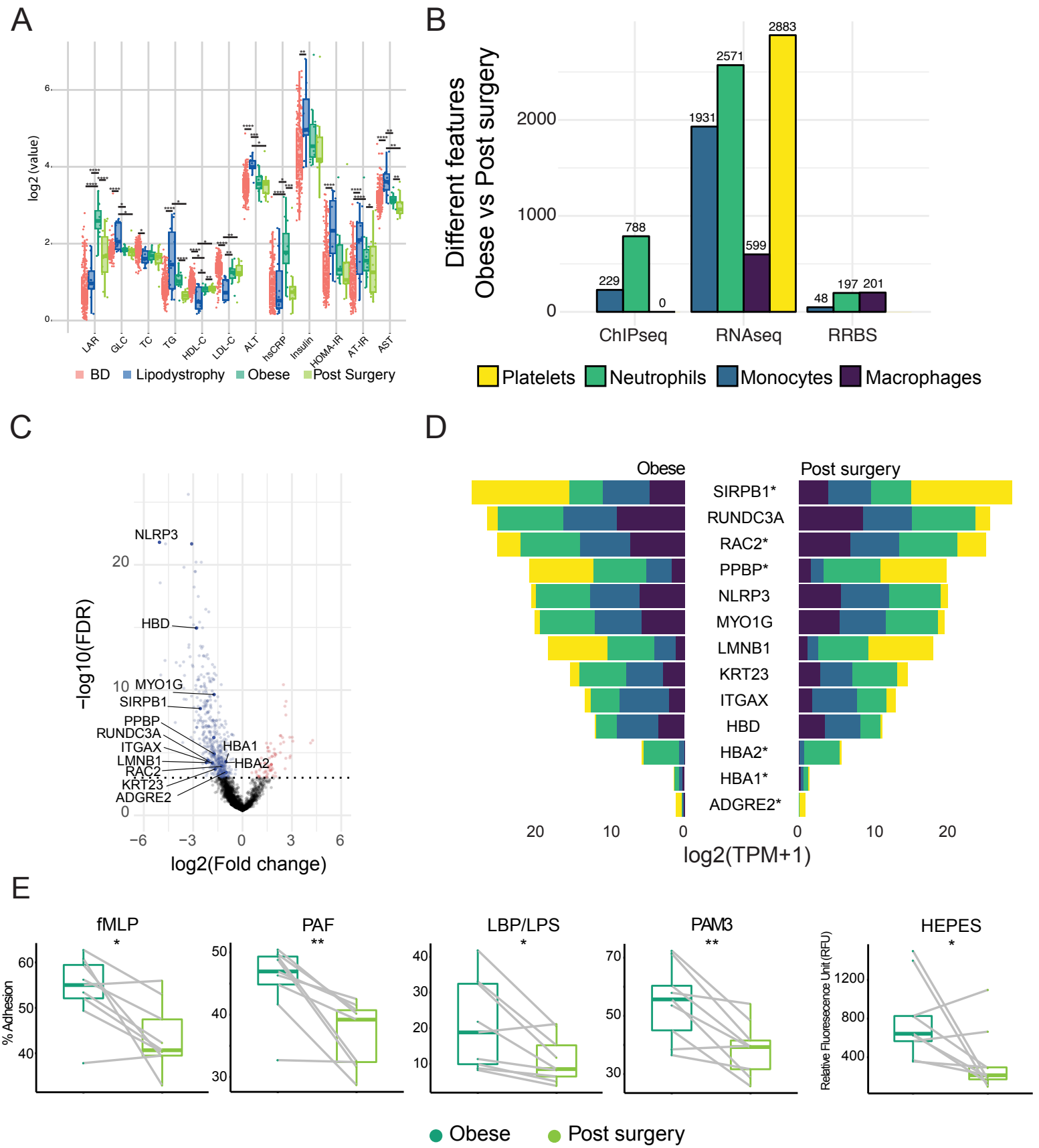
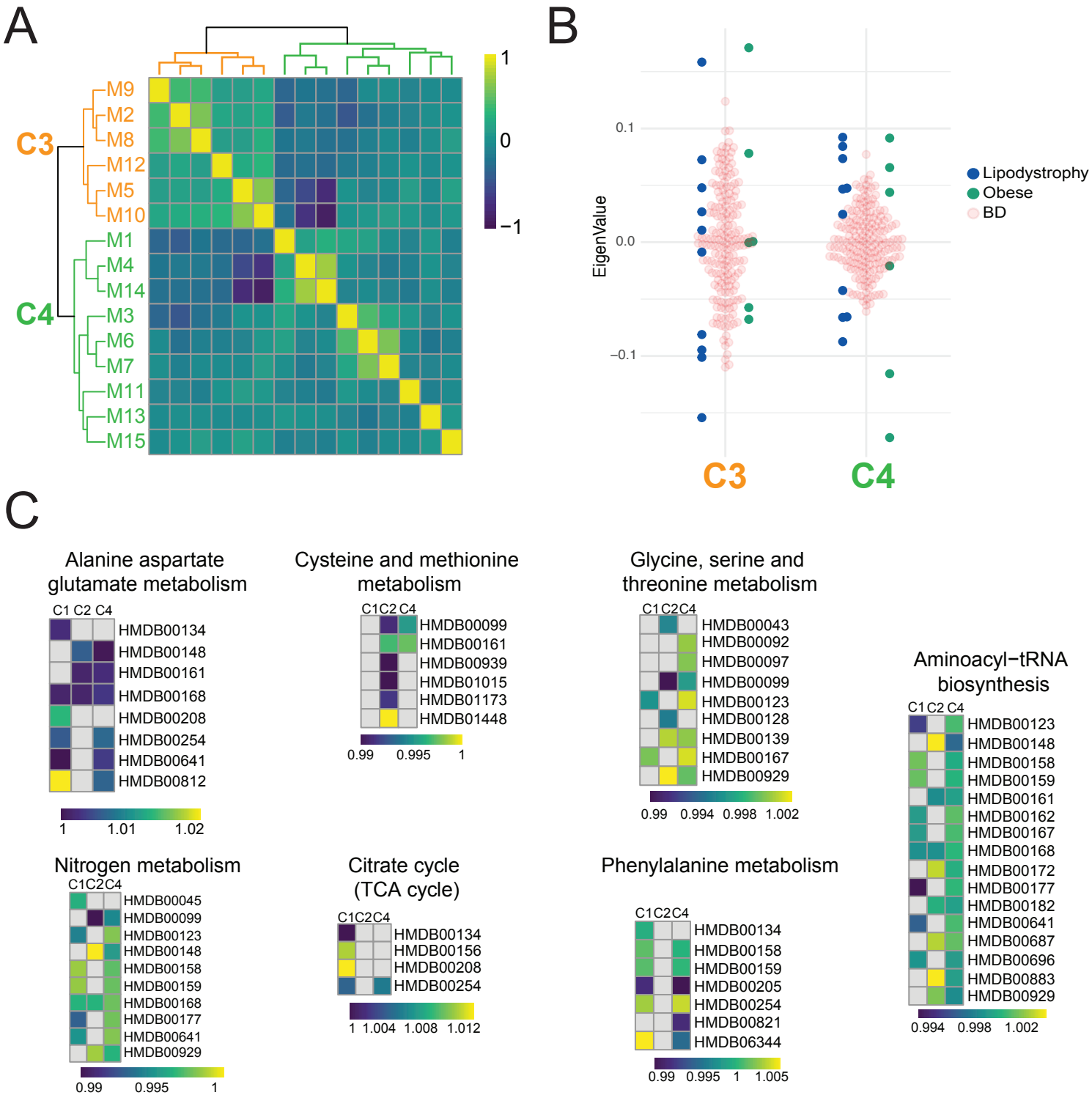
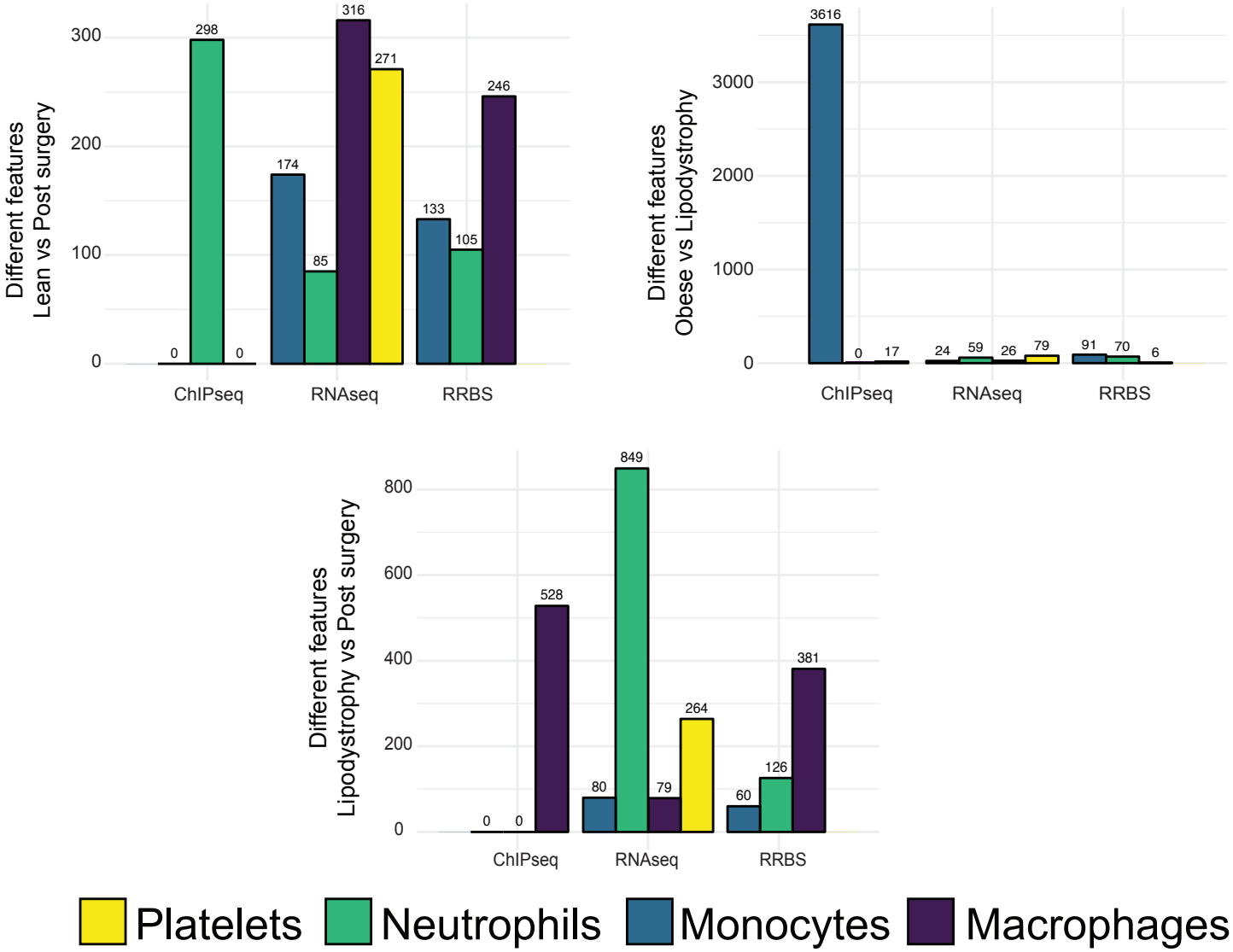


Figure 4 - Effect of bariatric surgery on transcriptional profile, epigenetic landscape and cell functions.

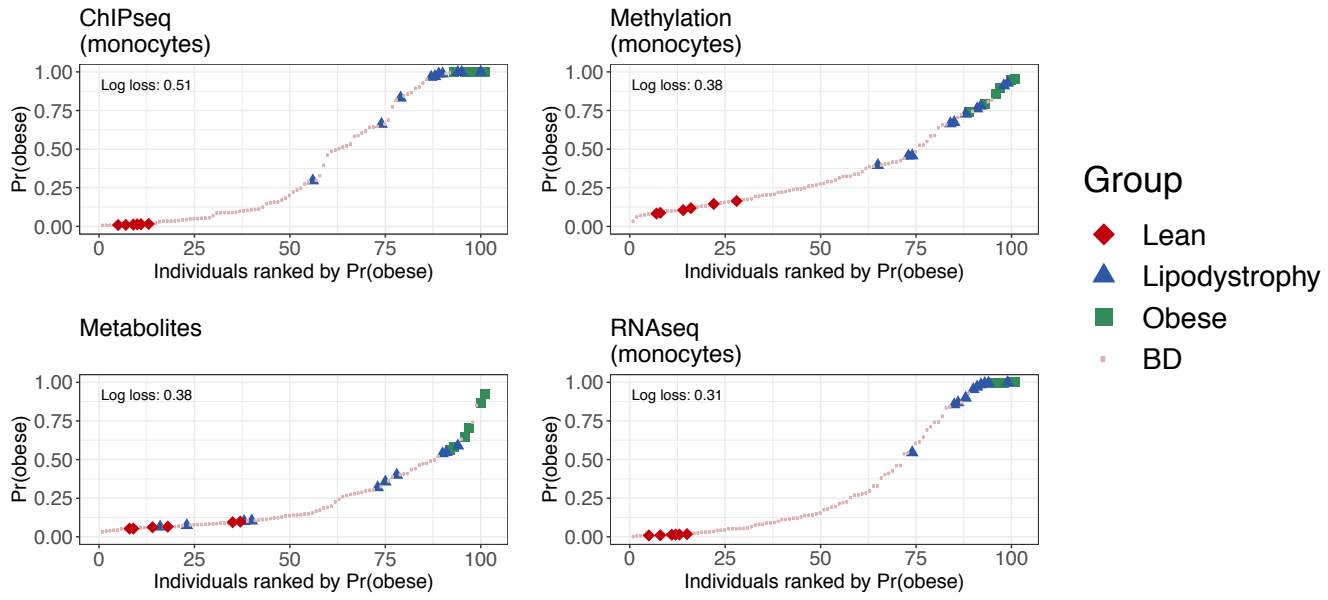


Supplemental figure 1 - Related to Figure 1 - WGCNA analysis with BD individuals metabolite values and cluster functional annotation.

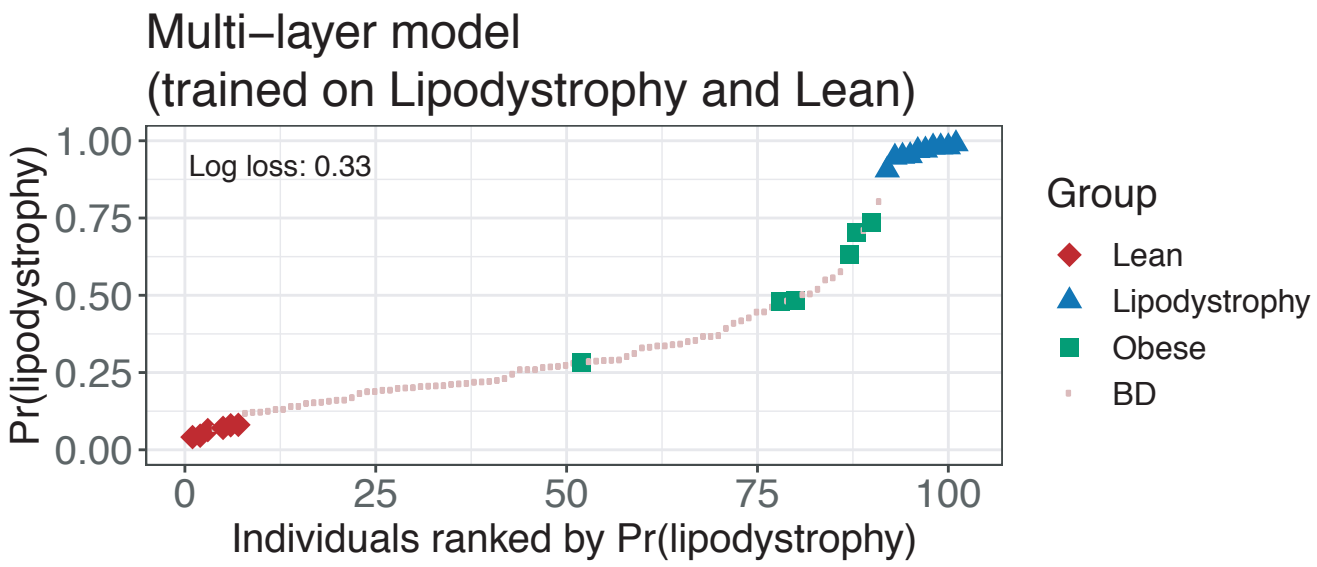


Supplemental figure 2 - Related to Figure 2 - Summary plots of different feature numbers in all comparisons.

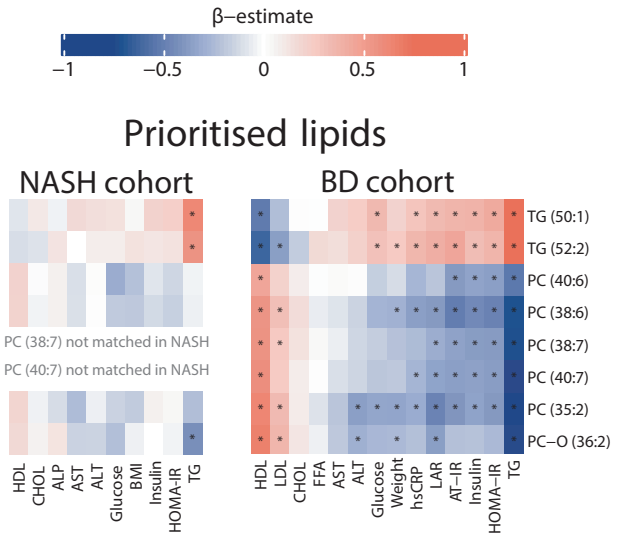
A



B



Supplemental figure 3 - Related to Figure 3 - Multi-omic signatures of extreme phenotype groups and their use in prediction.



Supplemental figure 4 - Related to Figure 3 - A common pattern of associations between the prioritised lipid species and known CMS risk factors.

Material and methods

The datasets generated during this study are available at EGA under study ID EGAS00001003780.
The codes generated during this study are available at GitLab 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 work package 10 (WP10) volunteers (representing the blood donors, "BD", cohort) 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).
"BioNASH" Cohort consisted of 73 consecutive patients recruited at the NASH Service at the Cambridge University Hospital. All the patients had a clinical diagnosis of NAFLD (patients with alternate diagnoses and etiologies were excluded) and histology scored. This study was approved by the local Ethics Committee; all patients gave their informed consent for the use of data (Biochemistry and clinical history) and samples for research purposes. The principles of the Declaration of Helsinki were followed.

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) for platelet isolation. Platelets were 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 red blood cell 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 ChIPseq (formaldehyde fixation, see below). Monocytes were isolated from the peripheral blood mononuclear cell (PBMC) layer by CD14 positive selection (Miltenyi) and aliquots prepared for RNA extraction (TRIzol), DNA extraction for RRBS (snap frozen pellet) and ChIPseq (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 ChIPseq (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 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 (SM$\boldsymbol{\Phi}$, FITC, 9605-02 Southern Biotech), CD45 (HI30, PE-Cy5.5, MHCD4518 Invitrogen). Samples whose purity was below 90% were discarded. BD samples isolation has been extensively described in Chen et al..

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) 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 resuspended in RNase free water. Purified

RNA was stored in single use aliquots. Each sample was quality controlled by a 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 of total RNA for neutrophils, monocytes and macrophages and 200 ng for platelets. o libraries were prepared for sequencing using the Kapa stranded RNAseq kit with riboerase (Roche) according to the manufacturer's instructions and sequenced 150bp paired end on Illumina HiSeq 2500 or Illumina HiSeq 4000. BD RNAseq data (extensively described in Chen et al.(Chen *et al.*, 2016)) 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 trimmed with TrimGalore! (v.0.3.7) [https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/]. Transcript-level abundance was estimated using Kallisto (v0.42)(Bray *et al.*, 2016) with 100 bootstrap iterations in single-end mode for extreme phenotype samples in order to minimize technical batch effect with BD cohort. Transcript abundances were then summarized to gene-level with Tximport R package (v1.9) (Soneson, Love and Robinson, 2015) by using tximport function and Ensembl reference transcriptome (Ensembl Genes 96)(Soneson, Love and Robinson, 2015; Zerbino *et al.*, 2018). This step provides an input count matrix for DESeq2 (v.1.21.21)(Love, Huber and Anders, 2014). 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))(Leek *et al.*, 2012; Love, Huber and Anders, 2014) and individual status as covariate. Non-autosomal genes and those with no or low variance across individuals were removed. The final gene sets (including coding and non-coding genes) were formed of 10,925 genes for monocytes and of 26,634 for neutrophils.

Differential analysis

For differential analysis, transcript-level abundance was 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 and we used age and gender as covariates. 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 the R package EnrichR (v.1.0) (E. Y. Chen *et al.*, 2013; Kuleshov *et al.*, 2016) 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)(Li and Durbin, 2010) *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)(Li *et al.*, 2009) and picard (http://broadinstitute.github.io/picard v.2.0.1).

A combination of quality metrics was used to assess sample 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)(Kharchenko, Tolstorukov and Park, 2008; Landt *et al.*, 2012), area under the curve (AUC), X-intercept and Elbow Point computed with plotFingerPrint function from deepTools suite (v.3.0.2)(Ramírez *et al.*, 2016) 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-10000)	>=(e-10000) and <(e-5000)	>=(e-5000) and <(e-2000)	>=(e-2000) and <e	>=e and <(e+25000)

To build ChIPseq layer for integrative analysis, we defined a master set of peaks and quantified H3K27ac ChIPseq signals under these peaks. Peaks shared by at least 5 individuals were merged using R package DiffBind (v2.9)(Ross-Innes *et al.*, 2012). 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 didn't 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 peaks across individuals were removed. The final master set of peaks counted 25,595 regions in monocytes and 26,300 regions in neutrophils.

Differential analysis

For differential analysis, we used DiffBind with the 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. We used age and gender as covariates. 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)(Heinz *et al.*, 2010), 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 RNAseq, 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

BD 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.(Chen *et al.*, 2016). 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)(Müller *et al.*, 2019). First, we performed background correction and dye-bias normalization using NOOB(Triche *et al.*, 2013), followed by normalization between Infinium probe types with SWAN(Maksimovic, Gordon and Oshlack, 2012). 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(Nordlund *et al.*, 2013); cross-reactive and SNP-overlapping probes(Y.-A. Chen *et al.*, 2013).

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)(Bolger, Lohse and Usadel, 2014) was used for trimming the adapter sequences. Trimmed short read sequences were aligned onto the GRCh38/hg38 human reference genome with BSMAP(v2.90)(Xi and Li, 2009) 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(Kent *et al.*, 2002), 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 844 and 1,127 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)(Du, Kibbe and Lin, 2008; Du *et al.*, 2010), 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. The final data matrix used for multi-omic integration, comprised DNA methylation M-values across 24,311 CpG sites and 210 samples in monocytes and 24,217 CpG sites and 203 samples in neutrophils.

Differential analysis

For differential analysis, we used the methylKit R package (v.1.8.1)(Akalin *et al.*, 2012) 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. We used age and gender as covariates. Q Values are then computed using the SLIM method(Wang, Tuominen and Tsai, 2011; Akalin *et al.*, 2012). 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 RNAseq and ChIPseq, 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) score as follows: (glucose (mg/dL) x insulin (mIU/L)) / 405, and adipose tissue insulin resistance (AT) score 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 (BD 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(O'Brien *et al.*, 2019). 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(Eiden *et al.*, 2015). Raw data were then converted to .mzXML (usingMSconvert(Race, Styles and Bunch, 2012) 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. Te fles were aligned using the XCMS(Smith *et al.*, 2006; Tautenhahn, Böttcher and Neumann, 2008) 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(Sanders *et al.*, 2018).

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. Deamination 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)(Ritchie *et al.*, 2015). 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(Carithers *et al.*, 2015) 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)(Jain and Tuteja, 2019). This package relies on Human Protein Atlas(Uhlén *et al.*, 2015) 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(Zhang and Horvath, 2005) 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. To increase statistical power, we merged the patient groups under the assumption that they share similar associations of metabolites and phenotypic traits. 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) (Langfelder and Horvath, 2008). 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(Langfelder and Horvath, 2007). We performed 3 analysis: extreme phenotypes (obese individuals and lipodystrophy patients were combined to get minimal sample size for network analysis), donors (all BD individuals) and a consensus analysis. We identified 8, 22 and 16 modules with donors, patients and consensus data respectively. Regarding consensus analysis, we considered 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. 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 (HOMA-IR) 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(Chong, Yamamoto and Xia, 2019) and in particular Pathway analysis module by submitting combined list HMDB identifiers for clusters C1 and C2, hyper-geometric 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 BD individuals 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. We used the same method to impute missing clinical values in the NASH cohort.

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. We adjusted for age and sex by including them as unpenalised covariates in the multivariable model. The elastic-net α parameter was fixed at $\alpha = 0.1$, 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 prioritised a reduced set of 9 lipid species that had been selected into the signature. These 9 species satisfied the following criteria: (1) they were selected into the lipidomic signature; and (2) using the Mann-Whitney test with Storey's q-value method to correct for multiple testing, we were able to reject the null hypothesis of no difference in distribution for these lipids in all of the following comparisons: (i) obese vs. control; (ii) lipodystrophy vs. control; and (iii) {obese and lipodystrophy} vs. control. All tests were performed using data from the present study only. Of these 9 species, we were able to match 8 with lipid species that had been quantified in a subset of 1,507 participants of the Fenland study(Sanders *et al.*, 2018; Lindsay *et al.*, 2019) 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. We used linear regression analysis to test for association between plasma levels of the 8 lipid species selected into the lipidomic signature and all relevant CMS parameters quantified in both the reduced Fenland cohort, and the BD cohort, adjusting for age and sex, and using the Bonferroni method to control for multiple testing. To create a negative control set, we identified lipids that satisfied the following criteria: (1) they were not selected into the lipidomic signature; (2) they could be matched with lipid species that had been quantified in the reduced Fenland cohort; and (3) using the Mann-Whitney test with Storey's q-value method to correct for multiple testing, we were unable to reject the null hypothesis of no difference in distribution for these lipids in any of the following comparisons: (i) obese vs. control; (ii) lipodystrophy vs. control; and (iii) {obese and lipodystrophy} vs. control. There were 37 lipid species that satisfied these criteria. We ranked these according to their mean absolute Pearson correlation with the 9 prioritised lipid species, and selected the 5 lowest ranking as our negative control set.

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

Adams, T. D. *et al.* (2017) 'Weight and Metabolic Outcomes 12 Years after Gastric Bypass', *The New England journal of medicine*, 377(12), pp. 1143–1155.

Akalin, A. *et al.* (2012) 'methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles', *Genome biology*, 13(10), p. R87.

Alberti, K. G. M. M. *et al.* (2009) '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(16), pp. 1640–1645.

Alves, A. *et al.* (2019) 'Glycine Metabolism and Its Alterations in Obesity and Metabolic Diseases', *Nutrients*, 11(6). doi: 10.3390/nu11061356.

Aron-Wisnewsky, J. *et al.* (2009) 'Human adipose tissue macrophages: m1 and m2 cell surface markers in subcutaneous and omental depots and after weight loss', *The Journal of clinical endocrinology and metabolism*, 94(11), pp. 4619–4623.

Artigao-Rodenas, L. M. *et al.* (2013) 'Framingham risk score for prediction of cardiovascular diseases: a population-based study from southern Europe', *PloS one*, 8(9), p. e73529.

Azzu, V. *et al.* (2020) 'Adipose tissue-liver cross talk in the control of whole-body metabolism: implications in non-alcoholic fatty liver disease', *Gastroenterology*. doi: 10.1053/j.gastro.2019.12.054.

Bedogni, G. *et al.* (2006) 'The Fatty Liver Index: a simple and accurate predictor of hepatic steatosis in the general population', *BMC gastroenterology*, 6, p. 33.

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

Berg, A. H. and Scherer, P. E. (2005) 'Adipose tissue, inflammation, and cardiovascular disease', *Circulation research*, 96(9), pp. 939–949.

Bobryshev, Y. V. *et al.* (2016) 'Macrophages and Their Role in Atherosclerosis: Pathophysiology and Transcriptome Analysis', *BioMed research international*, 2016, p. 9582430.

Bolger, A. M., Lohse, M. and Usadel, B. (2014) 'Trimmomatic: a flexible trimmer for Illumina sequence data', *Bioinformatics* , 30(15), pp. 2114–2120.

Bray, N. L. *et al.* (2016) 'Near-optimal probabilistic RNAseq quantification', *Nature biotechnology*,

34(5), pp. 525–527.

Brown, A. J., Sepuru, K. M. and Rajarathnam, K. (2017) 'Structural Basis of Native CXCL7 Monomer Binding to CXCR2 Receptor N-Domain and Glycosaminoglycan Heparin', *International journal of molecular sciences*, 18(3). doi: 10.3390/ijms18030508.

Busetto, L. *et al.* (2017) 'Practical Recommendations of the Obesity Management Task Force of the European Association for the Study of Obesity for the Post-Bariatric Surgery Medical Management', *Obesity facts*, 10(6), pp. 597–632.

Caielli, S., Banchereau, J. and Pascual, V. (2012) 'Neutrophils come of age in chronic inflammation', *Current opinion in immunology*, 24(6), pp. 671–677.

Carithers, L. J. *et al.* (2015) 'A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project', *Biopreservation and biobanking*, 13(5), pp. 311–319.

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

Chen, L. *et al.* (2016) 'Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells', *Cell*, 167(5), pp. 1398–1414.e24.

Chen, Y.-A. *et al.* (2013) 'Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray', *Epigenetics: official journal of the DNA Methylation Society*, 8(2), pp. 203–209.

Choe, S. S. *et al.* (2016) 'Adipose Tissue Remodeling: Its Role in Energy Metabolism and Metabolic Disorders', *Frontiers in endocrinology*, 7, p. 30.

Chong, J., Yamamoto, M. and Xia, J. (2019) 'MetaboAnalystR 2.0: From Raw Spectra to Biological Insights', *Metabolites*, 9(3). doi: 10.3390/metabo9030057.

Cirulli, E. T. *et al.* (2019) 'Profound Perturbation of the Metabolome in Obesity Is Associated with Health Risk', *Cell metabolism*, 29(2), pp. 488–500.e2.

Ciudin, A. *et al.* (2011) 'Successful treatment for the Dunnigan-type familial partial lipodystrophy with Roux-en-Y gastric bypass', *Clinical endocrinology*, 75(3), pp. 403–404.

Dai, J. *et al.* (2019) 'Serum 3-carboxy-4-methyl-5-propyl-2-furanpropanoic acid is associated with lipid profiles and might protect against non-alcoholic fatty liver disease in Chinese individuals', *Journal of diabetes investigation*, 10(3), pp. 793–800.

Das, S. K., Ma, L. and Sharma, N. K. (2015) 'Adipose tissue gene expression and metabolic health of obese adults', *International journal of obesity* , 39(5), pp. 869–873.

Drucker, D. J. and Nauck, M. A. (2006) 'The incretin system: glucagon-like peptide-1 receptor agonists and dipeptidyl peptidase-4 inhibitors in type 2 diabetes', *The Lancet*, 368(9548), pp. 1696–1705.

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

Du, P., Kibbe, W. A. and Lin, S. M. (2008) 'lumi: a pipeline for processing Illumina microarray', *Bioinformatics* , 24(13), pp. 1547–1548.

Eiden, M. *et al.* (2015) 'Mechanistic insights revealed by lipid profiling in monogenic insulin resistance

syndromes', *Genome medicine*, 7, p. 63.

Elagizi, A. *et al.* (2018) 'An Overview and Update on Obesity and the Obesity Paradox in Cardiovascular Diseases', *Progress in cardiovascular diseases*, 61(2), pp. 142–150.

Fiorenza, C. G., Chou, S. H. and Mantzoros, C. S. (2011) 'Lipodystrophy: pathophysiology and advances in treatment', *Nature reviews. Endocrinology*, 7(3), pp. 137–150.

Ford, E. S. (2005) 'Risks for all-cause mortality, cardiovascular disease, and diabetes associated with the metabolic syndrome: a summary of the evidence', *Diabetes care*, 28(7), pp. 1769–1778.

Ghosh, S. *et al.* (2010) 'Gene expression profiling in whole blood identifies distinct biological pathways associated with obesity', *BMC medical genomics*, 3, p. 56.

Go, A. S. *et al.* (2014) 'Heart disease and stroke statistics--2014 update: a report from the American Heart Association', *Circulation*, 129(3), pp. e28–e292.

Gros, A., Ollivier, V. and Ho-Tin-Noé, B. (2014) 'Platelets in inflammation: regulation of leukocyte activities and vascular repair', *Frontiers in immunology*, 5, p. 678.

Grundfest-Broniatowski, S. *et al.* (2017) 'Successful Treatment of an Unusual Case of FPLD2: The Role of Roux-en-Y Gastric Bypass-Case Report and Literature Review', *Journal of gastrointestinal surgery: official journal of the Society for Surgery of the Alimentary Tract*, 21(4), pp. 739–743.

Grundy, S. M. *et al.* (2005) 'Diagnosis and management of the metabolic syndrome: an American Heart Association/National Heart, Lung, and Blood Institute Scientific Statement', *Circulation*, 112(17), pp. 2735–2752.

GTEx Consortium (2015) 'Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans', *Science*, 348(6235), pp. 648–660.

Hall, Z. *et al.* (2017) 'Lipid zonation and phospholipid remodeling in nonalcoholic fatty liver disease', *Hepatology* , 65(4), pp. 1165–1180.

Heinz, S. *et al.* (2010) 'Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities', *Molecular cell*, 38(4), pp. 576–589.

Hippisley-Cox, J. *et al.* (2008) 'Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2', *BMJ* , 336(7659), pp. 1475–1482.

Horie, T. *et al.* (2013) 'MicroRNA-33 regulates sterol regulatory element-binding protein 1 expression in mice', *Nature communications*, 4, p. 2883.

Hotamisligil, G. S. (2006) 'Inflammation and metabolic disorders', *Nature*, 444(7121), pp. 860–867.

Hotamisligil, G. S., Shargill, N. S. and Spiegelman, B. M. (1993) 'Adipose expression of tumor necrosis factor-alpha: direct role in obesity-linked insulin resistance', *Science*, 259(5091), pp. 87–91.

Huang-Doran, I. *et al.* (2010) 'Lipodystrophy: metabolic insights from a rare disorder', *The Journal of endocrinology*, 207(3), pp. 245–255.

Hyötyläinen, T. *et al.* (2016) 'Genome-scale study reveals reduced metabolic adaptability in patients with non-alcoholic fatty liver disease', *Nature communications*, 7, p. 8994.

Innes, J. K. and Calder, P. C. (2018) 'The Differential Effects of Eicosapentaenoic Acid and

Docosahexaenoic Acid on Cardiometabolic Risk Factors: A Systematic Review', *International journal of molecular sciences*, 19(2). doi: 10.3390/ijms19020532.

Jain, A. and Tuteja, G. (2019) 'TissueEnrich: Tissue-specific gene enrichment analysis', *Bioinformatics* , 35(11), pp. 1966–1967.

Jain, S. K. *et al.* (2014) 'Vitamin D and L-cysteine levels correlate positively with GSH and negatively with insulin resistance levels in the blood of type 2 diabetic patients', *European journal of clinical nutrition*, 68(10), pp. 1148–1153.

Karlsson, J. *et al.* (2007) 'Ten-year trends in health-related quality of life after surgical and conventional treatment for severe obesity: the SOS intervention study', *International journal of obesity* , 31(8), pp. 1248–1261.

Katzmarzyk, P. T. *et al.* (2006) 'The importance of waist circumference in the definition of metabolic syndrome: prospective analyses of mortality in men', *Diabetes care*, 29(2), pp. 404–409.

Kelli, H. M. and Kassas, I. (2016) 'Cardio Metabolic Syndrome: A Global Epidemic', *Journal of diabetes & metabolism*, 6(3). doi: 10.4172/2155-6156.1000513.

Kent, W. J. *et al.* (2002) 'The human genome browser at UCSC', *Genome research*, 12(6), pp. 996–1006.

Kharchenko, P. V., Tolstorukov, M. Y. and Park, P. J. (2008) 'Design and analysis of ChIPseq experiments for DNA-binding proteins', *Nature biotechnology*, 26(12), pp. 1351–1359.

Kieffer-Kwon, K.-R. *et al.* (2013) 'Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation', *Cell*, 155(7), pp. 1507–1520.

Kip, K. E. *et al.* (2004) '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(6), pp. 706–713.

Kosmas, C. E. *et al.* (2019) 'Anti-inflammatory therapy for cardiovascular disease', *Annals of translational medicine*, 7(7), p. 147.

Koupenova, M. *et al.* (2018) 'Circulating Platelets as Mediators of Immunity, Inflammation, and Thrombosis', *Circulation research*, 122(2), pp. 337–351.

Kuan-Yu, I. *et al.* (2017) 'Activation of Adhesion GPCR EMR2/ADGRE2 Induces Macrophage Differentiation and Inflammatory Responses via Gα16/Akt/MAPK/NF-κB Signaling Pathways', *Frontiers in immunology*, 8, p. 463.

Kuleshov, M. V. *et al.* (2016) 'Enrichr: a comprehensive gene set enrichment analysis web server 2016 update', *Nucleic acids research*, 44(W1), pp. W90–7.

Kvaløy, K., Page, C. M. and Holmen, T. L. (2018) 'Epigenome-wide methylation differences in a group of lean and obese women - A HUNT Study', *Scientific reports*, 8(1), p. 16330.

Lancha, A., Frühbeck, G. and Gómez-Ambrosi, J. (2012) 'Peripheral signalling involved in energy homeostasis control', *Nutrition research reviews*, 25(2), pp. 223–248.

Landt, S. G. *et al.* (2012) 'ChIPseq guidelines and practices of the ENCODE and modENCODE consortia', *Genome research*, 22(9), pp. 1813–1831.

Langfelder, P. and Horvath, S. (2007) 'Eigengene networks for studying the relationships between co-

expression modules', *BMC systems biology*, 1, p. 54.

Langfelder, P. and Horvath, S. (2008) 'WGCNA: an R package for weighted correlation network analysis', *BMC bioinformatics*, 9, p. 559.

Lawler, H. M. *et al.* (2016) 'Adipose Tissue Hypoxia, Inflammation, and Fibrosis in Obese Insulin-Sensitive and Obese Insulin-Resistant Subjects', *The Journal of clinical endocrinology and metabolism*, 101(4), pp. 1422–1428.

Leal, J. *et al.* (2006) 'Economic burden of cardiovascular diseases in the enlarged European Union', *European heart journal*, 27(13), pp. 1610–1619.

Leek, J. T. *et al.* (2012) 'The sva package for removing batch effects and other unwanted variation in high-throughput experiments', *Bioinformatics* , 28(6), pp. 882–883.

Leentjens, J. *et al.* (2018) 'Trained Innate Immunity as a Novel Mechanism Linking Infection and the Development of Atherosclerosis', *Circulation research*, 122(5), pp. 664–669.

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

Li, H. *et al.* (2009) 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics* , 25(16), pp. 2078–2079.

Li, H. and Durbin, R. (2010) 'Fast and accurate long-read alignment with Burrows-Wheeler transform', *Bioinformatics* , 26(5), pp. 589–595.

Lindsay, T. *et al.* (2019) 'Descriptive epidemiology of physical activity energy expenditure in UK adults (The Fenland study)', *The international journal of behavioral nutrition and physical activity*, 16(1), p. 126.

Ling, C. and Rönn, T. (2019) 'Epigenetics in Human Obesity and Type 2 Diabetes', *Cell metabolism*, 29(5), pp. 1028–1044.

Liu, J. *et al.* (2018) 'Data integration by multi-tuning parameter elastic net regression', *BMC bioinformatics*, 19(1), p. 369.

Love, M. I., Huber, W. and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNAseq data with DESeq2', *Genome biology*, 15(12), p. 550.

Maggard, M. A. *et al.* (2005) 'Meta-analysis: surgical treatment of obesity', *Annals of internal medicine*, 142(7), pp. 547–559.

Majithia, A. and Bhatt, D. L. (2019) 'Novel Antiplatelet Therapies for Atherothrombotic Diseases', *Arteriosclerosis, thrombosis, and vascular biology*, 39(4), pp. 546–557.

Maksimovic, J., Gordon, L. and Oshlack, A. (2012) 'SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips', *Genome biology*, 13(6), p. R44.

Matthews, D. R. *et al.* (1985) 'Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man', *Diabetologia*, 28(7), pp. 412–419.

McGrath, N. M. and Krishna, G. (2006) 'Gastric bypass for insulin resistance due to lipodystrophy', *Obesity surgery*, 16(11), pp. 1542–1544.

McNeill, A. M. *et al.* (2005) 'The metabolic syndrome and 11-year risk of incident cardiovascular disease in the atherosclerosis risk in communities study', *Diabetes care*, 28(2), pp. 385–390.

Melvin, A. *et al.* (2017) 'Roux-en-Y Gastric Bypass Surgery in the Management of Familial Partial Lipodystrophy Type 1', *The Journal of clinical endocrinology and metabolism*, 102(10), pp. 3616–3620.

Moore, S. C. *et al.* (2014) 'Human metabolic correlates of body mass index', *Metabolomics: Official journal of the Metabolomic Society*, 10(2), pp. 259–269.

Müller, F. *et al.* (2019) 'RnBeads 2.0: comprehensive analysis of DNA methylation data', *Genome biology*, 20(1), p. 55.

Murphy, K. P. (2012) *Machine Learning: A Probabilistic Perspective*. MIT Press.

Nichols, G. A. *et al.* (2017) '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', *Preventing chronic disease*, 14, p. E22.

Ninomiya, J. K. *et al.* (2004) 'Association of the metabolic syndrome with history of myocardial infarction and stroke in the Third National Health and Nutrition Examination Survey', *Circulation*, 109(1), pp. 42–46.

Nording, H. M., Seizer, P. and Langer, H. F. (2015) 'Platelets in inflammation and atherogenesis', *Frontiers in immunology*, 6, p. 98.

Nordlund, J. *et al.* (2013) 'Genome-wide signatures of differential DNA methylation in pediatric acute lymphoblastic leukemia', *Genome biology*, 14(9), p. r105.

Novakovic, B. *et al.* (2016) 'β-Glucan Reverses the Epigenetic State of LPS-Induced Immunological Tolerance', *Cell*, 167(5), pp. 1354–1368.e14.

O'Brien, K. A. *et al.* (2019) 'Metabolomic and lipidomic plasma profile changes in human participants ascending to Everest Base Camp', *Scientific reports*, 9(1), p. 2297.

Onesi, S. O. and Ignatius, U. E. (2014) 'Metabolic syndrome: Performance of five different diagnostic criterias', *Indian journal of endocrinology and metabolism*, 18(4), pp. 496–501.

Pacana, T. *et al.* (2015) 'Dysregulated Hepatic Methionine Metabolism Drives Homocysteine Elevation in Diet-Induced Nonalcoholic Fatty Liver Disease', *PloS one*, 10(8), p. e0136822.

Pahan, K. (2006) 'Lipid-lowering drugs', *Cellular and molecular life sciences: CMLS*, 63(10), pp. 1165–1178.

Park, H. S., Park, J. Y. and Yu, R. (2005) 'Relationship of obesity and visceral adiposity with serum concentrations of CRP, TNF-alpha and IL-6', *Diabetes research and clinical practice*, 69(1), pp. 29–35.

Puhr-Westerheide, D. *et al.* (2019) 'Neutrophils promote venular thrombosis by shaping the rheological environment for platelet aggregation', *Scientific reports*, 9(1), p. 15932.

Quintin, J. *et al.* (2012) 'Candida albicans infection affords protection against reinfection via functional reprogramming of monocytes', *Cell host & microbe*, 12(2), pp. 223–232.

Race, A. M., Styles, I. B. and Bunch, J. (2012) 'Inclusive sharing of mass spectrometry imaging data requires a converter for all', *Journal of proteomics*, 75(16), pp. 5111–5112.

Ramírez, F. *et al.* (2016) 'deepTools2: a next generation web server for deep-sequencing data analysis', *Nucleic acids research*, 44(W1), pp. W160–5.

Ramirez, G. A., Manfredi, A. A. and Maugeri, N. (2019) 'Misunderstandings Between Platelets and Neutrophils Build in Chronic Inflammation', *Frontiers in immunology*, 10, p. 2491.

Rath, R. and Masek, J. (1966) 'Changes in the nitrogen metabolism in obese women after fasting and refeeding', *Metabolism: clinical and experimental*, 15(1), pp. 1–8.

Rendo-Urteaga, T. *et al.* (2015) 'Peripheral blood mononuclear cell gene expression profile in obese boys who followed a moderate energy-restricted diet: differences between high and low responders at baseline and after the intervention', *The British journal of nutrition*, 113(2), pp. 331–342.

Ritchie, M. E. *et al.* (2015) 'limma powers differential expression analyses for RNAsequencing and microarray studies', *Nucleic acids research*, 43(7), p. e47.

Ross-Innes, C. S. *et al.* (2012) 'Differential oestrogen receptor binding is associated with clinical outcome in breast cancer', *Nature*, 481(7381), pp. 389–393.

Sanchez-Rangel, E. and Inzucchi, S. E. (2017) 'Metformin: clinical use in type 2 diabetes', *Diabetologia*, 60(9), pp. 1586–1593.

Sanders, F. W. B. *et al.* (2018) 'Hepatic steatosis risk is partly driven by increased de novo lipogenesis following carbohydrate consumption', *Genome biology*, 19(1), p. 79.

Shah, D. *et al.* (2015) 'Obesity-induced adipokine imbalance impairs mouse pulmonary vascular endothelial function and primes the lung for injury', *Scientific reports*, 5, p. 11362.

Shimada, Y. J. *et al.* (2019) 'Bariatric surgery is associated with lower risk of acute care use for cardiovascular disease in obese adults', *Cardiovascular research*, 115(4), pp. 800–806.

Singer, K. *et al.* (2014) 'Diet-induced obesity promotes myelopoiesis in hematopoietic stem cells', *Molecular metabolism*, 3(6), pp. 664–675.

Smith, C. A. *et al.* (2006) 'XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification', *Analytical chemistry*, 78(3), pp. 779–787.

Søndergaard, E. *et al.* (2017) 'How to Measure Adipose Tissue Insulin Sensitivity', *The Journal of clinical endocrinology and metabolism*, 102(4), pp. 1193–1199.

Soneson, C., Love, M. I. and Robinson, M. D. (2015) 'Differential analyses for RNAseq: transcript-level estimates improve gene-level inferences', *F1000Research*, 4, p. 1521.

Sookoian, S. and Pirola, C. J. (2012) 'Alanine and aspartate aminotransferase and glutamine-cycling pathway: their roles in pathogenesis of metabolic syndrome', *World journal of gastroenterology: WJG*, 18(29), pp. 3775–3781.

Spangenberg, P. *et al.* (1993) 'The platelet glycoprotein IIb/IIIa complex is involved in the adhesion of activated platelets to leukocytes', *Thrombosis and haemostasis*, 70(3), pp. 514–521.

Stefan, N., Schick, F. and Häring, H.-U. (2017) 'Causes, Characteristics, and Consequences of Metabolically Unhealthy Normal Weight in Humans', *Cell metabolism*, 26(2), pp. 292–300.

Sterling, R. K. *et al.* (2006) 'Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection', *Hepatology* , 43(6), pp. 1317–1325.

Stienstra, R. and Stefan, N. (2013) 'Tipping the inflammatory balance: inflammasome activation distinguishes metabolically unhealthy from healthy obesity', *Diabetologia*, 56(11), pp. 2343–2346.

St-Pierre, A. C. *et al.* (2005) 'Insulin resistance syndrome, body mass index and the risk of ischemic heart disease', *CMAJ: Canadian Medical Association journal = journal de l'Association medicale canadienne*, 172(10), pp. 1301–1305.

Swystun, L. L. and Liaw, P. C. (2016) 'The role of leukocytes in thrombosis', *Blood*, 128(6), pp. 753–762.

Tautenhahn, R., Böttcher, C. and Neumann, S. (2008) 'Highly sensitive feature detection for high resolution LC/MS', *BMC bioinformatics*, 9, p. 504.

Triche, T. J., Jr *et al.* (2013) 'Low-level processing of Illumina Infinium DNA Methylation BeadArrays', *Nucleic acids research*, 41(7), p. e90.

van Tuijl, J. *et al.* (2019) 'Immunometabolism orchestrates training of innate immunity in atherosclerosis', *Cardiovascular research*, 115(9), pp. 1416–1424.

Turro, E. *et al.* (2020) 'Whole-genome sequencing of patients with rare diseases in a national health system', *Nature*, 583(7814), pp. 96–102.

Uhlén, M. *et al.* (2015) 'Proteomics. Tissue-based map of the human proteome', *Science*, 347(6220), p. 1260419.

Utzschneider, K. M. and Trence, D. L. (2006) 'Effectiveness of gastric bypass surgery in a patient with familial partial lipodystrophy', *Diabetes care*, 29(6), pp. 1380–1382.

Virtue, S. and Vidal-Puig, A. (2010) 'Adipose tissue expandability, lipotoxicity and the Metabolic Syndrome--an allostatic perspective', *Biochimica et biophysica acta*, 1801(3), pp. 338–349.

Vishvanath, L. and Gupta, R. K. (2019) 'Contribution of adipogenesis to healthy adipose tissue expansion in obesity', *The Journal of clinical investigation*, 129(10), pp. 4022–4031.

Visser, M. *et al.* (1999) 'Elevated C-reactive protein levels in overweight and obese adults', *JAMA: the journal of the American Medical Association*, 282(22), pp. 2131–2135.

Waldmann, P. *et al.* (2013) 'Evaluation of the lasso and the elastic net in genome-wide association studies', *Frontiers in genetics*, 4, p. 270.

Wang, H.-Q., Tuominen, L. K. and Tsai, C.-J. (2011) 'SLIM: a sliding linear model for estimating the proportion of true null hypotheses in datasets with dependence structures', *Bioinformatics* , 27(2), pp. 225–231.

Weisberg, S. P. *et al.* (2003) 'Obesity is associated with macrophage accumulation in adipose tissue', *The Journal of clinical investigation*, 112(12), pp. 1796–1808.

Wewer Albrechtsen, N. J. *et al.* (2018) 'Plasma Proteome Profiling Reveals Dynamics of Inflammatory and Lipid Homeostasis Markers after Roux-En-Y Gastric Bypass Surgery', *Cell systems*, 7(6), pp. 601–612.e3.

Wright, H. L. *et al.* (2010) 'Neutrophil function in inflammation and inflammatory diseases', *Rheumatology* , 49(9), pp. 1618–1631.

Wu, C. *et al.* (2019) 'A Selective Review of Multi-Level Omics Data Integration Using Variable Selection', *High-throughput*, 8(1). doi: 10.3390/ht8010004.

Xi, Y. and Li, W. (2009) 'BSMAP: whole genome bisulfite sequence MAPping program', *BMC bioinformatics*, 10, p. 232.

Zerbino, D. R. *et al.* (2018) 'Ensembl 2018', *Nucleic acids research*, 46(D1), pp. D754–D761.

Zhang, B. and Horvath, S. (2005) 'A general framework for weighted gene co-expression network analysis', *Statistical applications in genetics and molecular biology*, 4, p. Article17.

Zou, H. and Hastie, T. (2005) 'Regularization and variable selection via the elastic net', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, pp. 301–320. doi: 10.1111/j.1467-9868.2005.00503.x.