

UCLA

UCLA Electronic Theses and Dissertations

Title

Cross-tissue omics analysis discovers coiled-coil domain containing 80 and superoxide dismutase 3 as novel serum biomarker candidates for non-alcoholic fatty liver disease

Permalink

<https://escholarship.org/uc/item/2ns6n6k2>

Author

Darci-Maher, Nicholas Waxter

Publication Date

2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Cross-tissue omics analysis discovers coiled-coil domain containing 80 and superoxide dismutase 3 as novel serum biomarker candidates for non-alcoholic fatty liver disease

A thesis submitted in partial satisfaction
of the requirements for the degree
Master of Science in Bioinformatics

by

Nicholas Waxter Darci-Maher

2022

© Copyright by
Nicholas Waxter Darci-Maher
2022

ABSTRACT OF THE THESIS

Cross-tissue omics analysis discovers coiled-coil domain containing 80 and superoxide dismutase 3 as novel serum biomarker candidates for non-alcoholic fatty liver disease

by

Nicholas Waxter Darci-Maher

Master of Science in Bioinformatics

University of California, Los Angeles, 2022

Professor Paivi E. Pajukanta, Chair

Non-alcoholic fatty liver disease (NAFLD) is a fast growing epidemic, which remains grossly underdiagnosed due to the lack of affordable and practical diagnostic tools in the primary health care setting. Here, we utilize dual-tissue RNA-seq data in subcutaneous adipose tissue and liver paired with liver histology-based NAFLD diagnosis from a cohort of obese individuals to discover serum biomarker candidates (SBCs) for obesity-related NAFLD. We hypothesize that in some obese individuals, obesity and its accompanying low-grade inflammation compromise the key functions of subcutaneous adipose tissue, preventing efficient adipogenesis and storage of fat into the subcutaneous fat depot, and thus driving ectopic fat accumulation into the liver. To identify subcutaneous adipose tissue-origin SBCs for the three primary NAFLD histology traits, steatosis, fibrosis, and non-alcoholic steatohepatitis (NASH), we scan for genes that are transcriptome-wide significantly differentially expressed (DE) for these NAFLD traits in subcutaneous adipose tissue but not in the liver, encode proteins secreted to serum, and show preferential expression in subcutaneous adipose tissue over the liver. Using a best subsets analysis, we identify the secreted adipokines CCDC80,

SOD3, and COL6A2 as the key SBCs, the adipose expression of which explains the most significant amount of variance in steatosis, fibrosis, and NASH among the SBCs. We also show, by knockdown in human preadipocytes during adipogenesis, that the fibrosis and NASH SBCs, CCDC80 and SOD3, modulate the crucial adipogenesis genes, SREBF1 and LEP, emphasizing their premise as indicators for adipose-tissue origin, obesity-related NAFLD. Our results have a great translational potential to improve the diagnosis of obesity-related NAFLD by providing a blood panel of SBCs.

The thesis of Nicholas Waxter Darci-Maher is approved.

Joseph R. Pisegna

Sriram Sankararaman

Paivi E. Pajukanta, Committee Chair

University of California, Los Angeles

2022

To Theo and Toby, for making my footsteps feel so important.
To Tom and Beth, for being the giants whose shoulders I stand on.

TABLE OF CONTENTS

1	Introduction	1
2	Results	4
2.1	Study design	4
2.2	Co-expression networks of distinct functional pathways correlate across an individual's subcutaneous adipose tissue and liver	6
2.3	645 genes are DE in subcutaneous adipose tissue, but not liver, between individuals with and without NAFLD	7
2.4	Selection of 10 serum biomarker candidates (SBCs) for NAFLD	8
2.5	Determination of key SBCs using best subsets modeling approach	8
2.6	Effect of CCDC80 and SOD3 knockdown on human preadipocytes during adipogenesis	11
3	Discussion	17
4	Methods	22
4.1	Study cohorts	22
4.1.1	Kuopio Obesity Surgery Study (KOBS) cohort used for WGCNA and DE analysis	22
4.2	Adipose and liver bulk RNA sequencing in the KOBS cohort	22
4.3	Identification of adipose and liver cell-type marker genes	23
4.4	Weighted gene co-expression network analysis (WGCNA) of KOBS adipose and liver expression data	24

4.5	DE analysis of KOBS adipose and liver expression data	25
4.6	Filtering of adipose NAFLD DE genes for adipose-origin serum biomarker candidates	26
4.7	Best subsets approach to identify key SBCs	27
4.8	SGBS Cell culture	28
4.9	CCDC80 and SOD3 siRNA knockdown and sample collection for RNA-seq experiment	28
4.10	siRNA knockdown RNA-seq library preparation	29
4.11	Alignment and quantification of siRNA knockdown RNA-seq data	29
4.12	DE analysis of siRNA knockdown expression data	30
	References	31

LIST OF FIGURES

1	Study design to discover CCDC80, SOD3, and COL6A2 as novel serum biomarker candidates (SBCs) for obesity-related NAFLD	5
2	A total of 974 genes are differentially expressed (DE) in subcutaneous adipose tissue between the obese individuals with the three main NAFLD traits, steatosis, fibrosis and/or NASH, and the obese individuals with healthy livers	9
3	Filtering of subcutaneous adipose NAFLD DE genes to select serum biomarker candidates (SBCs)	10
4	Selection of the key SBCs using the best subsets analysis, motivated by our prior gene-gene correlations observed in the adipose expression of the SBCs	12
5	CCDC80 knockdown in human preadipocytes differentiated to adipocytes activates known drivers of adipogenesis	14
S1	A total of 698 genes are differentially expressed (DE) for steatosis, and 282 genes for fibrosis, in the obese adipose tissue from the KOBS participants	15
S2	Knockdown of SOD3 in human preadipocytes differentiated to adipocytes deactivates known drivers of healthy energy homeostasis	16

ACKNOWLEDGMENTS

We thank the individuals who participated in the KOBS and GTEx studies. We also thank the individuals who participated in the Finnish Twins and CRYO snRNA-seq studies. The work was supported by NIH grants R01HG010505 and R01DK132775. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this thesis were obtained from the GTEx Portal on 6/27/2020. The HPA data used for the analyses described in this thesis were obtained from the Human Protein Atlas on 6/27/2020.

This thesis is a version of the following manuscript:

Nicholas Darci-Maher¹, Marcus Alvarez¹, Uma Thanigai Arasu², Seung Hyuk T. Lee¹, David Z. Pan¹, Zong Miao¹, Sankha Subhra Das¹, Dorota Kaminska^{1,3,4}, Tiit Örd², Jieping Yang⁵, Zhaoping Li⁵, Brandon Han⁶, Michael P. Mellody^{6,7}, Robert Damoiseaux^{6,7,8,9}, Jihane N. Benhammou¹⁰, Joseph R. Pisegna¹¹, Janet S. Sinsheimer^{1,12,13}, Minna U. Kaikkonen², Jussi Pihlajamäki^{3,14}, Päivi Pajukanta^{1,15,16}. Cross-tissue omics analysis discovers CCDC80 and SOD3 as novel serum biomarker candidates for NAFLD. Manuscript in preparation for publication.

¹Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, USA.

²A. I. Virtanen Institute for Molecular Sciences, University of Eastern Finland, Kuopio, Finland.

³Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland.

⁴Turku PET Centre, Turku University Hospital, Turku, Finland.

⁵Department of Medicine, Center for Human Nutrition, David Geffen School of Medicine at UCLA, Los Angeles, USA.

⁶California NanoSystems Institute, University of California, Los Angeles, Los Angeles, USA.

⁷Department of Bioengineering, University of California, Los Angeles, Los Angeles, USA.

⁸Department of Molecular and Medical Pharmacology, University of California, Los Angeles, Los Angeles, USA.

⁹Jonsson Comprehensive Cancer Center, UCLA, Los Angeles, USA.

¹⁰Vatche and Tamar Manoukian Division of Digestive Diseases and Gastroenterology, Hepatology and Parenteral Nutrition, David Geffen School of Medicine at UCLA and VA Greater Los Angeles HCS, Los Angeles, USA.

¹¹Department of Medicine and Human Genetics, Division of Gastroenterology, Hepatology and Parenteral Nutrition, David Geffen School of Medicine at UCLA and VA Greater Los Angeles HCS, Los Angeles, USA.

¹²Department of Biostatistics, UCLA Fielding School of Public Health, Los Angeles, USA.

¹³Department of Computational Medicine, David Geffen School of Medicine at UCLA, Los Angeles, USA.

¹⁴Department of Medicine, Endocrinology and Clinical Nutrition, Kuopio University Hospital, Kuopio, Finland.

¹⁵Bioinformatics Interdepartmental Program, UCLA, Los Angeles, USA.

¹⁶Institute for Precision Health, David Geffen School of Medicine at UCLA, Los Angeles, USA.

Author Contributions: N. D., M. A., and P. P. designed the study. N. D., M. A., S. H. T. L., D. Z. P., and Z. M. performed the computational and statistical analyses. U. T. A., T. Ö., and M. U. K. conducted the siRNA knockdown experiments in SGBS preadipocytes and collected the adipogenesis RNA-seq samples. J. P., D. K., M. A., and P. P. collected the KOBS RNA-seq data. M. A., J. N. B, P. P., and J. R. P. generated the liver snRNA-seq data. J. Y., Z. L., B. H., M. P. M., and R. D. contributed to validation experiments. S. S. D. contributed to data visualization. N. D. and P. P. wrote the manuscript, and all other authors read, edited, and/or approved of the manuscript.

CHAPTER 1

Introduction

Non-alcoholic fatty liver disease (NAFLD) is a highly prevalent disorder that affects $\sim 25\%$ of people globally [1]. NAFLD represents a heterogeneous spectrum of liver disease, ranging from simple steatosis to liver fibrosis and non-alcoholic steatohepatitis (NASH) [2]. NAFLD can ultimately lead to liver cirrhosis, and is expected to become the leading cause of liver transplantation within this decade [3]. Heterogeneity in NAFLD etiology and pathogenesis is also reflected by the fact that while obesity is the key risk factor for NAFLD, 5-40% of NAFLD patients are normal weight, depending on the population [4–6].

NAFLD manifests in the liver, but prevailing theories suggest that the obesity-driven form of NAFLD originates in adipose tissue [2, 7–11]. It has been hypothesized that some obese individuals cannot generate new adipocytes (hypoplasia) effectively enough to store extra fat, and instead their existing adipocytes become larger (hypertrophy) [2, 10, 12]. These large adipocytes tend to undergo cellular death, attracting infiltrations of inflammatory cells, such as macrophages, which ultimately causes low grade inflammation and deteriorates adipose tissue functions [2, 10, 12]. As a result, adipose lipolysis and hepatic de novo lipogenesis increase, releasing free fatty acids into the bloodstream [2, 10, 13]. This drives ectopic fat deposits onto vital organs, including the liver. These deposits evoke macrophage infiltration and inflammation in the liver, which the liver attempts to repair with scar tissue, i.e. fibrosis [2, 10]. Without weight loss intervention, the obese adipose tissue becomes increasingly dysfunctional, and the liver becomes increasingly fibrotic, until the liver is permanently damaged [2, 7–10]. Existing evidence broadly supports this hypothesis [2, 11, 13–15], but many of the

exact molecular factors driving NAFLD pathogenesis remain unknown. Furthermore, the known common NAFLD variants, including the PNPL3, TM6SF2, HSD17B13, and GCKR variants, explain only a small proportion (10–20%) of its heritability [16].

Presently, there is no effective treatment for the obesity-driven advanced forms of NAFLD; however, as simple steatosis is still reversible through weight loss, early diagnosis would be extremely important [2, 8]. While a variety of diagnosis strategies currently exist for the various stages of the NAFLD spectrum, these strategies are either too invasive (e.g. liver biopsy) or too expensive (e.g. magnetic resonance imaging (MRI)) to be implemented in primary health care [2, 17]. Liver biopsy can accurately identify the stages of steatosis, fibrosis, and NASH by direct histological assessment of the liver tissue [2, 9]. However, a liver biopsy is invasive, relatively risky, and prone to bias [18, 19]. Imaging methods, including abdominal ultrasonography, MRI, and elastography, are less invasive [20–22]. However, MRIs are expensive, and ultrasonography and elastography have low sensitivity in detecting early steatosis cases and cannot robustly detect NASH [2]. Existing serum biomarker panels [23–29] are noninvasive and inexpensive. However, predictive power of these models remains limited, with area under the receiver operating curve (AUC) ranging from 0.66 to 0.87 [2, 17]. Because of these diagnostic challenges, the early stages of NAFLD go largely underdiagnosed, and many patients already exhibit fibrosis by the time NAFLD is detected.

In our study, we used a cross-tissue omics approach to search for effective subcutaneous adipose tissue-origin biomarkers for NAFLD and elucidate their roles in NAFLD pathogenesis, using dual-tissue transcriptomic data from a cohort of extremely obese individuals who underwent bariatric surgery. First, we found proof-of-concept evidence for the adipose origin of NAFLD by discovering an adipose co-expression network, enriched for fatty acid degradation and insulin signaling pathway genes, to be highly positively correlated with a liver co-expression network, enriched for amino acid biosynthesis genes. This indicates that normal healthy adipose tissue functions are correlated with normal healthy liver functions, given that the adipose network was negatively correlated with serum triglyceride (TG) lev-

els and the liver network was negatively correlated with NAFLD. Next, to find key adipose genes involved in NAFLD, we searched for genes differentially expressed (DE) in adipose tissue but not in liver between individuals with histology-based healthy liver and those with NAFLD. We then filtered these adipose NAFLD DE genes for 10 likely serum biomarker candidates (SBCs), which were DE in adipose but not liver, secreted to serum, and exhibited adipose aware expression. We defined key SBCs using a best subsets approach, where we searched for the minimum number of SBCs whose adipose expression explained the maximum variance in NAFLD. Next, we conducted siRNA knockdown experiments with the key fibrosis and NASH SBC genes in human preadipocytes differentiated to adipocytes. In these adipogenesis experiments, we found that knockdown of the SBC Coiled-Coil Domain Containing 80 (CCDC80) significantly increased the expression of the fatty acid synthesis master transcription factor (TF) Sterol Regulatory Element Binding Transcription Factor 1 (SREBF1) [30, 31], and knockdown of the SBC Superoxide Dismutase 3 (SOD3) significantly decreased the expression of the satiety signaling protein Leptin (LEP) [32]. Our study has the potential to substantially improve patient outcomes by discovering genes which may drive the pathogenesis of obesity-induced, adipose-origin NAFLD, and could be developed into a serum biomarker panel to noninvasively detect NAFLD.

CHAPTER 2

Results

2.1 Study design

We developed an integrative cross-tissue transcriptomics approach to search for adipose-origin secreted serum biomarker candidates for NAFLD (Figure 1), leveraging a cohort of extremely obese individuals with RNA-seq data available from both adipose (n=262) and liver tissue (n=267), as well as NAFLD traits diagnosed by liver histology. We defined NAFLD using steatosis, fibrosis, and NASH, which were measured by histological assessment of liver tissue. First, we searched for correlation between adipose and liver co-expression networks. Next, we scanned for adipose aware DE genes, which were genes DE for NAFLD traits in subcutaneous adipose tissue but not liver. We then identified SBCs in this list of adipose aware DE genes by selecting genes that coded for secreted proteins, were expressed highly enough in adipose tissue to be detected in serum, and were expressed much higher in adipose tissue than in liver. We selected key SBCs from this list using best subsets analysis. Finally, we investigated the functions of key SBCs in adipose tissue by knocking them down in a culture of human preadipocytes differentiating to adipocytes.

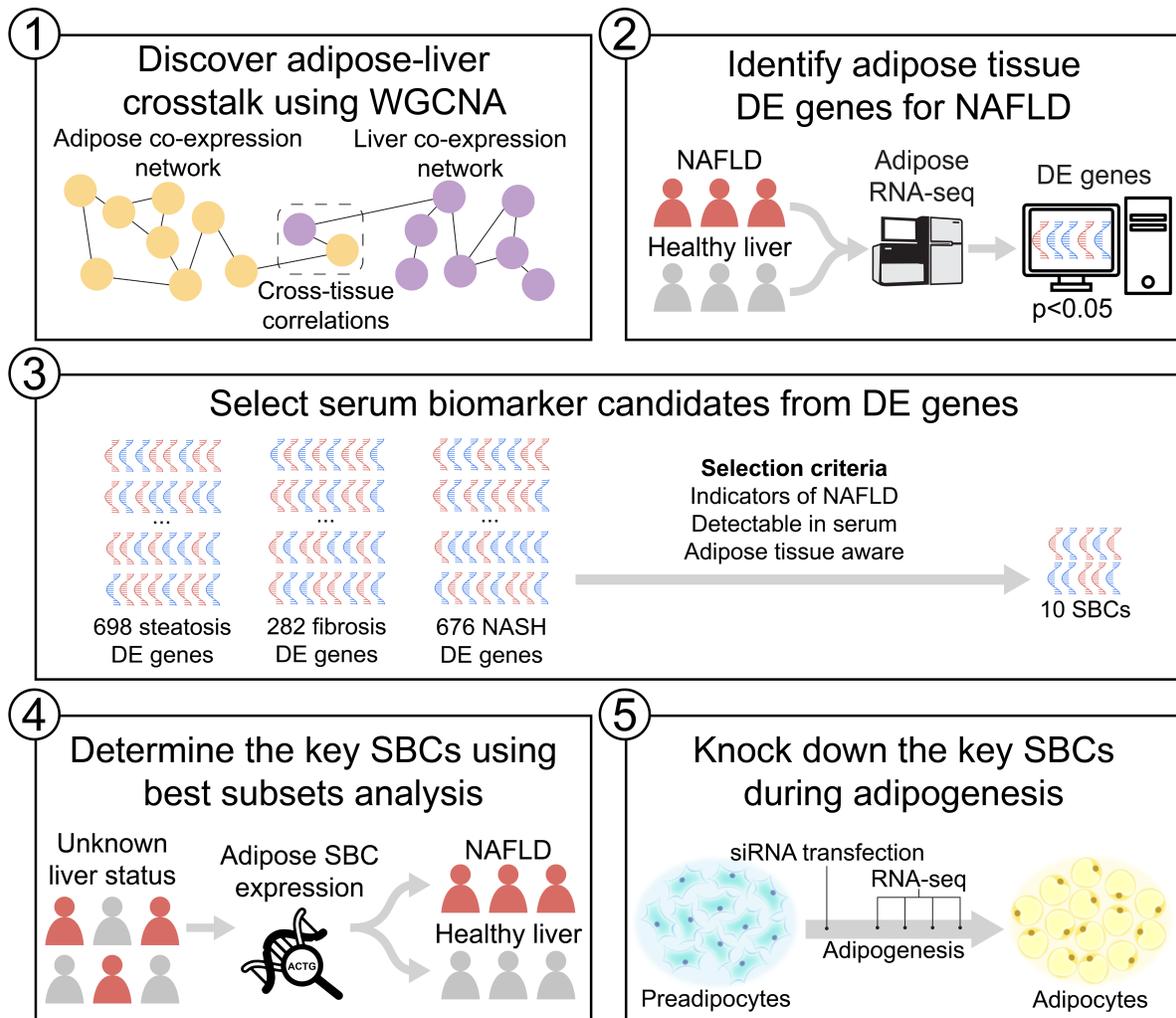


Figure 1: Study design to discover *CCDC80*, *SOD3*, and *COL6A2* as novel serum biomarker candidates (SBCs) for obesity-related NAFLD. To discover novel SBCs for obesity-related NAFLD, we leveraged a unique dual-tissue transcriptomic cohort with histology-based diagnosis of steatosis, fibrosis, and NASH. First, we discovered molecular crosstalk between adipose tissue and liver using WGCNA. Next, we scanned for genes DE in adipose tissue for the three NAFLD traits. We filtered these DE genes for secreted proteins, i.e. SBCs, using a set of selection criteria, and determined the key SBCs using best subsets analysis. Finally, we followed up the key SBCs functionally by knocking them down in human preadipocytes during adipogenesis.

2.2 Co-expression networks of distinct functional pathways correlate across an individual’s subcutaneous adipose tissue and liver

To search for signatures of molecular crosstalk between adipose and liver tissue related to NAFLD, we used our dual-tissue cohort to construct gene co-expression networks separately in adipose and liver tissue using the R [33] package WGCNA [34, 35], and related these networks to each other (see Methods). To investigate the functional significance of the modules (i.e. networks), we correlated all adipose and liver module eigengenes (MEs) with common metabolic traits and NAFLD liver histology measurements. To identify networks involved in tissue crosstalk, we correlated every adipose ME with every liver ME.

In agreement with our hypothesis, we found evidence that the healthy functions of adipose tissue, consisting of storing and burning fat, are positively associated with the healthy functions of liver tissue, consisting of synthesizing biomolecules into fatty acids. This was represented by positive correlation between the adipose lightyellow network (ALY) and the liver saddlebrown network (LSB) ($r=0.331$, $p=2.296 * 10^{-7}$). ALY is negatively correlated with serum TG ($r=-0.252$, $p=9.979 * 10^{-5}$), and is enriched for regulation of lipolysis in adipocytes (false discovery rate (FDR)=0.0156), insulin signaling pathways (FDR=0.0156), and adipocyte cell-type marker genes ($p_{\text{hypergeometric}}=1.453 * 10^{-7}$). LSB is negatively correlated with steatosis ($r=-0.378$, $p=2.572 * 10^{-9}$), fibrosis ($r=-0.306$, $p=1.982 * 10^{-6}$), NASH ($r=-0.384$, $p=1.302 * 10^{-9}$), type 2 diabetes mellitus (T2D) ($r=-0.291$, $p=6.284 * 10^{-6}$), and body mass index (BMI) ($r=-0.277$, $p=1.825 * 10^{-5}$), and is enriched for the biosynthesis of amino acids pathway (FDR=1.725x10-9) and hepatocyte cell-type marker genes ($p_{\text{hypergeometric}}=1.842 * 10^{-3}$, $2.202 * 10^{-3}$, $1.292 * 10^{-4}$, $7.143 * 10^{-4}$, and $1.513 * 10^{-3}$ for Hep-7, Hep-9, Hep-10, Hep-11, and Hep-13, respectively).

Additionally, our results suggest that inflamed and dysfunctional adipose tissue is associated with a decrease in normal healthy liver function, and an increase of NAFLD and other

adverse metabolic trait functions. This was represented by negative correlation between the adipose cyan network (AC) and LSB ($r=-0.351$, $p=3.561 * 10^{-8}$). AC is positively correlated with NASH ($r=0.268$, $p=3.303 * 10^{-5}$), and is enriched for autoimmune and inflammatory pathways, including inflammatory bowel disease ($FDR=1.824 * 10^{-11}$) and autoimmune thyroid disease ($FDR=5.558 * 10^{-13}$).

Taken together, our gene co-expression network results indicate that adipose and liver tissue communicate at the transcriptional level, and that this communication is grossly inverted in NAFLD. To investigate the details of this tissue crosstalk, we aimed to move past the co-expression network level and discover individual genes as indicators for adipose dysfunction-related NAFLD.

2.3 645 genes are DE in subcutaneous adipose tissue, but not liver, between individuals with and without NAFLD

To first identify genes whose adipose expression was associated with three key histology-based NAFLD traits (steatosis, fibrosis, and NASH), we ran DE analysis with the R [33] limma-voom pipeline [36–38]. We compared the adipose expression of individuals with steatosis, fibrosis, and NASH to those with healthy livers, while correcting for common demographic and technical confounders (see Methods). We identified 974 genes DE for at least one NAFLD histology trait (698, 282, and 676 DE genes for steatosis, fibrosis, and NASH, respectively) (Figure 2). To select genes with adipose tissue aware expression, we filtered out all liver DE genes for the same three NAFLD traits. This resulted in 645 total adipose aware DE genes (437, 187, and 469 adipose aware DE genes for steatosis, fibrosis, and NASH, respectively). These adipose aware DE genes were significantly enriched in both the ALY ($p_{\text{hypergeometric}}=1.467 * 10^{-7}$) and AC ($p_{\text{hypergeometric}}=6.430 * 10^{-13}$) adipose co-expression networks, enforcing our hypothesis that dysfunctional adipose tissue is associated with NAFLD. Next, we focused on the genes among these 645 that were most likely to be detectable in

serum to discover potential adipose-origin serum biomarkers for NAFLD.

2.4 Selection of 10 serum biomarker candidates (SBCs) for NAFLD

We reasoned that the adipose aware NAFLD DE genes that are effective SBCs must leave the cell, be expressed at sufficient levels in their source tissue to be detectable in serum, and have predominantly adipose aware expression. To implement these constraints, we filtered the 645 adipose aware DE genes for the ones present in the Human Protein Atlas (HPA) [39] list of secreted proteins, and with median transcripts per million (TPM) greater than 30 in subcutaneous adipose tissue, using data from the GTEx portal. Additionally, we filtered out all genes whose ratio of subcutaneous adipose median TPM to liver median TPM was less than 10. This design resulted in a final list of 10 SBCs: CCDC80, CD300LG, COL6A1, COL6A2, GPX3, MGP, SFRP2, SOD3, TIMP3, and VEGFB (Figure 3). Taken together, all SBCs are DE in subcutaneous adipose tissue for at least one NAFLD trait (steatosis, fibrosis, or NASH), are not DE in liver for any of the same three NAFLD traits, code for secreted proteins, have median TPM >30 in subcutaneous adipose tissue, and have $>10x$ higher median TPM in subcutaneous adipose tissue than in liver tissue.

2.5 Determination of key SBCs using best subsets modeling approach

To find the best subset of these 10 genes for evaluating NAFLD risk, we modeled the effect of their adipose expression on NAFLD. First, we observed that there are significant adipose expression gene-gene correlations among the 10 SBCs (Figure 4a), indicating that they are not fully independently expressed in the adipose tissue. To avoid redundancy, we then searched for the minimum set among the 10 SBCs whose adipose expression explained

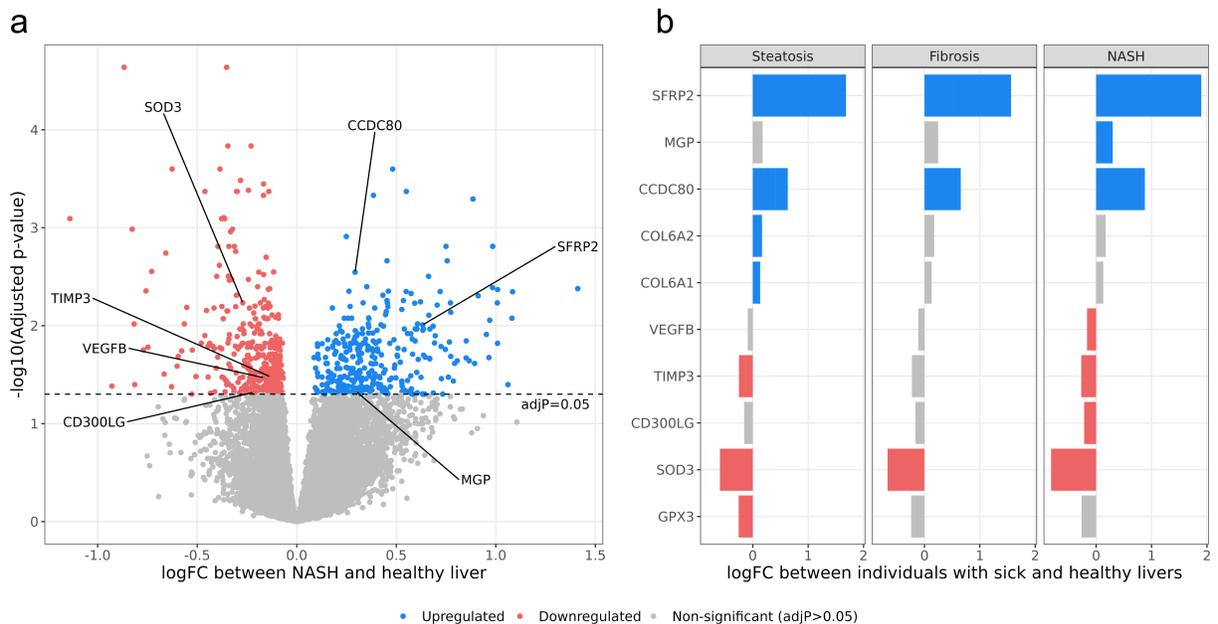


Figure 2: A total of 974 genes are differentially expressed (DE) in subcutaneous adipose tissue between the obese individuals with the three main NAFLD traits, steatosis, fibrosis and/or NASH, and the obese individuals with healthy livers. We performed DE analysis on bulk RNA-seq data from subcutaneous adipose biopsies in the KOBs cohort, comparing individuals with the NAFLD traits diagnosed by liver histology to those with healthy livers. Of the 974 adipose DE genes, 698, 282, and 676 genes are DE for steatosis, fibrosis, and NASH, respectively. (a) Volcano plot showing the results of the NASH DE analysis in the adipose tissue. The X-axis represents log fold-change (logFC) from individuals with NASH and those with healthy livers. The Y-axis represents the negative log of the DE p-value, adjusted for multiple testing. Significant SBCs identified in our subsequent filtering steps (Figure 3) are highlighted. Volcano plots of steatosis and fibrosis DE results are shown in Figure S1. (b) Bar plot showing the DE direction of the SBCs in steatosis, fibrosis, and NASH. X-axis represents logFC from individuals with each NAFLD trait and those with healthy livers. Y-axis represents the SBC name. Blue SBCs have increased adipose expression in individuals with NAFLD when compared to the individuals with healthy livers, while red SBCs have decreased adipose expression.

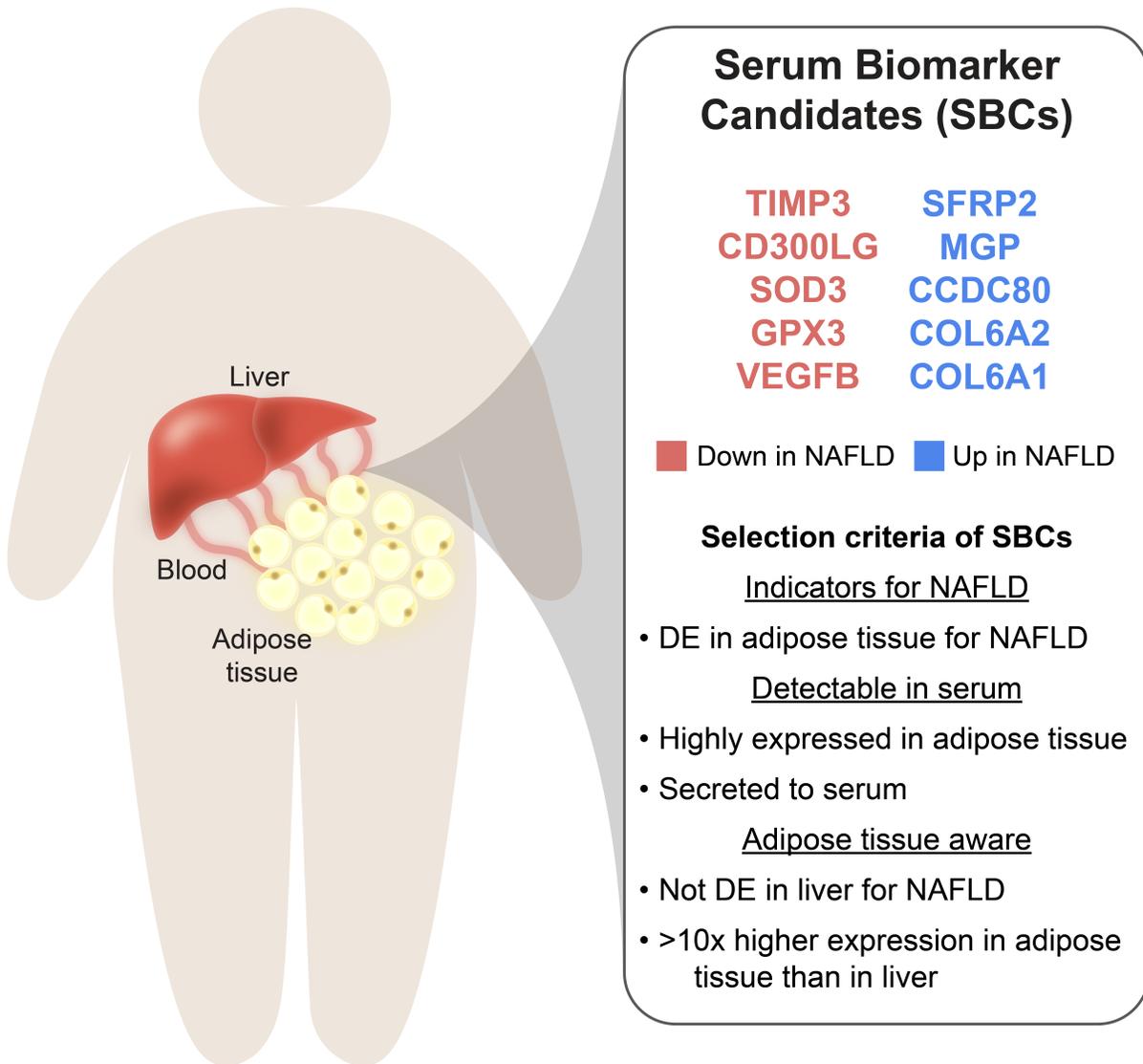


Figure 3: Filtering of subcutaneous adipose NAFLD DE genes to select serum biomarker candidates (SBCs). To identify SBCs among the list of 974 adipose NAFLD DE genes, we selected the genes that were DE for NAFLD in adipose tissue but not in the liver, coded for proteins secreted to serum, had moderate to high expression in adipose tissue, and had >10x higher expression in subcutaneous adipose tissue than in the liver. These filters reduced the list of 974 total adipose DE genes across steatosis, fibrosis, and NASH to 10 SBCs. Blue genes are upregulated in steatosis, fibrosis, and/or NASH in adipose tissue, while red genes are downregulated.

the maximum amount of variance in NAFLD, using a best subsets approach with the R [33] package leaps [40]. In this approach, we fit linear models for different combinations and numbers of SBCs, and tested the variance in steatosis, fibrosis, and NASH explained by each combination of genes, while correcting for the same covariates we used in the WGCNA and DE analyses (see Methods). We discovered that the gene COL6A2 explains the most variation in steatosis ($r^2=0.055$, $p_{\text{permutation}}=0.012$), and the genes CCDC80 and SOD3 explain the most variation in both fibrosis and NASH ($r^2=0.102$, $p_{\text{permutation}}=1.970 * 10^{-3}$ for fibrosis; $r^2=0.166$, $p_{\text{permutation}}=3.500 * 10^{-4}$ for NASH) (Figure 4b). This result further strengthens the premise of CCDC80 and SOD3 as biomarkers, because our permutation results show that their adipose expression explains more variation in fibrosis and NASH than virtually all other pairs of genes.

2.6 Effect of CCDC80 and SOD3 knockdown on human preadipocytes during adipogenesis

Because CCDC80 and SOD3 were observed as the strongest SBCs for both fibrosis and NASH, we next investigated their effects on adipogenesis *in vitro* using an siRNA knockdown experiment. In this experiment, we cultured human SGBS preadipocytes over the course of differentiation from preadipocytes to adipocytes, and collected bulk RNA-seq data at four time points. We first confirmed that the knockdown was effective, as evidenced by the downregulation of both CCDC80 and SOD3 ($p<0.05$) in their respective knockdown conditions (Figure 5b, S2). Next, we searched for DE genes between control and separate knockdown of CCDC80 and SOD3 at each adipogenesis time point from baseline to seven days (see Methods). Because we were interested in the impact of CCDC80 and SOD3 knockdown on adipogenesis specifically, we restricted the genes tested for DE to a list of preadipocyte, adipocyte, and adipogenesis marker genes ($n=492$ genes tested, see Methods).

We found evidence suggesting that CCDC80 contributes to NAFLD progression by in-

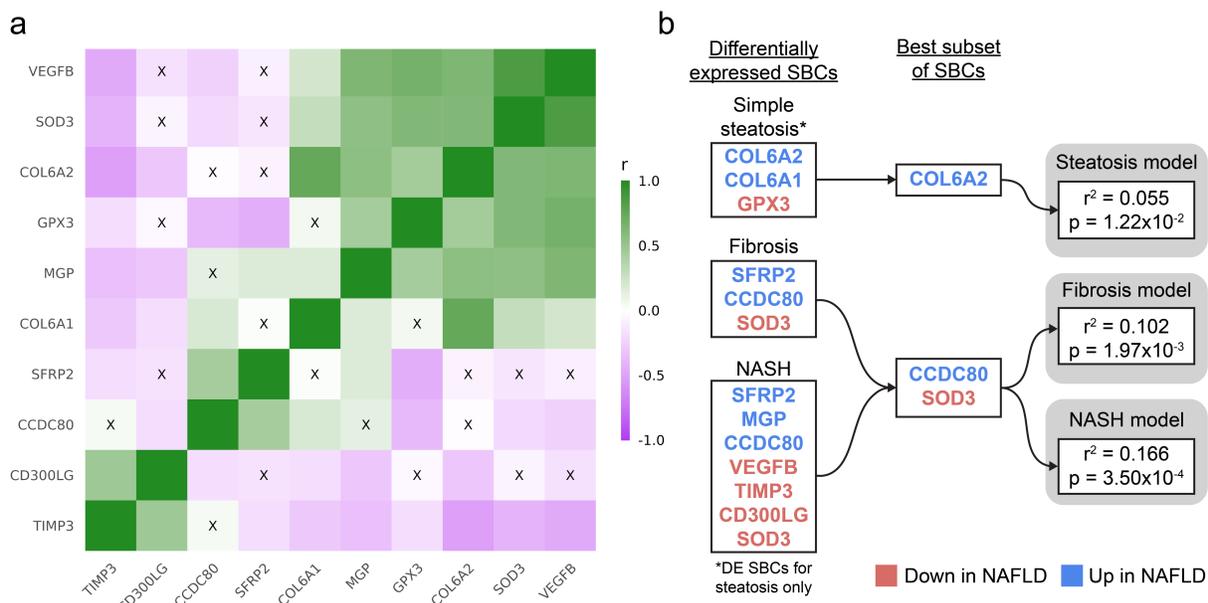


Figure 4: Selection of the key SBCs using the best subsets analysis, motivated by our prior gene-gene correlations observed in the adipose expression of the SBCs. We filtered the list of 10 SBCs further by testing the proportion of variance explained in steatosis, fibrosis, and NASH by the adipose expression of the SBCs. (a) Pairwise gene-gene correlation structure between the adipose expression of the SBC genes. Each box represents the strength of the pairwise Pearson correlation between the adipose expression of the SBCs. Green boxes correspond to a positive correlation, and purple boxes correspond to a negative correlation. “X” indicates that the correlation is non-significant. Genes are ordered by the first principal component (PC). These correlations motivate the idea that a small subset of the SBCs can capture most of the expression of all 10 SBCs, which we then tested in the best subsets analysis. (b) Results of the best subsets analysis. For steatosis, fibrosis, and NASH, the best subset of significant SBCs was chosen by the leaps algorithm, based on the variance in the NAFLD trait explained by each combination of genes. P-values were calculated based on a permutation test ($B=100,000$) (see Methods). To capture genes involved in the early onset of NAFLD, only the 3 genes that were uniquely DE for steatosis in the adipose tissue were considered for the steatosis model.

hibiting the ability of adipose tissue to produce new adipocytes to store fat. This was supported by the observation that knockdown of CCDC80 during adipogenesis significantly increased the expression of fatty acid master transcription factor SREBF1 at 7 days (log fold-change in knockdown compared to control (logFC)=1.547, $p=8.608 * 10^{-4}$), as well as TG hydrolysis enzyme Lipoprotein Lipase (LPL) at 7 days (logFC=2.597, $p=7.215 * 10^{-4}$) (Figure 5). Of the 141 adipogenesis pathway genes we downloaded from WikiPathways, 13 were DE during at least one timepoint in the CCDC80 knockdown.

We also found evidence suggesting that SOD3 protects against NAFLD by promoting a healthy satiety feedback loop. This was supported by the fact that the knockdown of SOD3 during adipogenesis significantly decreased the expression of the satiety signaling protein LEP at 4 days (logFC=-0.651, $p=1.966 * 10^{-4}$) (Figure S2). 17 of the adipogenesis pathway genes were DE during at least one timepoint in the SOD3 knockdown.

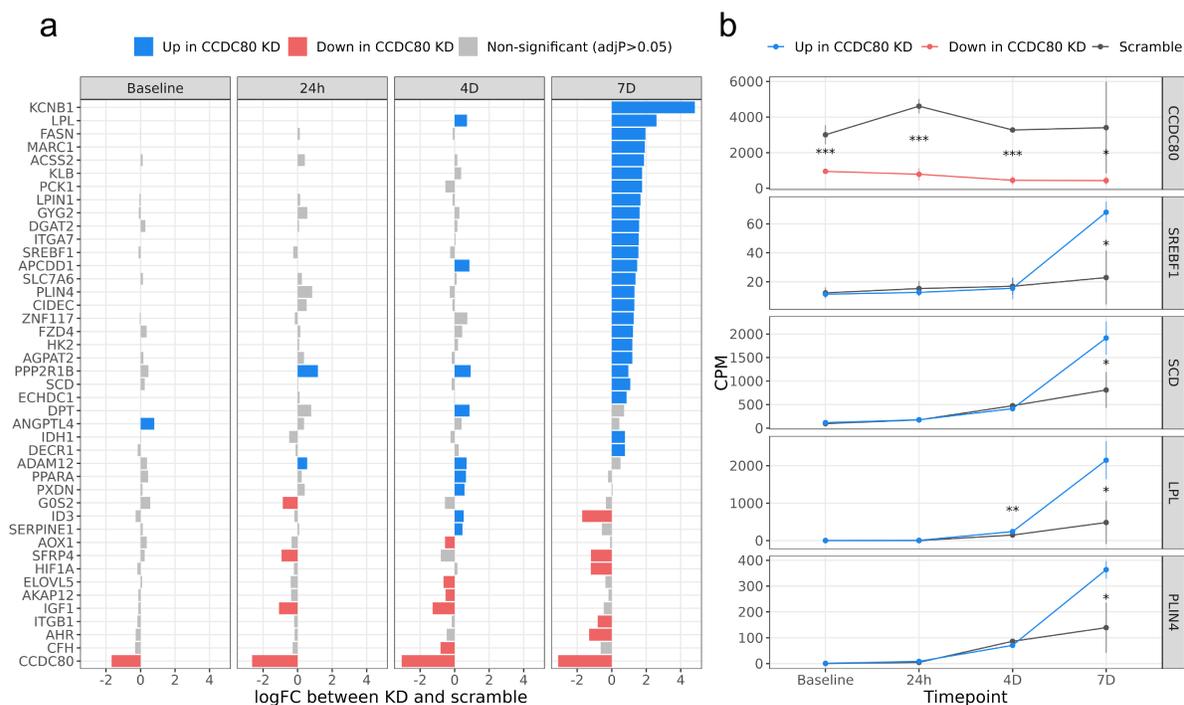


Figure 5: CCDC80 knockdown in human preadipocytes differentiated to adipocytes activates known drivers of adipogenesis. We knocked down CCDC80 using siRNA transfection in a culture of human SGBS preadipocyte cells (see Methods), and measured expression via RNA-seq at 4 time points during adipogenesis. We then performed a differential expression (DE) analysis between the CCDC80 gene knockdown and scramble conditions at each time point. (a) Results of the DE analysis. The X-axis represents the log fold-change (logFC) of all 43 genes which were DE in at least one time point; the Y-axis the gene names; and facets the time points. Blue genes were expressed significantly more in the CCDC80 knockdown, and red genes were expressed less. (b) Mean expression of CCDC80 and selected well known examples of adipogenesis genes in the scramble and knockdown samples. The X-axis represents the time point; the Y-axis counts per million (CPM); facets the gene name; error bars the mean \pm standard deviation; and colors the direction of DE at 7D. Stars indicate the significance of DE between the knockdown and scramble samples: “***”=adjP<0.001; “**”=adjP<0.01; “*”=adjP<0.05; “.”=adjP < 0.1.

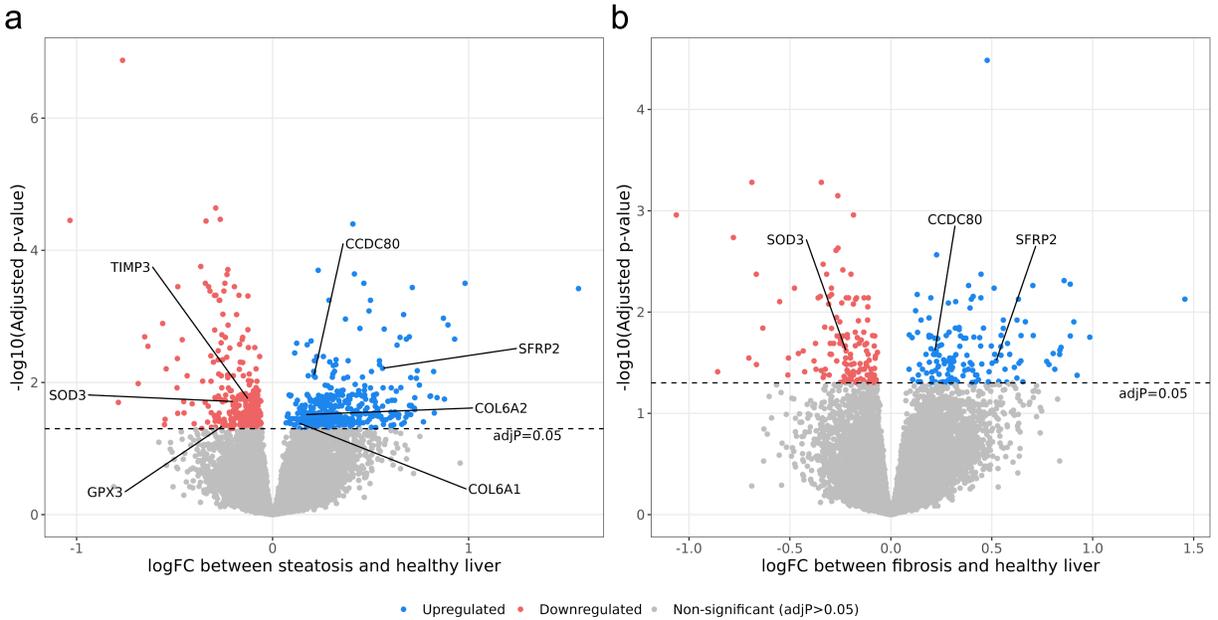


Figure S1: A total of 698 genes are differentially expressed (DE) for steatosis, and 282 genes for fibrosis, in the obese adipose tissue from the KOBS participants. We performed DE analysis on bulk RNA-seq data from subcutaneous adipose biopsies in the KOBS cohort, as described in Figure 2. Volcano plots show the results of the steatosis (a) and fibrosis (b) DE analyses in the adipose tissue. In both plots, the X-axis represents log fold-change (logFC) in adipose bulk RNA-seq data from individuals with NASH and those with healthy livers. The Y-axis represents the negative log of the DE p-value, adjusted for multiple testing with the Benjamini-Hochberg procedure. Significant SBCs identified in our filtering steps (Figure 3) are highlighted.

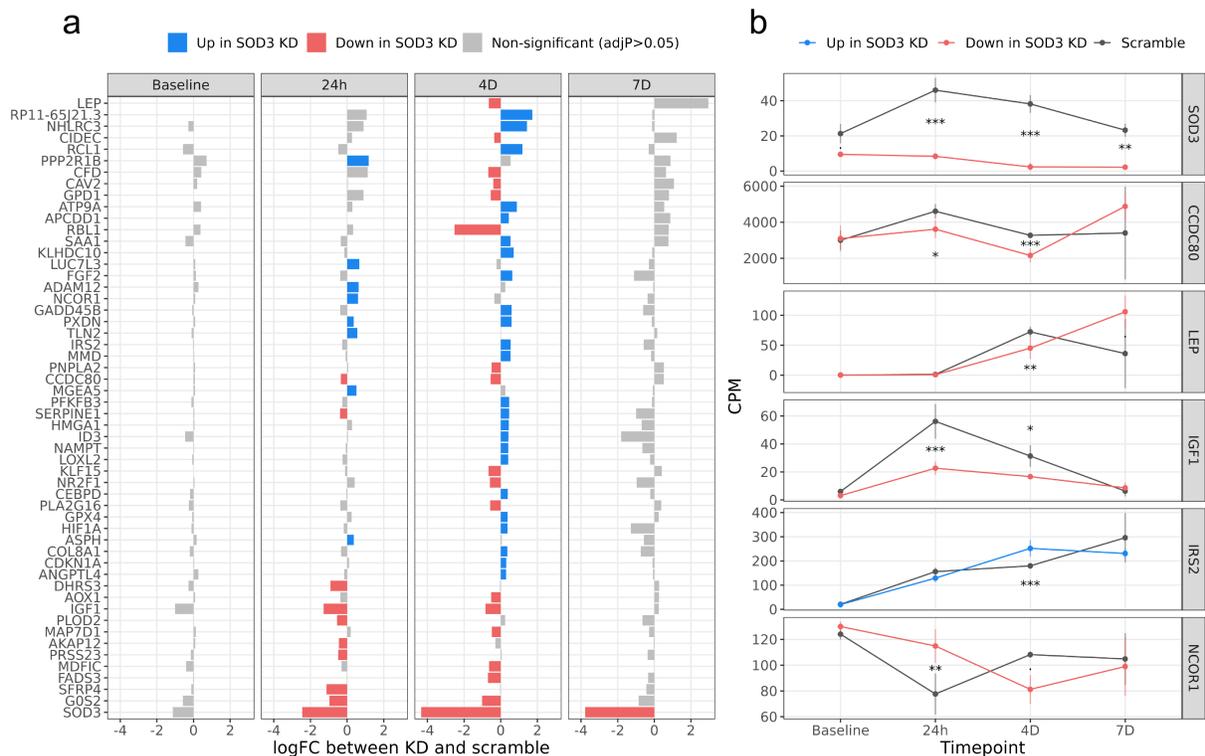


Figure S2: Knockdown of SOD3 in human preadipocytes deactivates known drivers of healthy energy homeostasis. We knocked down SOD3 using siRNA transfection in a culture of human SGBS preadipocyte cells (see Methods), and performed differential expression (DE) analysis in the same way as the CCDC80 knockdown analysis described in Figure 5. (a) Results of the DE analysis. The X-axis represents the log fold-change (logFC) of all 54 genes which were DE in at least one time point during the differentiation; the Y-axis the gene names; and facets the time points. Blue genes were expressed significantly more in the SOD3 knockdown, and red genes were expressed less. (b) Mean expression of SOD3 and selected well known examples of adipogenesis and satiety signaling genes in the scramble and knockdown samples during differentiation. The X-axis represents the time point; the Y-axis counts per million (CPM); facets the gene name; error bars the mean±standard deviation; and colors the direction of DE at 4D. Stars indicate the significance of DE between the knockdown and scramble samples: “***”=adjP<0.001; “**”=adjP<0.01; “*”=adjP<0.05; “.”=adjP < 0.1.

CHAPTER 3

Discussion

We hypothesized that in some obese individuals, obesity induces pathological inflammatory changes in the subcutaneous adipose tissue, leading to ectopic deposition of fat into the liver and driving the development of NAFLD, and that these mechanisms could be traced without liver biopsy or abdominal imaging using adipose aware SBCs. To test our hypothesis, we first demonstrated that there is crosstalk between the adipose tissue and liver by identifying correlation between adipose and liver gene co-expression networks associated with NAFLD and related metabolic traits. Next, we identified 645 adipose tissue aware DE genes for liver histology-based NAFLD phenotypes in extremely obese individuals with and without NAFLD. Filtering these adipose aware DE genes resulted in the identification of 10 SBCs, which are DE in adipose, are not DE in liver, show adipose aware expression, and encode proteins secreted to serum. We then identified CCDC80 and SOD3 as the key SBCs for fibrosis and NASH based on the best subsets analysis, in which they explain the maximum variance in fibrosis and NASH compared to all other subsets of DE SBCs, as well as other pairs of genes genome-wide in our permutation analysis. Finally, we demonstrated that both CCDC80 and SOD3 influence adipogenesis by knocking them down during human preadipocyte differentiation.

Previous work has utilized transcriptomics data and direct serum protein measurements paired with NAFLD diagnosis to search for noninvasive biomarkers for NAFLD [17, 41–43]. However, to the best of our knowledge, our study is the first to leverage a dual-tissue cohort with both adipose tissue and liver RNA-seq available, paired to a gold-standard NAFLD

diagnosis using liver histology. With our cross-tissue analysis, which scanned genome-wide for SBCs, we add this unique dual-tissue perspective, focused on obesity-driven, adipose-origin, NAFLD, to the body of previous NAFLD studies [17, 41–43]. While previous work has assessed the connection between either NAFLD and the adipose transcriptome [44, 45] or NAFLD and the liver transcriptome [43], our study made use of RNA-seq data measured from both adipose and liver in the same individuals. This was crucial in identifying our SBCs, because we were able to pinpoint genes involved in the adipose origin of NAFLD, specifically by removing the liver NAFLD DE genes. We reasoned that a large proportion of the liver NAFLD DE genes would be driven by the immune responses of the liver to the ectopic fat deposits present in NAFLD, and thus they would not be ideal biomarkers for detecting the onset of adipose-origin NAFLD. Detecting adipose aware SBCs is important, because adipose tissue is known to secrete a wide array of signaling proteins [46, 47], which opens up the possibility for capturing the specific adipokines associated with NAFLD in serum.

Our selection criteria for SBCs, which are specifically tailored to a cross-tissue transcriptomic design, provide additional novelty to our study. The HPA secretome has been applied previously to search for biomarkers [48], and the GTEx median TPM data has been applied to study NAFLD [49]; however, ours is the first to apply the two resources in combination as a set of filtering criteria for SBCs. These publicly available datasets allowed us to implement a crucial element of our study design, i.e. the selection of the adipose aware NAFLD DE genes which are most likely to be detectable in serum. Thus, our SBCs are DE in adipose tissue but not the liver for the three key NAFLD traits, steatosis, fibrosis and NASH, and in addition are expressed highly in adipose tissue, secreted to serum, and expressed substantially more in the adipose tissue than in the liver.

Due to our integrative filtering design, CCDC80 and SOD3 are likely to be strong indicators of NAFLD. Both of these genes meet all of our filtering criteria for SBCs, and our significant genome-wide permutation results show that they explain more variance in fibrosis

and NASH than all other pairs of genes. While our knockdown results do not prove that CCDC80 and SOD3 are causal in the pathogenesis of NAFLD, a biomarker can be effective whether or not it is causal. Our evidence that SOD3 and CCDC80 modulate healthy adipose tissue functions demonstrates their premise as indicators of the onset of NAFLD, either as responsive or causal players, and validates our computational methods of discovery. Additionally, previous work on the function of both genes aligns with our results [50–53].

As we observed in our study, CCDC80 has been shown to associate positively with adverse metabolic traits, including fatty liver [50–52]. In several mouse models described previously, CCDC80 knockdown decreased plasma TGs [50], and CCDC80 knockout promoted adipogenesis [51]. In a human study, CCDC80 was quantified in serum, and serum CCDC80 levels correlated positively with obesity risk, inflammation markers, and liver steatosis [52]. It has been proposed that CCDC80 increases hypertriglyceridemia by decreasing the expression of LPL, a key catalyst in hydrolysis of TGs [50]. Gong et al. observed that CCDC80 knockdown in vascular smooth muscle cells in vitro increased LPL expression, while CCDC80 overexpression decreased LPL [50]. In agreement with this evidence, we observed an upregulation of LPL in our CCDC80 knockdown in preadipocytes during adipogenesis at day 7. Additionally, we observed that the key TF of fatty acid biosynthesis, SREBF1, was significantly upregulated at day 7 in the CCDC80 knockdown during adipogenesis. SREBF1 is widely accepted as a TF promoting adipogenesis [30, 31, 54]. SREBF1 exhibits a steep and sustained increase in expression during the induction stage of adipogenesis [30], preceding the increases in expression of other known master adipogenesis regulators PPARG [55] and CEBPA [30, 56]. Kim et al. demonstrated the direct nature of the interaction between SREBF1 and PPARG by showing that cells expressing SREBF1 preferentially synthesize ligands for PPARG [54]. Thus, our results suggest that CCDC80 contributes to the pathogenesis of NAFLD by preventing adipose tissue from performing its vital functions through adipogenesis.

Our results also corroborate evidence supporting SOD3 as an adipose origin biomarker for

NAFLD. SOD3 is seen as a protective factor against oxidative stress, which has been shown to be a major contributor to the pathogenesis of NAFLD [53, 57–59]. SOD3 knockdown in human adipocytes caused increased accumulation of TGs [53], and global SOD3 knockout mice exhibited increased obesity and insulin resistance [53]. Adipocyte diameters in the white adipose tissue of mice overexpressing SOD3 on a high-fat diet were significantly smaller than those of control mice on a high-fat diet, and were almost identical to control mice on a regular chow diet [60]. Previous work suggests that SOD3 functions as a protective mechanism against NAFLD development by inhibiting the expression of inflammatory genes in adipose tissue [60], which aligns well with our hypothesis that NAFLD onset is triggered by dysfunctional and inflamed adipose tissue. Gao et al. also detected SOD3 protein in the supernatant of human adipocytes, suggesting it is secreted by the adipose tissue [53]. Additionally, we observed that LEP was significantly downregulated in the SOD3 knockdown at day 4 during adipogenesis. LEP encodes an adipokine secreted from adipose tissue that acts on the brain, playing a major role in energy homeostasis and satiety signaling [61–63]. It has been shown that LEP acts as the primary link between adipose tissue and the brain, in a negative feedback loop that decreases hunger urges with increasing energy intake and fat accumulation [61]. However, this same system has been demonstrated to break down in obesity, where LEP deficiency and/or LEP resistance hinder the ability of the body to balance energy intake and expenditure [61]. Taken together, our SOD3 knockdown results suggest that the normal function of SOD3 is to protect against NAFLD by promoting effective energy homeostasis.

CCDC80 and SOD3, along with the full list of 10 SBCs, should be considered for inclusion in a future serum biomarker panel to diagnose NAFLD. Several SBCs have been previously measured in serum, using ELISA kits and other related methods [52, 53, 64–68], however further investigation is needed. While our study leveraged the list of secreted proteins from the HPA, the majority of our analysis was done in a transcriptomic paradigm, under the assumption that mRNA abundance of the SBCs in adipose tissue is an effective proxy for

their corresponding protein levels in serum. Future vetting of our SBCs should involve comparing their protein levels in serum between individuals with and without NAFLD.

Our present study is limited by the fact that our discovery cohort originates from a genetically relatively homogenous European population, the Finns [69]. It is crucial to follow up our work with future studies in populations underrepresented in genomics, including Latin American, African, and Southeast Asian populations. Here, we made use of the European KOBS cohort because it was the first to make our dual-tissue NAFLD study design possible.

Currently, diagnosis of NAFLD requires either a liver biopsy, which necessitates an invasive surgery or inpatient procedure, or abdominal imaging (MRI or elastography), which is costly and time consuming [2]. Furthermore, NAFLD often remains undiagnosed and is therefore grossly underdiagnosed [2], emphasizing the pressing need for SBCs. Our cell culture-validated SBCs have strong potential to be developed into an effective blood panel for NAFLD, which could be used in the primary care setting as an initial screening step. This could allow for more efficient primary care screenings for NAFLD and its precursors, including steatosis, and ultimately improve patient health by catching and treating NAFLD earlier in its development. We envision that high-risk patients for NAFLD, specifically obese patients, could greatly benefit from this diagnostic option.

In conclusion, leveraging dual-tissue RNA-seq data from adipose tissue and liver, paired with liver histology-based NAFLD diagnosis, we discover 10 serum biomarker candidates for NAFLD. We identified *CCDC80* and *SOD3*, which explain maximum variance in fibrosis and NASH compared to all SBCs, as the key SBCs, and followed up this conclusion with functional knockdown experiments throughout adipogenesis. Our methodology can be extended to study cross-tissue communications and discover SBCs in any complex disease, providing that a cohort of RNA-seq data from multiple tissues in the same individuals is available along with detailed phenotype data. Overall, identifying SBCs involved in tissue-tissue crosstalk using our integrative transcriptomics pipeline can contribute to improved understanding and earlier clinical detection and diagnosis of complex diseases in the future.

CHAPTER 4

Methods

4.1 Study cohorts

4.1.1 Kuopio Obesity Surgery Study (KOBS) cohort used for WGCNA and DE analysis

The KOBS cohort was recruited at the University of Eastern Finland among extremely obese Finnish individuals who underwent bariatric surgery, as described in detail previously [70–72]. During the surgery, subcutaneous adipose and liver biopsies were collected for bulk RNA-sequencing as well as serum samples for clinical measurements, as described in detail previously [73]. The KOBS participants have detailed phenotype data measured for histological, metabolic, and anthropometric traits. These include age, sex, BMI, serum lipid and glucose levels, and liver histology measurements (i.e. liver fibrosis, liver steatosis, and NASH diagnosis). All individuals in the KOBS cohort provided a written informed consent, and the study protocols were approved by the local ethics committee.

4.2 Adipose and liver bulk RNA sequencing in the KOBS cohort

The adipose RNA-seq data [70] (n=262) were generated by sequencing TruSeq stranded libraries on the HiSeq4000 sequencing platform, producing an average of 42.38 M reads [70]. The liver RNA-seq data [72] (n=267) were generated by sequencing Ribo-Zero stranded libraries on the HiSeq2500 sequencing platform, producing an average of 39.73 M reads [72].

We aligned both the adipose and liver bulk reads to the GRCh37/hg19 reference using a 2-pass pipeline with STAR [74] and performed QC using PicardTools [75].

4.3 Identification of adipose and liver cell-type marker genes

To identify cell-type marker genes in adipose tissue and liver, we leveraged two additional cohorts where single nucleus RNA sequencing (snRNA-seq) was performed.

In the adipose cohort, snRNA-seq was performed on subcutaneous adipose biopsies from 15 individuals in the Finnish Twin and CRYO studies, as described in detail previously [76]. All individuals provided written informed consent, and the study protocols were approved by the local ethics committee. Filtering was performed with DIEM [77], and clusters were identified using Seurat v3.2.3 [78]. Cell-type annotation was performed using SingleR v1.2.4 [79], and cell-type marker genes were selected based on a Wilcoxon rank-sum test [76].

In the liver cohort, female patients (n=3) underwent surgery at the Dumont-UCLA Liver Cancer Center to treat hepatocellular carcinoma (HCC), as described in detail previously [80]. All participants in the study provided written informed consent, and the study protocols were approved by the UCLA IRB. During the surgery, tumor and adjacent non-tumor biopsies were collected. In the present study, we used only the snRNA-seq samples from non-tumor tissue. To identify marker genes for each liver cell-type, we tested normalized expression between nuclei within and outside a cluster. We normalized raw counts by first scaling all nuclei to sum to 1,000, then log-transforming. Next, we used the FindAllMarkers function from Seurat [78] to run differential expression. For each cell-type, we performed a logistic regression for each gene testing expression of nuclei within the cell-type against those classified as any other cell-type. We kept marker genes with an average \log_2 fold change of at least 0.1. We corrected p-values for multiple testing across all genes and cell-types using FDR.

4.4 Weighted gene co-expression network analysis (WGCNA) of KOBS adipose and liver expression data

To investigate molecular crosstalk between subcutaneous adipose tissue and liver, we used the KOBS expression data to construct weighted gene correlation networks with the R [33] package WGCNA v1.70 [34, 35]. In this and all other statistical analyses, we utilized the R packages ggplot2 [81] and tidyverse [82] extensively. Before creating the networks, we first normalized the expression data according to the developers’ instructions for RNA-seq data. Briefly, we selected genes with nonzero expression in 90% of samples (as described previously [83]), calculated their counts per million (CPM), performed a rank-based inverse normal transformation (INT), regressed out common RNA-seq covariates (age, sex, RNA integrity number (RIN), percent uniquely mapped reads, percent intronic bases, and median 3’ bias), and thereafter performed a second INT. This resulted in 21,408 and 22,500 input genes in the adipose tissue and liver, respectively. After normalization, we verified that no extreme outliers existed in the data by hierarchically clustering the samples.

Next, we constructed two independent co-expression networks, one in the subcutaneous adipose tissue and one in the liver, using WGCNA [34, 35, 84]. We followed the “step-by-step network construction” tutorial from the WGCNA website, which involved calculating an adjacency matrix, converting it to a dissimilarity topological overlap matrix (TOM), clustering genes hierarchically based on the TOM, performing a dynamic tree cut, and merging modules based on their module eigengene (ME) correlation. When constructing the adjacency matrix, we used a soft threshold power of 7 and 10 in adipose and liver, respectively, based on inspection of the plots showing the effect of soft threshold power on mean connectivity and scale free topology model fit. When merging modules, we used a cut height of 0.10 and 0.25 for adipose and liver, respectively, based on inspection of the ME dendrograms. The completed networks contained 57 and 28 modules for adipose and liver, respectively.

With the two networks constructed, we followed the “Relating modules to external clinical

traits” tutorial to correlate all MEs in both networks with relevant metabolic and histological phenotypes: liver steatosis, liver fibrosis, NASH diagnosis, type 2 diabetes mellitus (T2D), statin usage, BMI, triglycerides, and fasting glucose adjusted for T2D. We assessed the significance of these correlations after Bonferroni correction. Additionally, we correlated the MEs of both networks with each other, and assessed the significance of these correlations after Bonferroni correction.

Finally, we calculated the functional enrichment of modules in both networks. First, we calculated the KEGG pathway enrichment using an overrepresentation analysis (ORA) in WebGestalt 2019 [85, 86]. Next, we calculated the enrichment (compared to all genes with nonzero expression in 90% of samples) of adipose aware DE genes; unique cell-type marker genes for adipocytes, preadipocytes, and hepatocytes; and genes which were both DE and unique cell-type markers, respectively, using a hypergeometric test. We also identified transcription factors in the modules using PANTHER v16 [87]. We calculated the module membership of key genes identified in the functional enrichment tests using a Pearson correlation with the ME.

4.5 DE analysis of KOBS adipose and liver expression data

To identify genes differentially expressed (DE) between the KOBS participants with and without NAFLD, we performed case-control DE analysis on KOBS adipose and liver expression data for steatosis, fibrosis, and NASH, diagnosed by liver histology. In each analysis, the cases were patients with a nonzero grade for the liver histology phenotype being tested (n=158, 118, and 85 for steatosis, fibrosis, and NASH, respectively). The controls were patients with a grade of zero in all three liver histology phenotypes (n=87 for all tests).

To prepare for the DE analysis, we performed trimmed mean of M values (TMM) normalization on the adipose and liver bulk RNA-seq data using edgeR v3.32.1 [88–90]. To run the DE analysis, we then input these normalized expression values into the limma-voom

pipeline v3.46.0 [36–38], correcting for the same covariates that were regressed out in the WGCNA analysis. We assessed the significance of DE genes after Benjamini-Hochberg correction. After identifying the DE genes for steatosis, fibrosis, and NASH in adipose and liver tissue, we calculated the enrichment of cell-type marker genes in all DE gene lists using a hypergeometric test.

4.6 Filtering of adipose NAFLD DE genes for adipose-origin serum biomarker candidates

To identify adipose-origin serum biomarker candidates (SBCs) for NAFLD, we applied a filtering approach that focused on the adipose NAFLD DE genes. We started with the list of genes which were DE for any of the three liver histology traits (steatosis, fibrosis, or NASH) in the subcutaneous adipose tissue. Next, we removed the genes that were also DE for any of the same NAFLD traits in the liver. Next, we downloaded tissue-specific median transcripts per million (TPM) data from GTEx, and selected the genes that had both the median TPM > 30 in subcutaneous adipose tissue and whose ratio of subcutaneous adipose median TPM to liver median TPM was > 10. Finally, we selected the genes that encoded proteins secreted to serum, based on the Human Protein Atlas list of secreted proteins [39]. We designated the adipose NAFLD DE genes that satisfied all of these filters as SBCs.

To assess the relationship of the SBCs to each other, we correlated their adipose expression. First, we normalized the data by calculating the log-CPM of all SBCs. Then, we computed the Pearson correlation of every pairwise combination of SBCs using the R package Hmisc v4.6 [91], and assessed the significance of each correlation after Bonferroni correction.

4.7 Best subsets approach to identify key SBCs

To find the most effective subset of SBCs, we tested the proportion of variance in steatosis, fibrosis, and NASH explained by the adipose expression of different combinations of SBC, using the leaps algorithm. To normalize the data, we first calculated the adipose counts per million (CPM) of the SBCs, then performed an INT. Next, we used linear models in a best subsets analysis to test the variance in NAFLD traits explained by adipose SBC expression, using the regsubsets function from the R package leaps v3.1 [40]. This package uses an iterative algorithm to identify the best-fitting linear model with each number of genes included, ranging from a single variable to every variable provided.

For fibrosis and NASH, we tested all SBC genes DE for the target phenotype as possible inputs to the model. For steatosis, to identify genes involved in the early onset of NAFLD, we only tested SBC genes exclusively DE for steatosis, and not fibrosis or NASH. We included RNA-seq covariates (the same used in WGCNA and DE) in these analyses by regressing them out of the transformed CPMs before running leaps.

We identified the models that explained maximum variance in steatosis, fibrosis, and NASH using the Bayesian Information Criterion (BIC), and assessed the significance of these models with a permutation test ($B=100,000$). For each permutation, we selected a random set of adipose genes with nonzero expression in 90% of samples, equal to the number of genes in the best subset model chosen by leaps. We then used a linear model to test the variance in the phenotype being assessed that was explained by the adipose expression of those genes. The p-value for each SBC model was defined as the proportion of random permuted models whose r^2 value was greater than the SBC model.

4.8 SGBS Cell culture

Simone Golabi Behmel Syndrome (SGBS) pre-adipocyte cells [92] were maintained in DMEM/F-12 Nut media (Lonza # BE12-719F) with 4 μ g/ml Pantothenate (Sigma, # P-5155), 8 μ g/ml Biotin (Sigma # B-4639), 10% fetal bovine serum (FBS), 1% penicillin-streptomycin. These cells undergo complete differentiation into mature adipocytes in 14 days [93]. When pre-adipocytes reached confluence they were treated with serum free differentiation medium DMEM/F-12 supplemented with 4 μ g/ml Pantothenate, 8 μ g/ml Biotin, 1% penicillin-streptomycin, 2 μ mol/l rosiglitazone (Cayman Chemical # CAT 71740), 25 nmol/l dexamethasone (Sigma # D-4902), 0.5 mmol/l methylisobutylxanthine (Sigma # I5879), 0.1 μ mol/l cortisol (Sigma # H0888), 0.01 mg/ml transferrin (Sigma # T8158), 0.2 nmol/l triiodothyronin (Sigma # T6397), and 20 nmol/l human insulin (Sigma # I9278) for 7 days. This was followed with cell culture in adipogenic medium DMEM/F-12 supplemented with 4 μ g/ml Pantothenate, 8 μ g/ml Biotin, 1% penicillin-streptomycin, 0.1 μ mol/l cortisol, 0.01 mg/ml transferrin, 0.2 nmol/l triiodothyronine, and 20 nmol/l human insulin for an additional 7 days.

4.9 CCDC80 and SOD3 siRNA knockdown and sample collection for RNA-seq experiment

The cells were seeded in a 6-well plate at 1.6×10^6 cells per well. Once the cells reached 50% confluency, they were transfected with siRNA using lipofectamine RNAiMAX (Invitrogen) according to the manufacturer's instructions. Predesigned siRNAs from Thermo Fisher Scientific were used [scrambled (control) siRNA (30 nM) (ref no: 4390843), CCDC80 (60 nM) (ID: s45625), SOD3 (60 nM) (ID: s13272)].

During differentiation, the cells are devoid of serum and they stop dividing. This enables the cells to retain the siRNA transfection mix for up to 14 days, as previously shown [94]. In this study the cells were differentiated, and the samples were collected at different timepoints.

The cells were incubated with the transfection mix for 48h, after which the baseline samples were collected. The rest of the samples were treated with differentiation media (as described above), and collection was done at 24h, 4 days and 7 days.

4.10 siRNA knockdown RNA-seq library preparation

Cells were lysed and RNA was extracted using miRNeasy micro kit (Qiagen). Library samples were prepared using QuantSeq 3' mRNA-seq library prep kit FWD (Lexogen) according to the manufacturer's instructions, amplified for 18 cycles, and then sequenced with Illumina Next seq 500 for 75 cycles.

4.11 Alignment and quantification of siRNA knockdown

RNA-seq data

We aligned raw QuantSeq RNA-seq reads from the siRNA knockdown experiment to the GENCODE GRCh37 human reference genome and annotation v19 using STAR v2.5.2 [74]. We measured control, scrambled (control) siRNA, CCDC80 siRNA knockdown, and SOD3 siRNA knockdown conditions across the four differentiation time points, with 3-4 replicates per condition, resulting in a total of 59 samples. Before running the alignment, we first trimmed the raw reads with cutadapt v3.5, using a polyA sequence concatenated to the standard Illumina adapter as the trimming target. We used a 2-pass method to align the trimmed reads, which had an average read length of 83.5bp. After alignment, we verified the quality of our data using FastQC, based on statistics including sequence quality, GC content, and adapter content. Finally, we quantified gene expression using the Subread v1.6.2 package featureCounts, and selected only uniquely mapped reads for the expression data.

4.12 DE analysis of siRNA knockdown expression data

To identify genes DE between the CCDC80 and SOD3 knockdown samples and control samples, we performed DE analysis of the knockdown experiment expression data. First, we removed lowly expressed genes by selecting only those which had a total count of >10 summed across the samples within one group (control or knockdown). Next, we restricted the genes being tested for DE to SOD3, CCDC80, unique cell-type marker genes for adipocytes and preadipocytes, and adipogenesis pathway genes downloaded from WikiPathways WP236 [95] (n=492 genes tested). We also excluded the non-transfected control samples, resulting in a final sample size of 28 for both CCDC80 and SOD3.

We ran the limma-voom pipeline on the knockdown expression data in the same way as described for the KOBS DE analyses, except without including any covariates, thus comparing the knockdown samples to the scrambled (control) siRNA samples independently at each time point and for each knockdown condition (n=7 in all 8 tests). Our rationale for not including typical technical covariates of RNA-seq is that this was an *in vitro* cell-line experiment with isogenic replicates, in which the expression was assessed by performing QuantSeq 3' tag-based sequencing instead of the regular, highly dynamic bulk RNA-seq analysis.

To interpret the results of the DE tests, we analyzed the lists of DE genes at each time point. First, we verified that the knockdown was successful using the nominal p-values for CCDC80 and SOD3 in all timepoints from the corresponding experiments. Next, we overlapped the DE gene lists with the adipogenesis pathway genes from WikiPathways, and focused on those marked as transcription factors, especially SREBF1. Finally, we overlapped the lists of DE genes with SREBF1 binding targets downloaded from WikiPathways WP2706 and the TRANSFAC [96, 97] database.

REFERENCES

1. Younossi, Z. M. *et al.* Global epidemiology of nonalcoholic fatty liver disease—Meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology* **64**, 73–84 (2016).
2. Powell, E. E., Wong, V. W.-S. & Rinella, M. Non-alcoholic fatty liver disease. *Lancet* **397**, 2212–2224 (2021).
3. Wong, R. J. *et al.* Nonalcoholic steatohepatitis is the second leading etiology of liver disease among adults awaiting liver transplantation in the United States. *Gastroenterology* **148**, 547–555 (2015).
4. Albhaisi, S., Chowdhury, A. & Sanyal, A. J. Non-alcoholic fatty liver disease in lean individuals. *JHEP Rep.* **1**, 329–341 (2019).
5. Ahadi, M. *et al.* A review of non-alcoholic fatty liver disease in non-obese and lean individuals. *J. Gastroenterol. Hepatol.* **36**, 1497–1507 (2021).
6. Bilic-Curcic, I., Cigrovski Berkovic, M., Virovic-Jukic, L. & Mrzljak, A. Shifting perspectives - interplay between non-alcoholic fatty liver disease and insulin resistance in lean individuals. *World J. Hepatol.* **13**, 80–93 (2021).
7. Neuschwander-Tetri, B. A. Non-alcoholic fatty liver disease. *BMC Med.* **15**, 45 (2017).
8. Maurice, J. & Manousou, P. Non-alcoholic fatty liver disease. *Clin. Med.* **18**, 245–250 (2018).
9. Cobbina, E. & Akhlaghi, F. Non-alcoholic fatty liver disease (NAFLD) - pathogenesis, classification, and effect on drug metabolizing enzymes and transporters. *Drug Metab. Rev.* **49**, 197–211 (2017).
10. Friedman, S. L., Neuschwander-Tetri, B. A., Rinella, M. & Sanyal, A. J. Mechanisms of NAFLD development and therapeutic strategies. *Nat. Med.* **24**, 908–922 (2018).

11. Azzu, V., Vacca, M., Virtue, S., Allison, M. & Vidal-Puig, A. Adipose Tissue-Liver Cross Talk in the Control of Whole-Body Metabolism: Implications in Nonalcoholic Fatty Liver Disease. *Gastroenterology* **158**, 1899–1912 (2020).
12. Chang, M.-L., Yang, Z. & Yang, S.-S. Roles of Adipokines in Digestive Diseases: Markers of Inflammation, Metabolic Alteration and Disease Progression. *Int. J. Mol. Sci.* **21**, 8308 (2020).
13. Donnelly, K. L. *et al.* Sources of fatty acids stored in liver and secreted via lipoproteins in patients with nonalcoholic fatty liver disease. *J. Clin. Invest.* **115**, 1343–1351 (2005).
14. Lomonaco, R. *et al.* Effect of adipose tissue insulin resistance on metabolic parameters and liver histology in obese patients with nonalcoholic fatty liver disease. *Hepatology* **55**, 1389–1397 (2012).
15. Matsumoto, M. *et al.* An improved mouse model that rapidly develops fibrosis in non-alcoholic steatohepatitis. *Int. J. Exp. Pathol.* **94**, 93–103 (2013).
16. Sookoian, S. & Pirola, C. J. Genetic predisposition in nonalcoholic fatty liver disease. *Clin. Mol. Hepatol.* **23**, 1–12 (2017).
17. Wong, V. W.-S., Adams, L. A., de Lédinghen, V., Wong, G. L.-H. & Sookoian, S. Noninvasive biomarkers in NAFLD and NASH — current progress and future promise. *Nat Rev. Gastroenterol. Hepatol.* **15**, 461–478 (2018).
18. Davison, B. A. *et al.* Suboptimal reliability of liver biopsy evaluation has implications for randomized clinical trials. *J. Hepatol.* **73**, 1322–1332 (2020).
19. Ratziu, V. *et al.* Sampling Variability of Liver Biopsy in Nonalcoholic Fatty Liver Disease. *Gastroenterology* **128**, 1898–1906 (2005).
20. Brill, F. *et al.* Clinical value of liver ultrasound for the diagnosis of nonalcoholic fatty liver disease in overweight and obese patients. *Liver Int.* **35**, 2139–2146 (2015).

21. Caussy, C., Reeder, S. B., Sirlin, C. B. & Loomba, R. Noninvasive, Quantitative Assessment of Liver Fat by MRI-PDFF as an Endpoint in NASH Trials. *Hepatology* **68**, 763–772 (2018).
22. Park, C. C. *et al.* Magnetic Resonance Elastography vs Transient Elastography in Detection of Fibrosis and Noninvasive Measurement of Steatosis in Patients With Biopsy-Proven Nonalcoholic Fatty Liver Disease. *Gastroenterology* **152**, 598–607.e2 (2017).
23. Angulo, P. *et al.* The NAFLD Fibrosis Score: A Noninvasive System That Identifies Liver Fibrosis in Patients with NAFLD. *Hepatology* **45**, 846–854 (2007).
24. Sterling, R. K. *et al.* Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. *Hepatology* **43**, 1317–1325 (2006).
25. Vali, Y. *et al.* Enhanced liver fibrosis test for the non-invasive diagnosis of fibrosis in patients with NAFLD: A systematic review and meta-analysis. *J. Hepatol.* **73**, 252–262 (2020).
26. Miao, Z. *et al.* Identification of 90 NAFLD GWAS loci and establishment of NAFLD PRS and causal role of NAFLD in coronary artery disease. *HGG Adv.* **3**, 100056 (2022).
27. Daniels, S. J. *et al.* ADAPT: An Algorithm Incorporating PRO-C3 Accurately Identifies Patients With NAFLD and Advanced Fibrosis. *Hepatology* **69**, 1075–1086 (2019).
28. Newsome, P. N. *et al.* FibroScan-AST (FAST) score for the non-invasive identification of patients with non-alcoholic steatohepatitis with significant activity and fibrosis: a prospective derivation and global validation study. *Lancet Gastroenterol. Hepatol.* **5**, 362–373 (2020).
29. Harrison, S. A. *et al.* A blood-based biomarker panel (NIS4) for non-invasive diagnosis of non-alcoholic steatohepatitis and liver fibrosis: a prospective derivation and global validation study. *Lancet Gastroenterol. Hepatol.* **5**, 970–985 (2020).
30. Ayala-Sumuano, J.-T. *et al.* Srebf1a is a key regulator of transcriptional control for adipogenesis. *Sci. Rep.* **1**, 178 (2011).

31. Wang, X *et al.* Nuclear protein that binds sterol regulatory element of low density lipoprotein receptor promoter. II. Purification and characterization. *J. Biol. Chem.* **268**, 14497–14504 (1993).
32. Münzberg, H. & Morrison, C. D. Structure, production and signaling of leptin. *Metabolism* **64**, 13–23 (2015).
33. R Core Team. R: A Language and Environment for Statistical Computing. (2021).
34. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9** (2008).
35. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **4**, 17 (2005).
36. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
37. Phipson, B., Lee, S., Majewski, I. J., Alexander, W. S. & Smyth, G. K. Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. *Ann. Appl. Stat.* **10**, 946–963 (2016).
38. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
39. Uhlén, M. *et al.* The human secretome. *Sci. Signal.* **12**, eaaz0274 (2019).
40. Thomas Lumley based on Fortran code by Alan Miller. leaps: Regression Subset Selection. (2020).
41. Castera, L., Friedrich-Rust, M. & Loomba, R. Noninvasive Assessment of Liver Disease in Patients With Nonalcoholic Fatty Liver Disease. *Gastroenterology* **156**, 1264–1281.e4 (2019).
42. Piazzolla, V. A. & Mangia, A. Noninvasive Diagnosis of NAFLD and NASH. *Cells* **9**, 1005 (2020).

43. Kozumi, K. *et al.* Transcriptomics identify thrombospondin-2 as a biomarker for NASH and advanced liver fibrosis. *Hepatology* **74**, 2452–2466 (2021).
44. Sheldon, R. D. *et al.* Transcriptomic differences in intra-abdominal adipose tissue in extremely obese adolescents with different stages of NAFLD. *Physiol. Genomics* **48**, 897–911 (2016).
45. Osorio-Conles, Ó. *et al.* A Distinctive NAFLD Signature in Adipose Tissue from Women with Severe Obesity. *Int. J. Mol. Sci.* **22**, 10541 (2021).
46. Unamuno, X. *et al.* Adipokine dysregulation and adipose tissue inflammation in human obesity. *Eur. J. Clin. Invest.* **48**, e12997 (2018).
47. Fasshauer, M. & Blüher, M. Adipokines in health and disease. *Trends Pharmacol. Sci.* **36**, 461–470 (2015).
48. Wu, C.-C. *et al.* Candidate serological biomarkers for cancer identified from the secretomes of 23 cancer cell lines and the human protein atlas. *Mol. Cell. Proteomics* **9**, 1100–1117 (2010).
49. Johnson, K. *et al.* Increased serum miR-193a-5p during non-alcoholic fatty liver disease progression: Diagnostic and mechanistic relevance. *JHEP Rep.* **4**, 100409 (2022).
50. Gong, D. *et al.* Coiled-coil domain-containing 80 accelerates atherosclerosis development through decreasing lipoprotein lipase expression via ERK1/2 phosphorylation and TET2 expression. *Eur. J. Pharmacol.* **843**, 177–189 (2019).
51. Grill, J. I. *et al.* Loss of DRO1/CCDC80 results in obesity and promotes adipocyte differentiation. *Mol. Cell. Endocrinol.* **439**, 286–296 (2017).
52. Osorio-Conles, O *et al.* Adipose tissue and serum CCDC80 in obesity and its association with related metabolic disease. *Mol. Med.* **23**, 225–234 (2017).
53. Gao, D. *et al.* SOD3 Is Secreted by Adipocytes and Mitigates High-Fat Diet-Induced Obesity, Inflammation, and Insulin Resistance. *Antioxid. Redox Signal.* **32**, 193–212 (2020).

54. Kim, J. B., Wright, H. M., Wright, M. & Spiegelman, B. M. ADD1/SREBP1 activates PPAR γ through the production of endogenous ligand. *Proc. Natl. Acad. Sci. USA* **95**, 4333–4337 (1998).
55. Tontonoz, P., Hu, E. & Spiegelman, B. M. Stimulation of adipogenesis in fibroblasts by PPAR γ 2, a lipid-activated transcription factor. *Cell* **79**, 1147–1156 (1994).
56. Tang, Q. Q. & Lane, M. D. Activation and centromeric localization of CCAAT/enhancer-binding proteins during the mitotic clonal expansion of adipocyte differentiation. *Genes Dev.* **13**, 2231–2241 (1999).
57. Kim, J., Kim, H. & Son, C. Tissue-Specific Profiling of Oxidative Stress-Associated Transcriptome in a Healthy Mouse Model. *Int. J. Mol. Sci.* **19**, 3174 (2018).
58. Gongora, M. C. *et al.* Loss of extracellular superoxide dismutase leads to acute lung damage in the presence of ambient air: a potential mechanism underlying adult respiratory distress syndrome. *Am. J. Pathol.* **173**, 915–926 (2008).
59. Delli Bovi, A. P. *et al.* Oxidative Stress in Non-alcoholic Fatty Liver Disease. An Updated Mini Review. *Front. Med.* **8**, 595371 (2021).
60. Cui, R, Gao, M, Qu, S & Liu, D. Overexpression of superoxide dismutase 3 gene blocks high-fat diet-induced obesity, fatty liver and insulin resistance. *Gene Ther.* **21**, 840–848 (2014).
61. Obradovic, M. *et al.* Leptin and Obesity: Role and Clinical Implication. *Front. Endocrinol.* **12**, 585887 (2021).
62. Crujeiras, A. B. *et al.* Leptin resistance in obesity: An epigenetic landscape. *Life Sci.* **140**, 57–63 (2015).
63. De Luis, D. A., Perez Castrillón, J. L. & Dueñas, A. Leptin and obesity. *Minerva Med.* **100**, 229–236 (2009).

64. Soyama, H. *et al.* Rapid decrease in serum VEGF-A levels may be a worse prognostic biomarker for patients with platinum-resistant recurrent ovarian cancer treated with bevacizumab and gemcitabine. *Cancer Chemother. Pharmacol.* **85**, 941–947 (2020).
65. Crowley, R. K. *et al.* SFRP2 Is Associated with Increased Adiposity and VEGF Expression. *PLoS One* **11**, e0163777 (2016).
66. Cymbaluk-Płoska, A. *et al.* Suitability assessment of baseline concentration of MMP3, TIMP3, HE4 and CA125 in the serum of patients with ovarian cancer. *J. Ovarian Res.* **11**, 1 (2018).
67. Karsli-Ceppioglu, S *et al.* Association of Genetic Polymorphisms in the Gene with Coronary Artery Disease and Serum MGP Levels. *Balkan J. Med. Genet.* **22**, 43–50 (2019).
68. Baez-Duarte, B. G. *et al.* Glutathione peroxidase 3 serum levels and GPX3 gene polymorphisms in subjects with metabolic syndrome. *Arch. Med. Res.* **45**, 375–382 (2014).
69. Lao, O. *et al.* Correlation between genetic and geographic structure in Europe. *Curr. Biol.* **18**, 1241–1248 (2008).
70. Kolk, B. W. v. d. *et al.* Differential Mitochondrial Gene Expression in Adipose Tissue Following Weight Loss Induced by Diet or Bariatric Surgery. *J. Clin. Endocrinol. Metab.* **106**, 1312–1324 (2021).
71. Pihlajamäki, J. *et al.* Cholesterol absorption decreases after Roux-en-Y gastric bypass but not after gastric banding. *Metabolism* **59**, 866–872 (2010).
72. Benhammou, J. N. *et al.* Novel Lipid Long Intervening Noncoding RNA, Oligodendrocyte Maturation-Associated Long Intergenic Noncoding RNA, Regulates the Liver Steatosis Gene Stearoyl-Coenzyme A Desaturase As an Enhancer RNA. *Hepatol Commun.* **3**, 1356–1372 (2019).

73. Pan, D. Z. *et al.* Integration of human adipocyte chromosomal interactions with adipose gene expression prioritizes obesity-related genes from GWAS. *Nat. Commun.* **9**, 1512 (2018).
74. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
75. Broad Institute. Picard. (2022).
76. Pan, D. Z. *et al.* Identification of TBX15 as an adipose master trans regulator of abdominal obesity genes. *Genome Med.* **13**, 123 (2021).
77. Alvarez, M. *et al.* Enhancing droplet-based single-nucleus RNA-seq resolution using the semi-supervised machine learning classifier DIEM. *Sci. Rep.* **10**, 11019 (2020).
78. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902.e21 (2019).
79. Aran, D. *et al.* Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* **20**, 163–172 (2019).
80. Rao, S. *et al.* β 2-spectrin (SPTBN1) as a therapeutic target for diet-induced liver disease and preventing cancer development. *Sci. Transl. Med.* **13**, eabk2267 (2021).
81. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer New York, NY, 2016).
82. Wickham, H. *et al.* Welcome to the tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
83. Lappalainen, T. *et al.* Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**, 506–511 (2013).
84. Mateus, I. D. *et al.* Dual RNA-seq reveals large-scale non-conserved genotype x genotype specific genetic reprogramming and molecular crosstalk in the mycorrhizal symbiosis. *ISME J.* **13**, 1226–1238 (2019).

85. Zhang, B., Kirov, S. & Snoddy, J. WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res.* **33**, W741–8 (2005).
86. Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z. & Zhang, B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* **47**, W199–W205 (2019).
87. Mi, H. *et al.* PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res.* **49**, D394–D403 (2021).
88. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
89. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research* **40**, 4288–4297 (2012).
90. Chen, Y., Lun, A. T. L. & Smyth, G. K. From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Res.* **5**, 1438 (2016).
91. Frank E. Harrell Jr. Hmisc: Harrell Miscellaneous. (2021).
92. Wabitsch, M *et al.* Characterization of a human preadipocyte cell strain with high capacity for adipose differentiation. *Int. J. Obes. Relat. Metab. Disord.* **25**, 8–15 (2001).
93. Fischer-Posovszky, P., Newell, F. S., Wabitsch, M. & Tornqvist, H. E. Human SGBS Cells – a Unique Tool for Studies of Human Fat Cell Biology. *Obesity Facts* **1**, 184–189 (2008).
94. Mai, Y. *et al.* BMP and activin membrane-bound inhibitor (BAMBI) inhibits the adipogenesis of porcine preadipocytes through Wnt/ β -catenin signaling pathway. *Biochem. Cell Biol.* **92**, 172–182 (2014).

95. Martens, M. *et al.* WikiPathways: connecting communities. *Nucleic Acids Res.* **49**, D613–D621 (2021).
96. Matys, V *et al.* TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res.* **31**, 374–378 (2003).
97. Matys, V *et al.* TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.* **34**, D108–10 (2006).