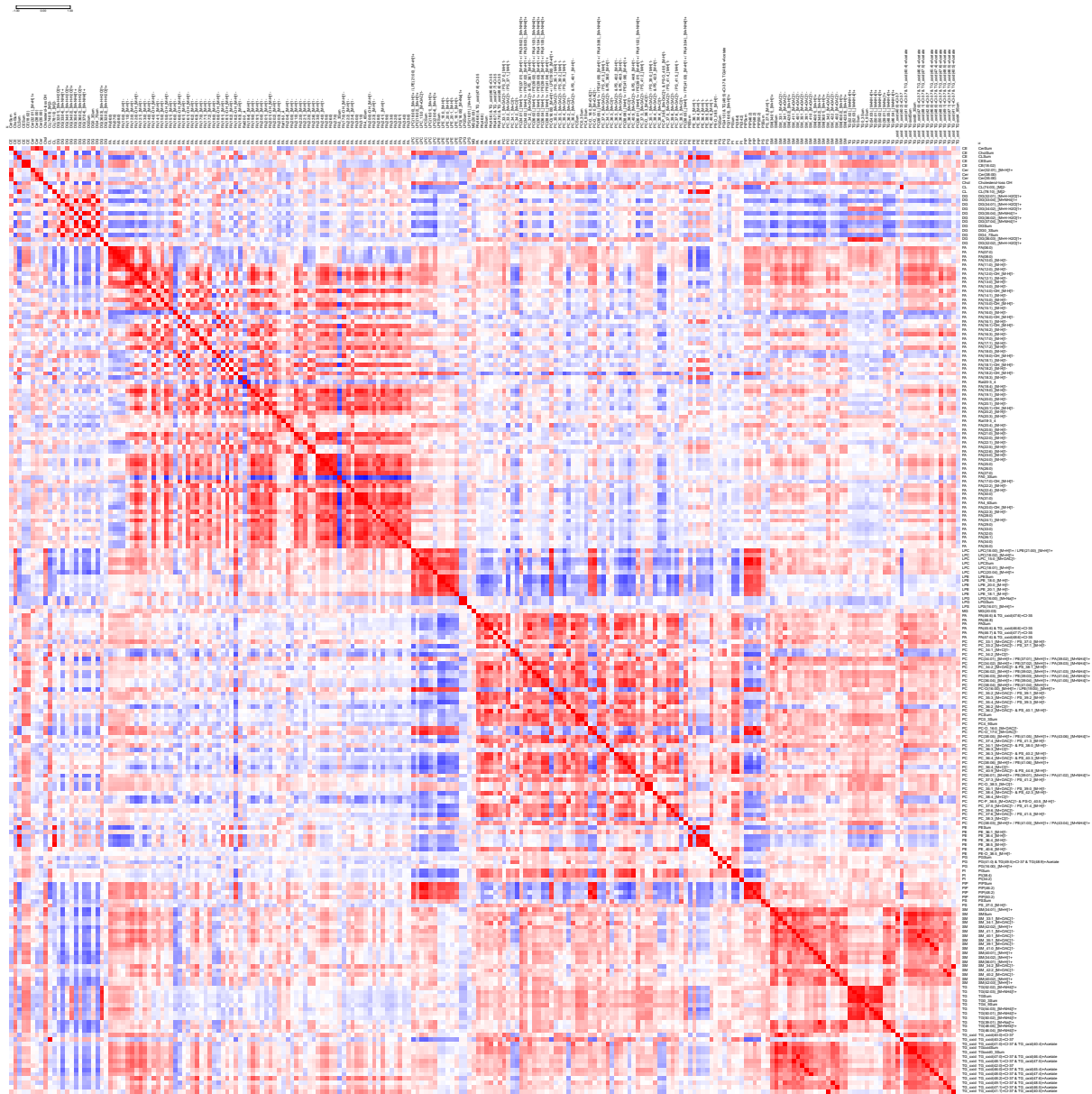
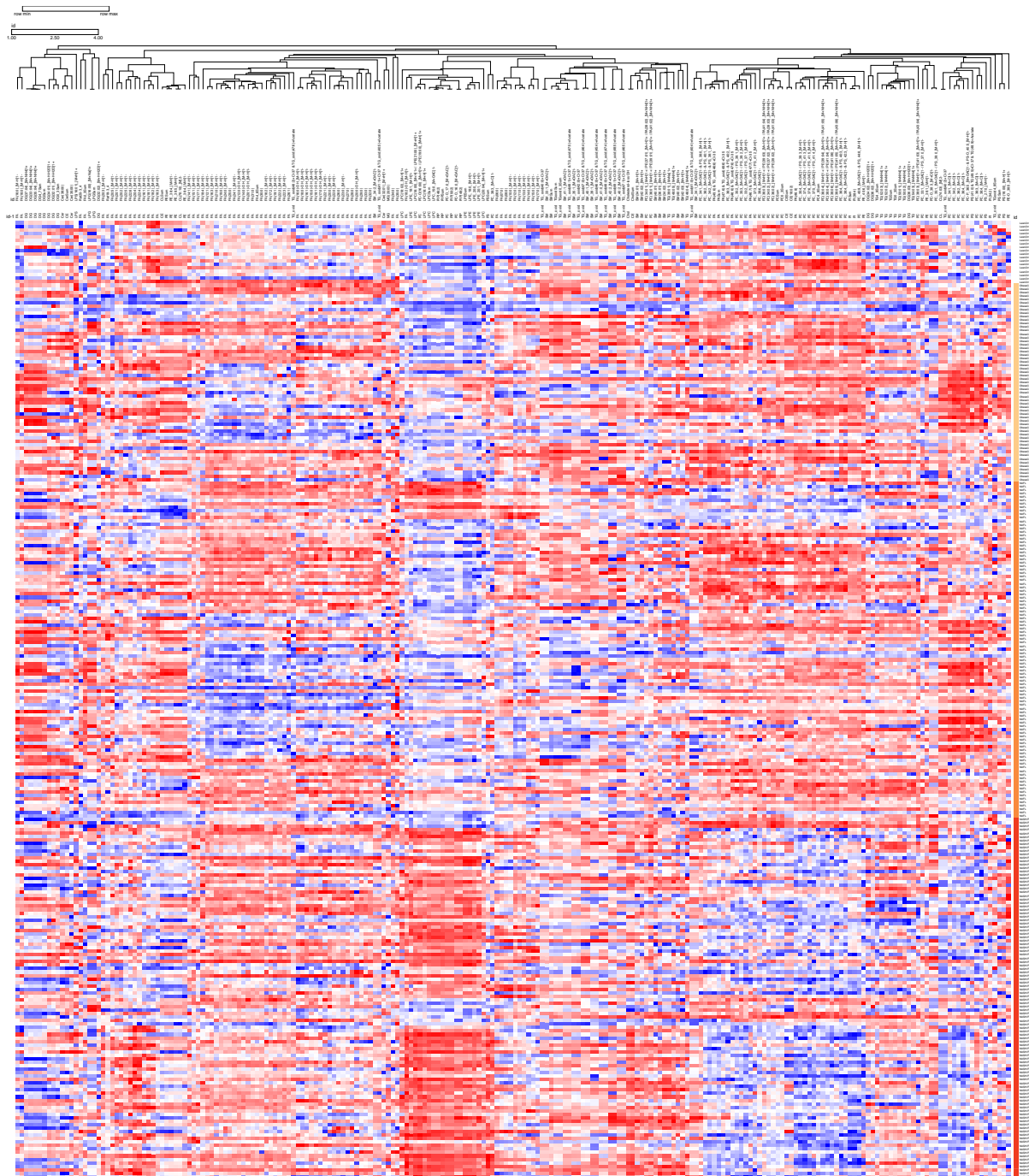


Plasma lipidomics distinguishes NASH and fibrosis from simple fatty liver in children

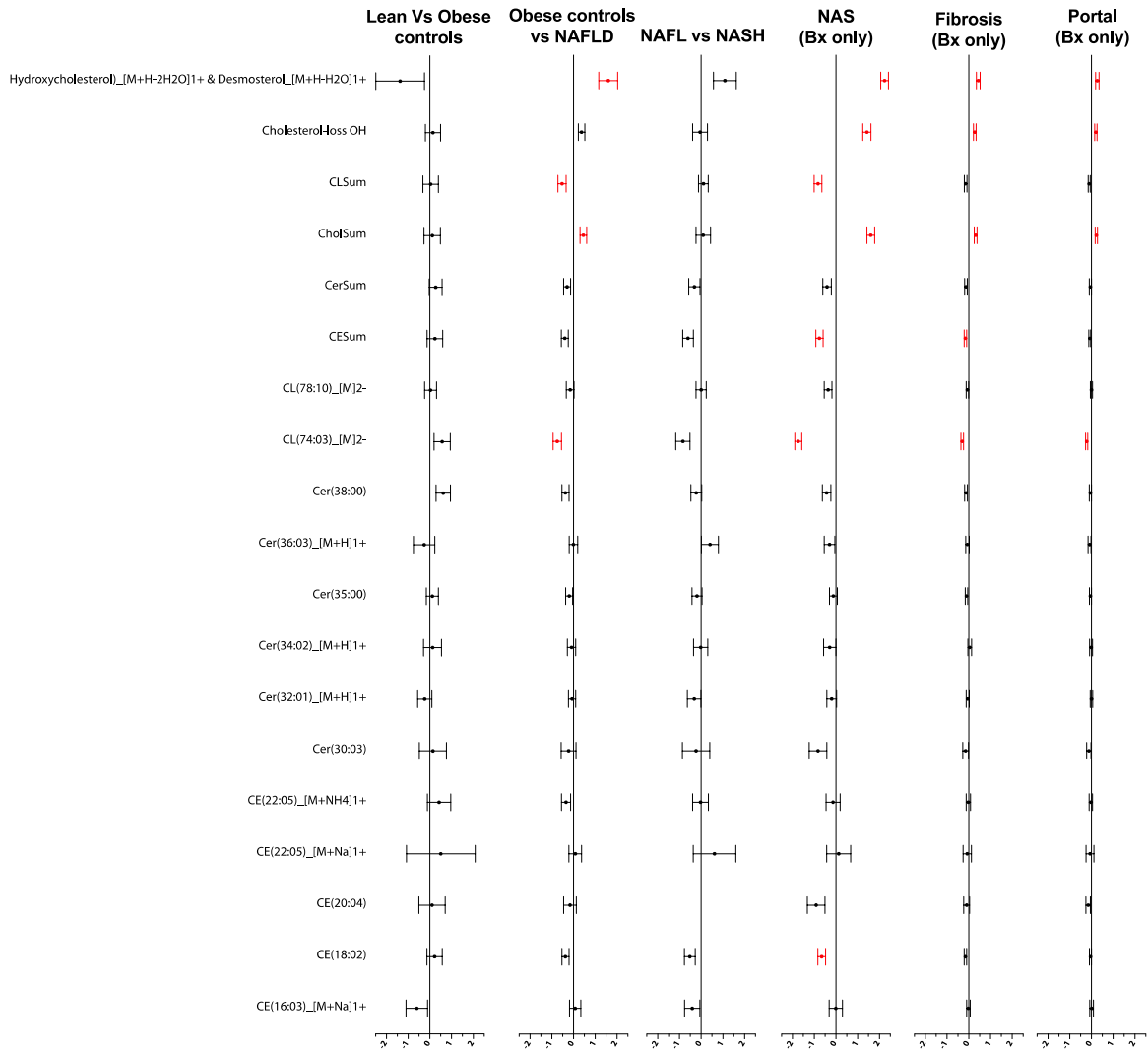
Supplementary Figures



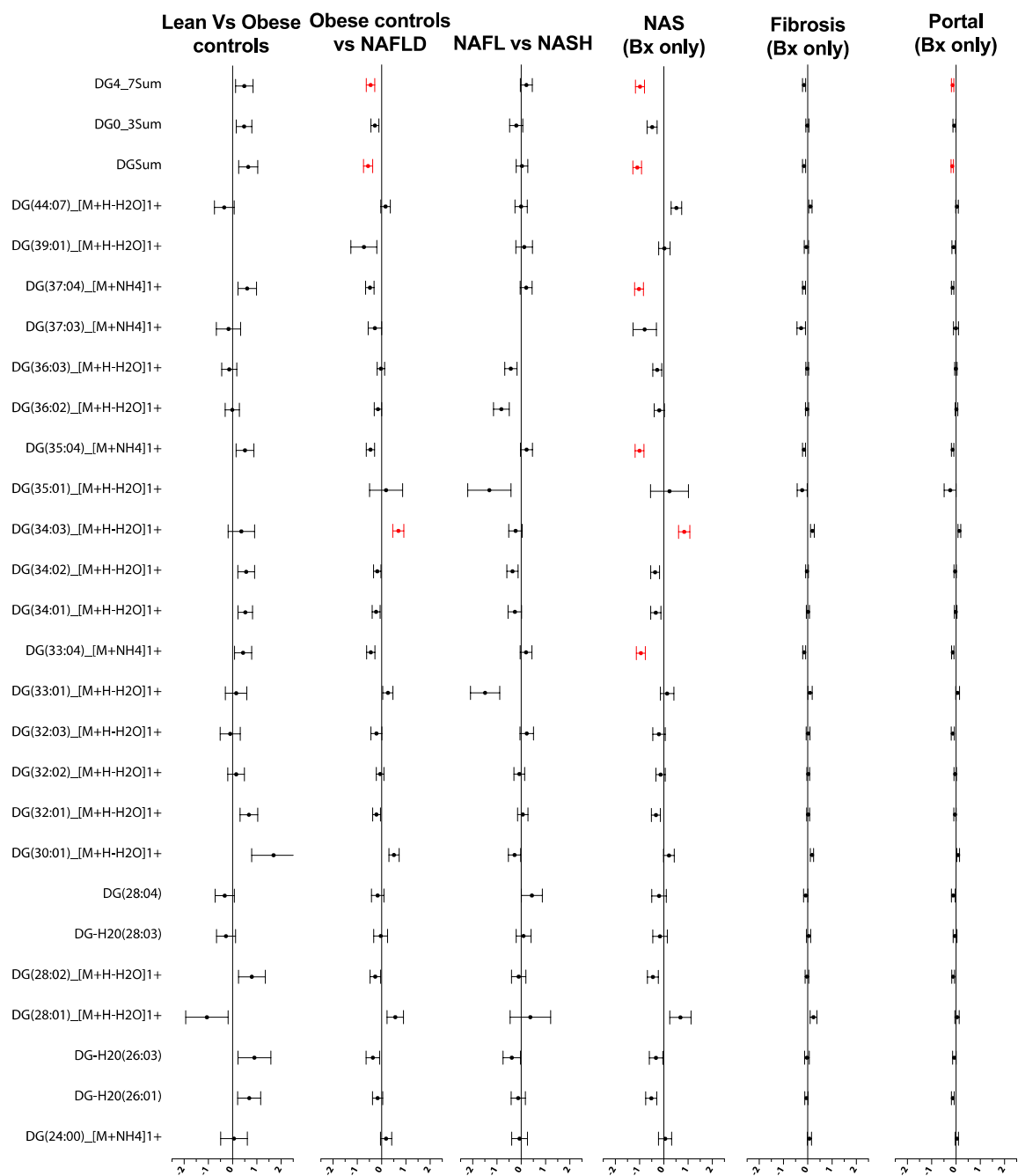
Supplementary Figure 1. Correlation matrix for lipid species identified on plasma lipidomics. Lipids are arranged alphabetically by class (or sub-class), for example lysophosphatidylcholines, then lysophosphoglycerols etc. Red indicates stronger positive correlation, blue indicates stronger negative correlation, white indicates no correlation.



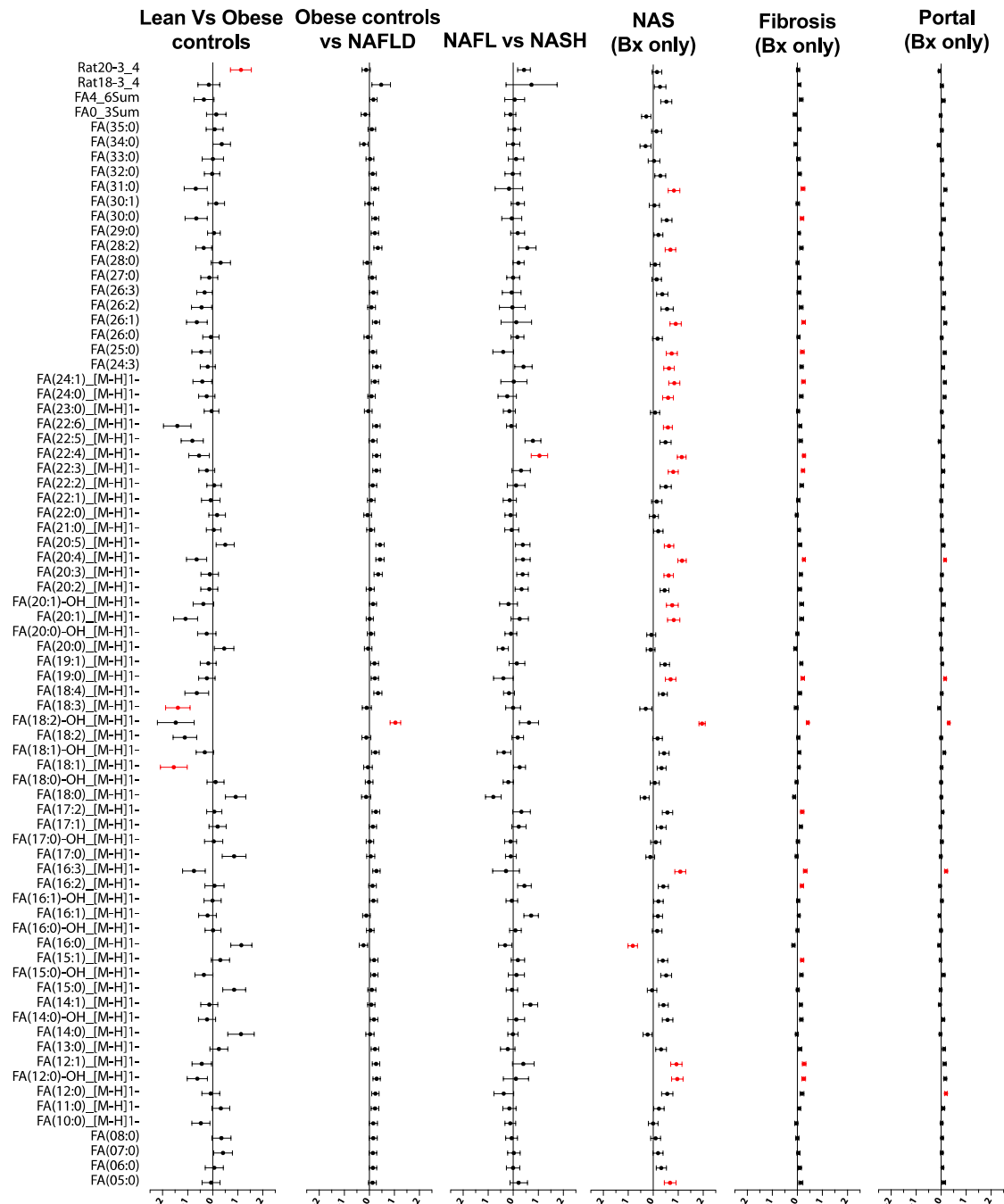
Supplementary Figure 2. Heatmap showing the relative abundances of lipid species across all participants. Lipids are shown on the x-axis and arranged by hierarchical clustering. participants are shown on the y-axis and are arranged by diagnosis, from top to bottom: lean control (n=19), obese control (n=65), non-alcoholic fatty liver (n=105), non-alcoholic steatohepatitis or fibrosis (n=113). The abundance of each lipid is standardised to mean = 0 and standard deviation = 1. Red represents a relatively higher level of each lipid and blue represents a lower level.



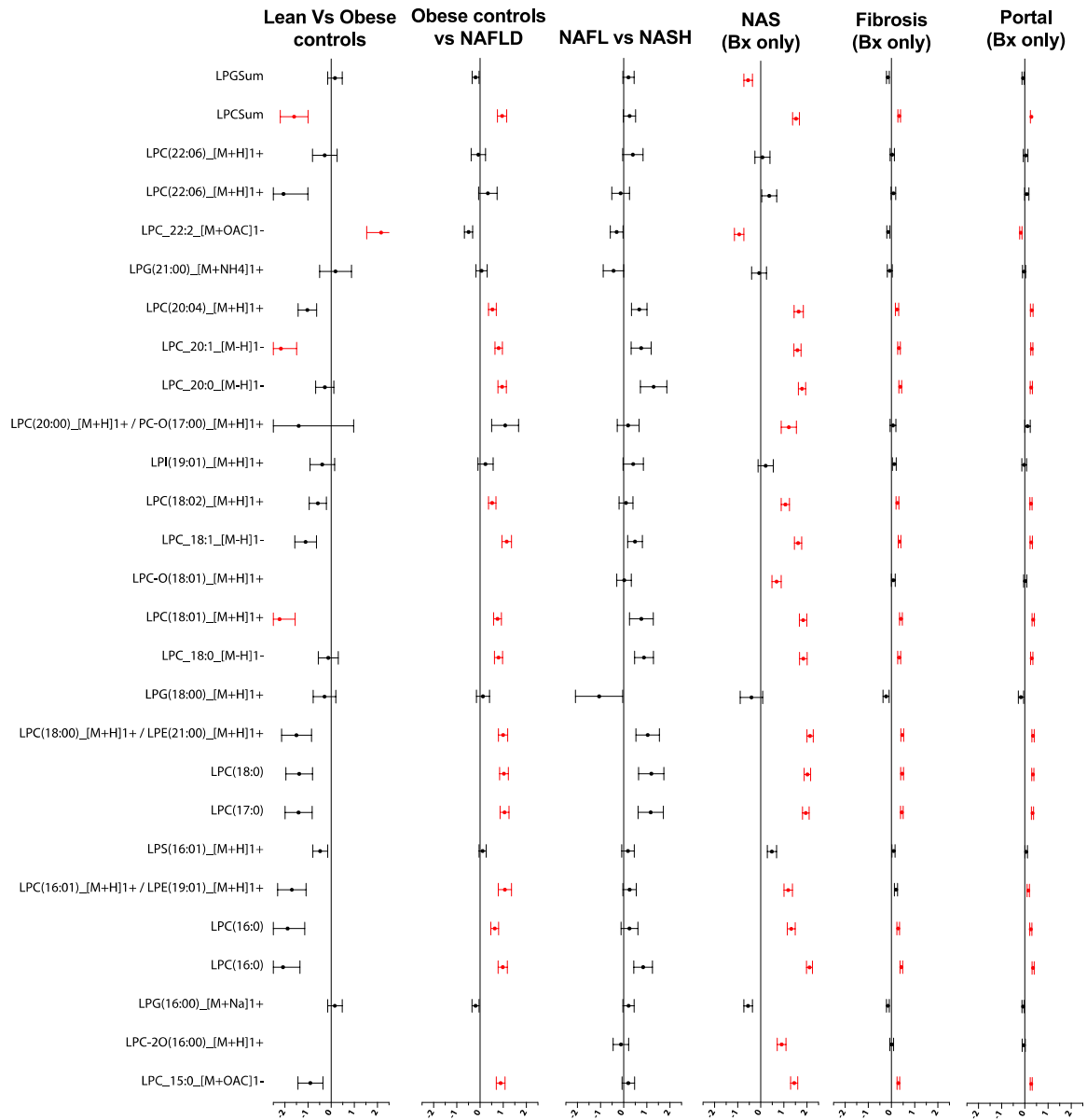
Supplementary Figure 3. Forest plots with results from regression analyses for cholesterol, ceramides, and cardiolipins. Each data point shows the beta with standard error, where beta represents the log-odds for categorical outcomes (lean vs obese controls, obese controls vs NAFLD, and NAFL vs NASH) or the unit change in histology score (NAS, fibrosis, or portal inflammation) for each standard deviation change in lipid. Data points shown in red indicate a significant association after adjusting for multiple comparisons. All regressions were adjusted for age and sex. Bx, biopsy; CE, cholesterol ester; Cer, ceramide; Chol, cholesterol; CL, cardiolipin.



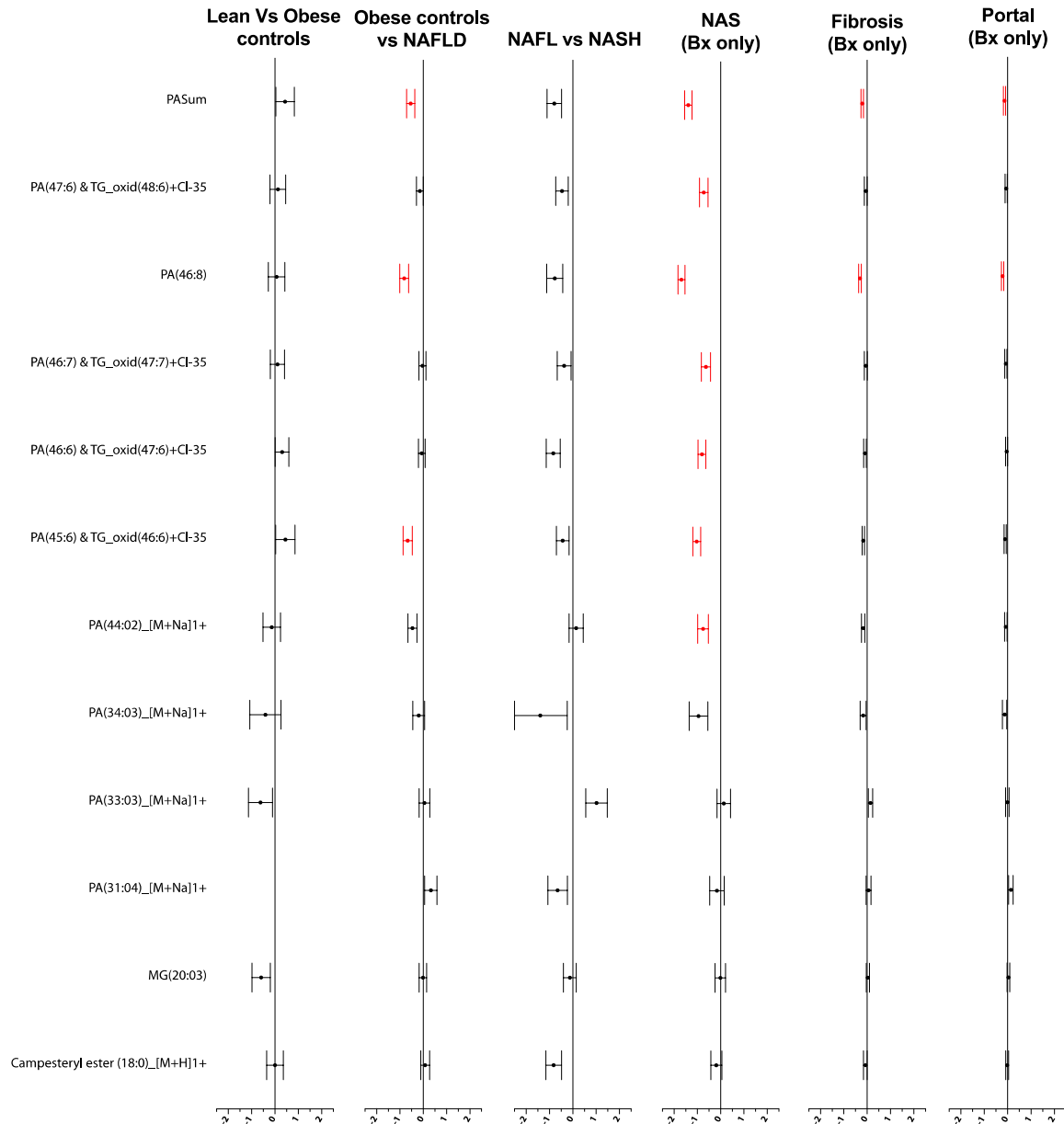
Supplementary Figure 4. Forest plots with results from regression analyses for diglycerols (DG). Each data point shows the beta with standard error, where beta represents the log-odds for categorical outcomes (lean vs obese controls, obese controls vs NAFLD, and NAFL vs NASH) or the unit change in histology score (NAS, fibrosis, or portal inflammation) for each standard deviation change in lipid. Data points shown in red indicate a significant association after adjusting for multiple comparisons. All regressions were adjusted for age and sex. 'DG0_3Sum' and 'DG4_7Sum' represent the total (sum) of all DG species with 0-3 and 4-7 double-bonds, respectively.



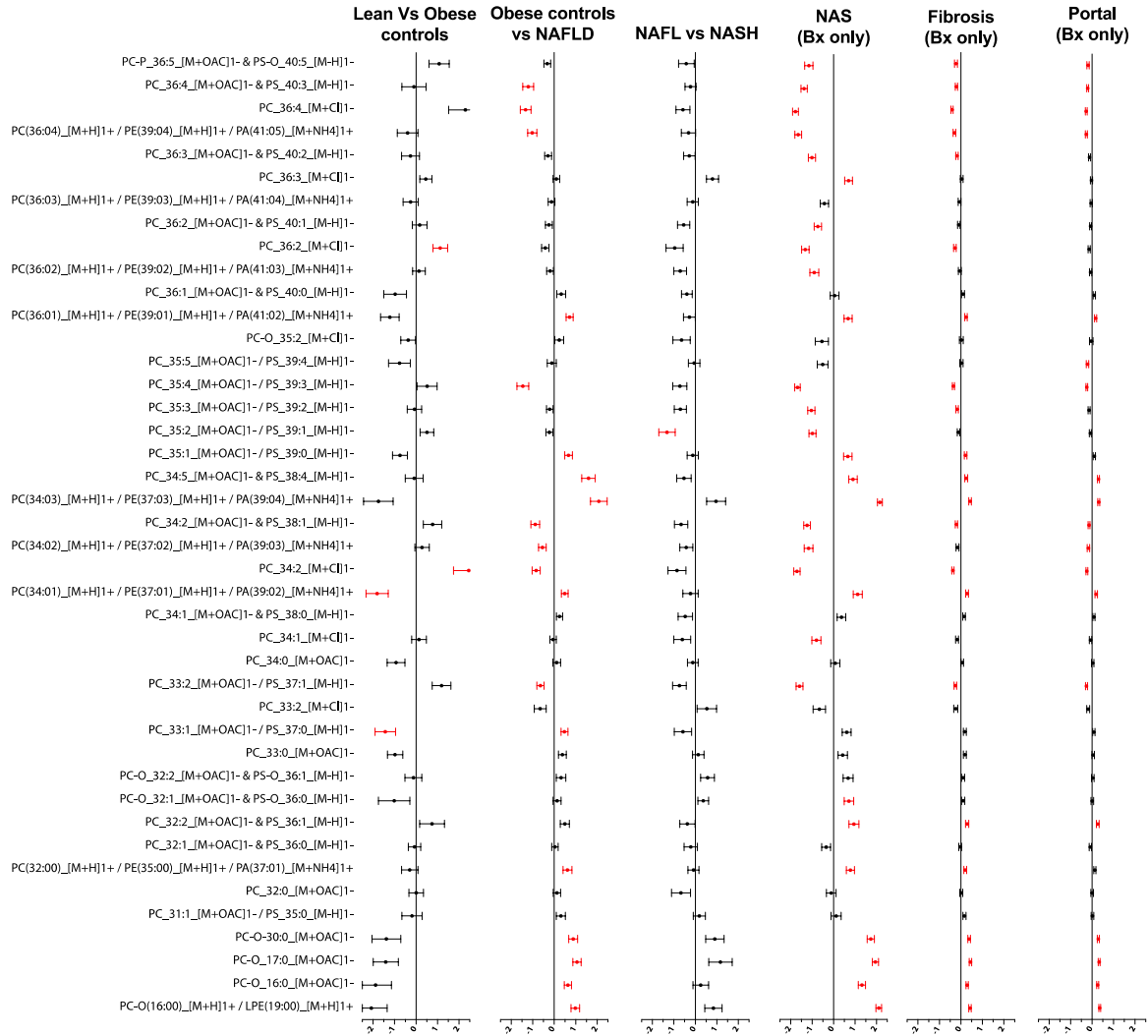
Supplementary Figure 5. Forest plots with results from regression analyses for fatty acids (FA) derived from the phospholipid fraction of plasma. Each data point shows the beta with standard error, where beta represents the log-odds for categorical outcomes (lean vs obese controls, obese controls vs NAFLD, and NAFL vs NASH) or the unit change in histology score (NAS, fibrosis, or portal inflammation) for each standard deviation change in lipid. Data points shown in red indicate a significant association after adjusting for multiple comparisons. All regressions were adjusted for age and sex.



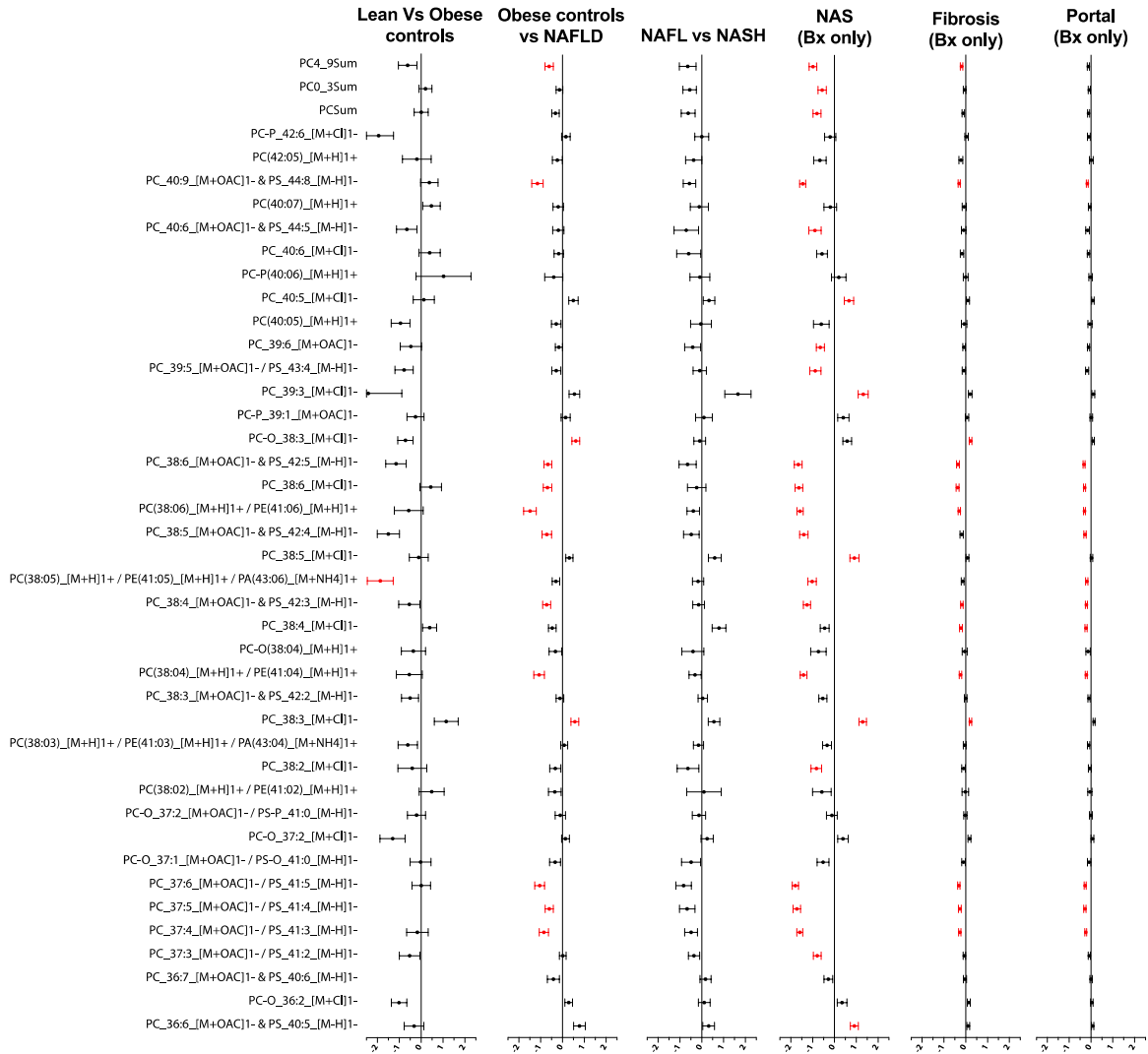
Supplementary Figure 6. Forest plots with results from regression analyses for lysophospholipids. Each data point shows the beta with standard error, where beta represents the log-odds for categorical outcomes (lean vs obese controls, obese controls vs NAFLD, and NAFL vs NASH) or the unit change in histology score (NAS, fibrosis, or portal inflammation) for each standard deviation change in lipid. Data points shown in red indicate a significant association after adjusting for multiple comparisons. All regressions were adjusted for age and sex. LysoPC, lysophosphatidylcholine; lysoPG, lysophosphoglycerol; lysoPS, lysophosphatidylserine.



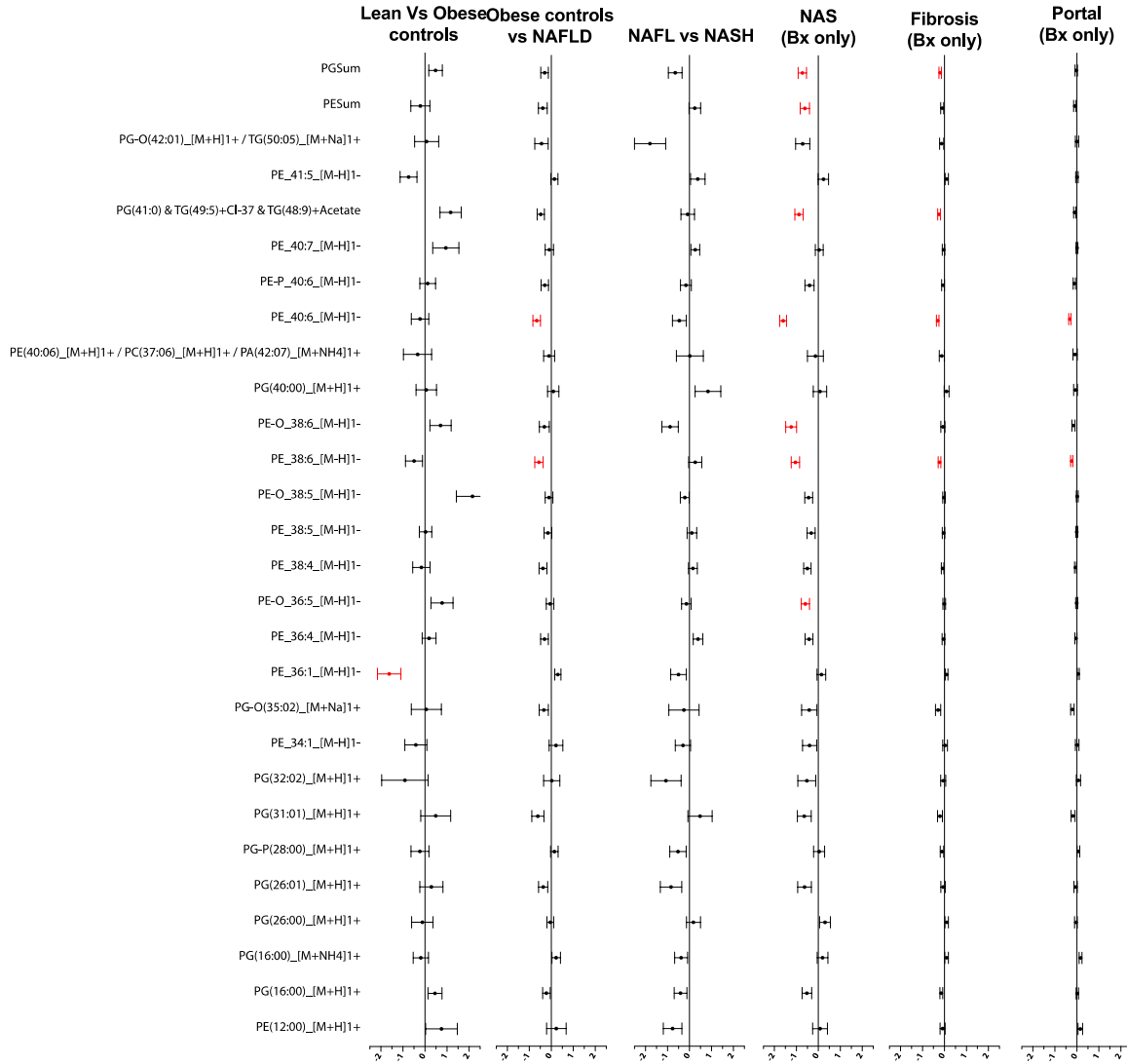
Supplementary Figure 7. Forest plots with results from regression analyses for phosphatidic acid (PA) species. Each data point shows the beta with standard error, where beta represents the log-odds for categorical outcomes (lean vs obese controls, obese controls vs NAFLD, and NAFL vs NASH) or the unit change in histology score (NAS, fibrosis, or portal inflammation) for each standard deviation change in lipid. Data points shown in red indicate a significant association after adjusting for multiple comparisons. All regressions were adjusted for age and sex.



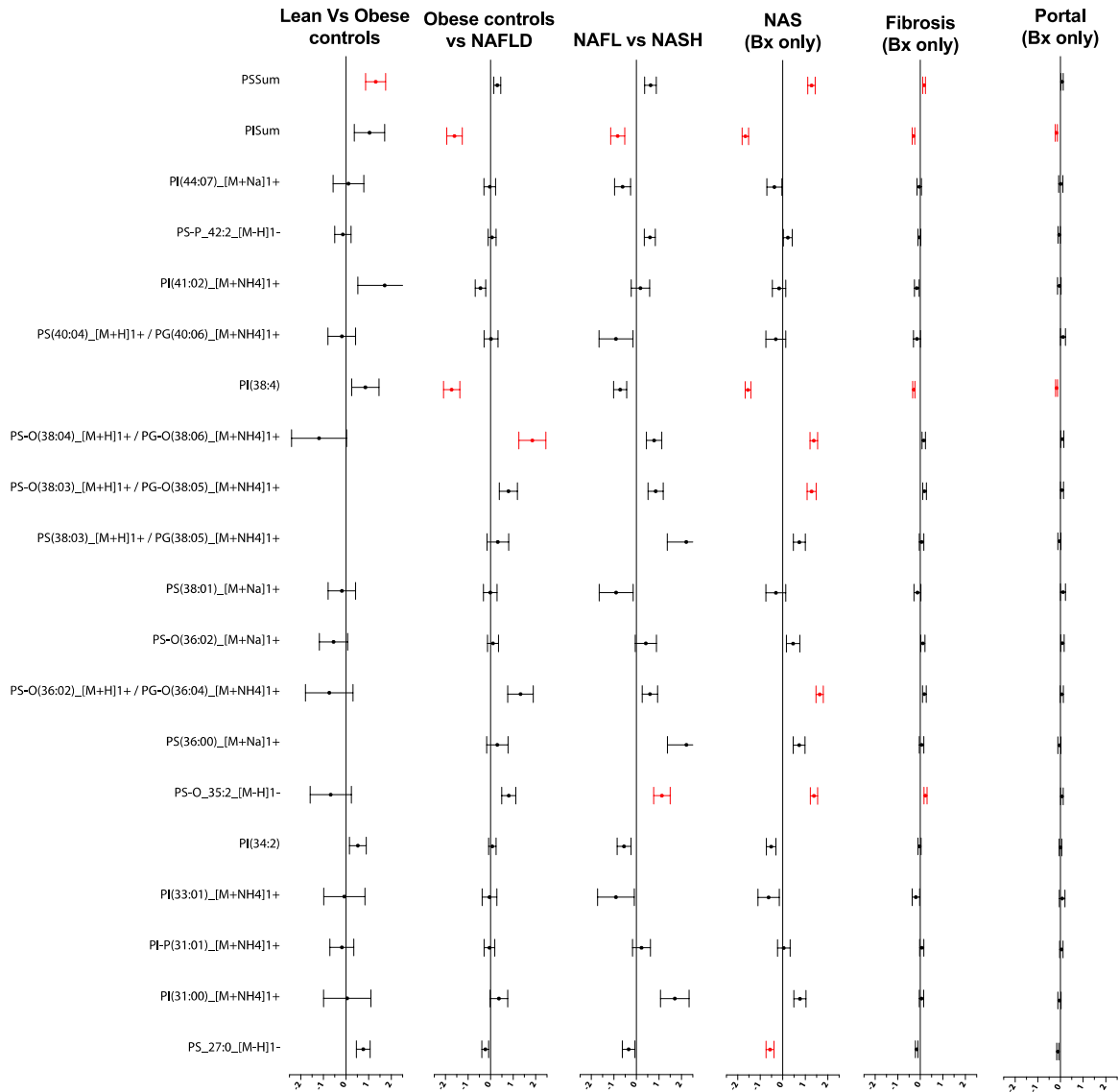
Supplementary Figure 8. Forest plots with results from regression analyses for phosphatidylcholine (PC) species (part 1). Each data point shows the beta with standard error, where beta represents the log-odds for categorical outcomes (lean vs obese controls, obese controls vs NAFLD, and NAFL vs NASH) or the unit change in histology score (NAS, fibrosis, or portal inflammation) for each standard deviation change in lipid. Data points shown in red indicate a significant association after adjusting for multiple comparisons. All regressions were adjusted for age and sex.



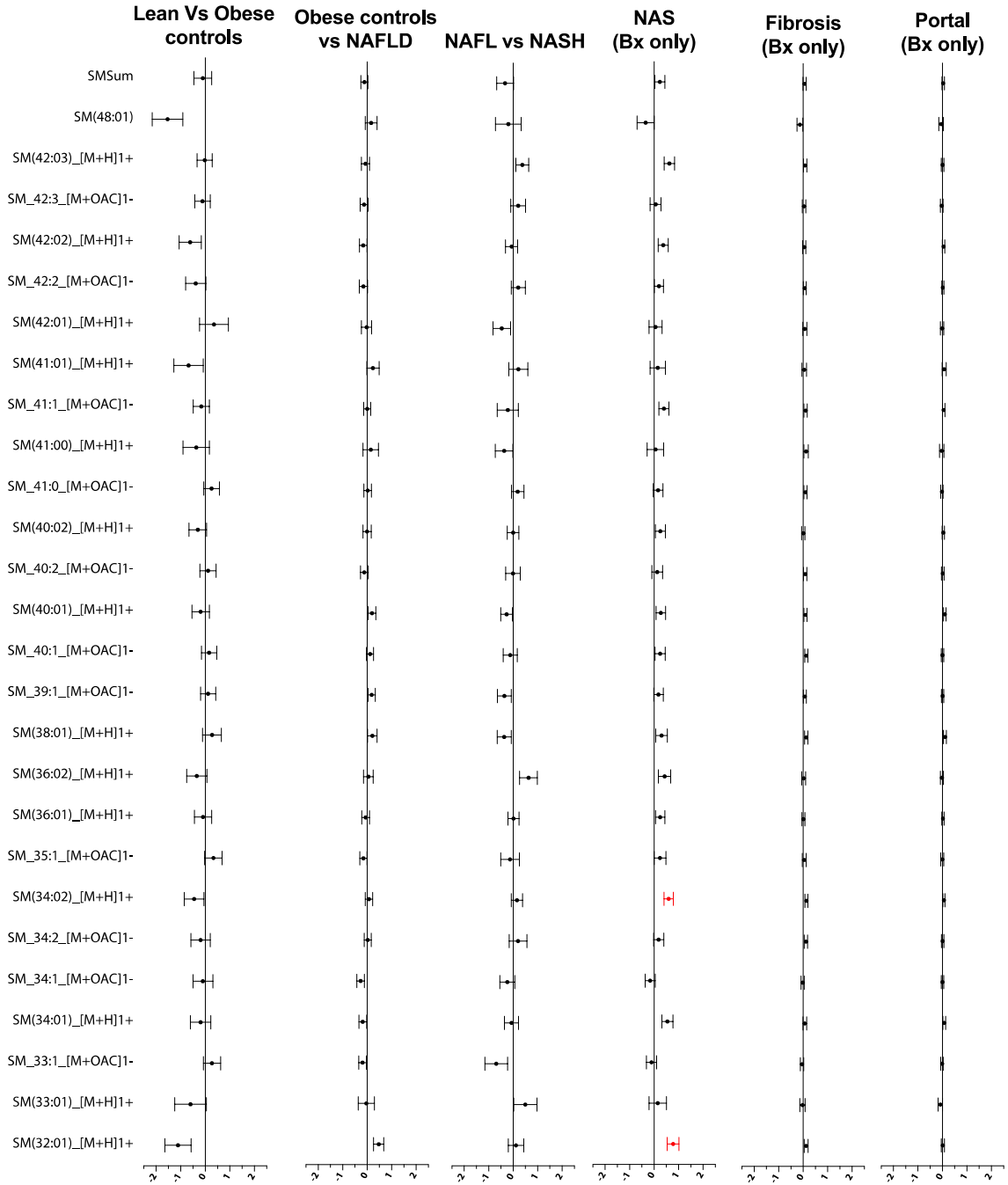
Supplementary Figure 9. Forest plots with results from regression analyses for phosphatidylcholine (PC) species (part 2). Each data point shows the beta with standard error, where beta represents the log-odds for categorical outcomes (lean vs obese controls, obese controls vs NAFLD, and NAFL vs NASH) or the unit change in histology score (NAS, fibrosis, or portal inflammation) for each standard deviation change in lipid. Data points shown in red indicate a significant association after adjusting for multiple comparisons. All regressions were adjusted for age and sex.



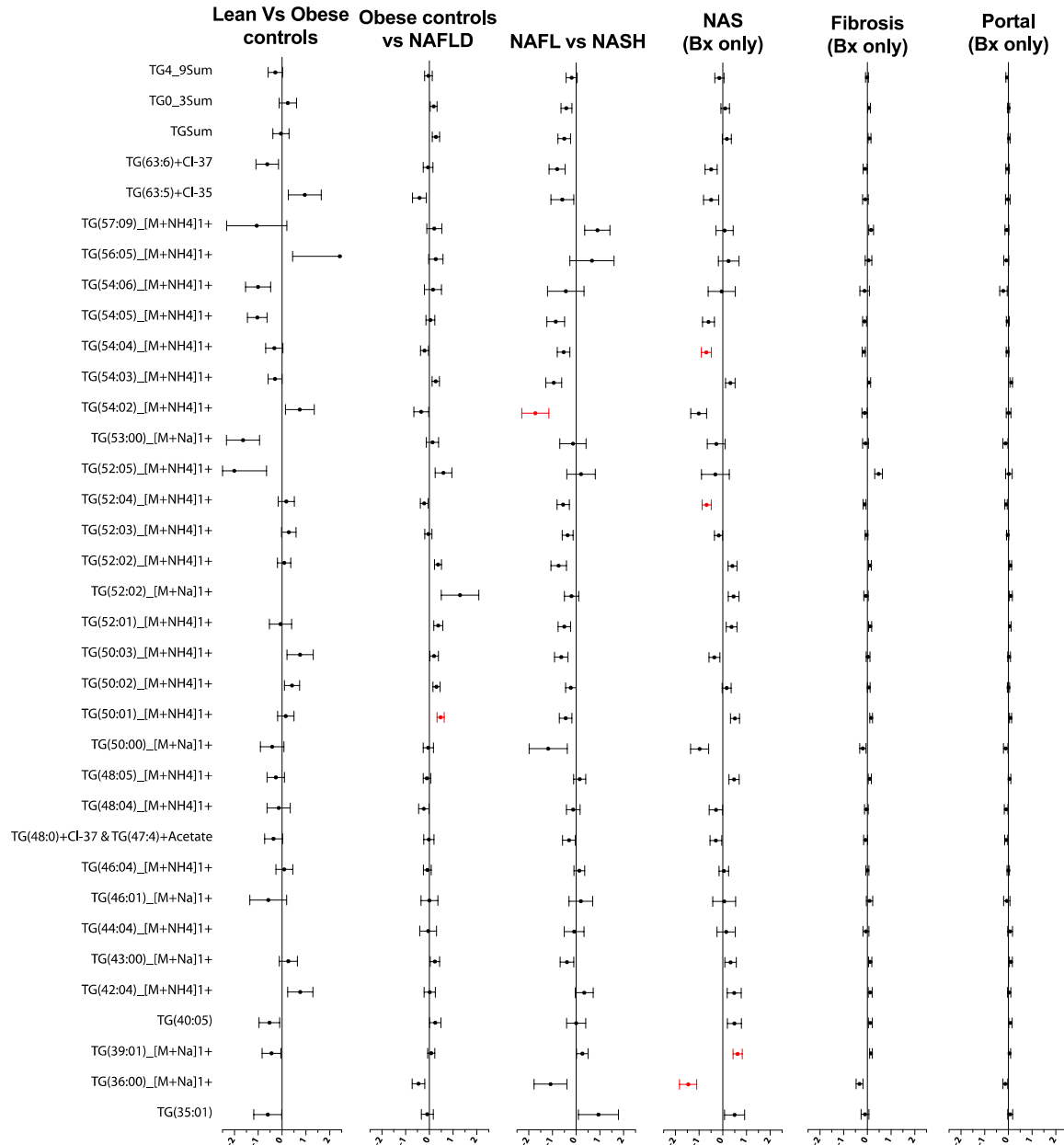
Supplementary Figure 10. Forest plots with results from regression analyses for phosphatidylethanolamines (PE) and phosphatidylglycerol (PG). Each data point shows the beta with standard error, where beta represents the log-odds for categorical outcomes (lean vs obese controls, obese controls vs NAFLD, and NAFL vs NASH) or the unit change in histology score (NAS, fibrosis, or portal inflammation) for each standard deviation change in lipid. Data points shown in red indicate a significant association after adjusting for multiple comparisons. All regressions were adjusted for age and sex.



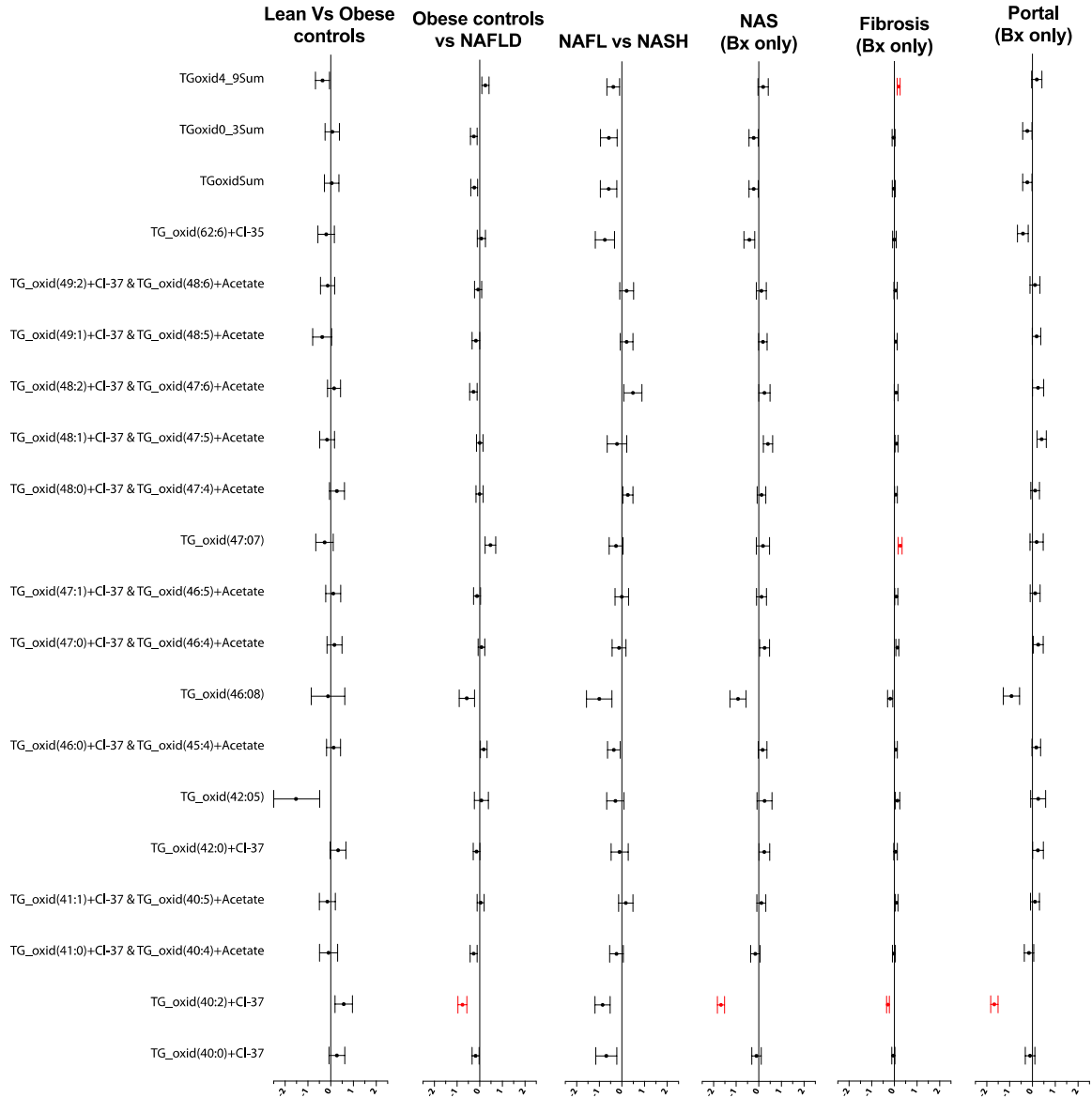
Supplementary Figure 11. Forest plots with results from regression analyses for phosphatidylinositols (PI) and phosphatidylserines (PS). Each data point shows the beta with standard error, where beta represents the log-odds for categorical outcomes (lean vs obese controls, obese controls vs NAFLD, and NAFL vs NASH) or the unit change in histology score (NAS, fibrosis, or portal inflammation) for each standard deviation change in lipid. Data points shown in red indicate a significant association after adjusting for multiple comparisons. All regressions were adjusted for age and sex.



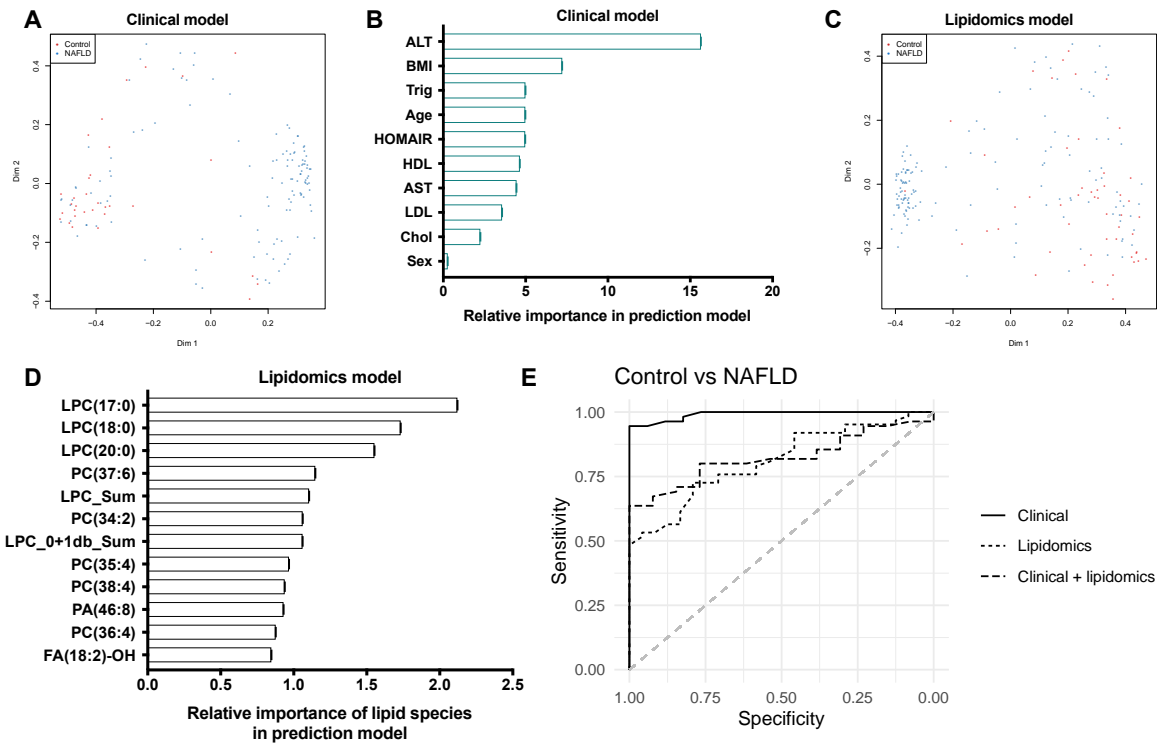
Supplementary Figure 12. Forest plots with results from regression analyses for sphingomyelins (SM). Each data point shows the beta with standard error, where beta represents the log-odds for categorical outcomes (lean vs obese controls, obese controls vs NAFLD, and NAFL vs NASH) or the unit change in histology score (NAS, fibrosis, or portal inflammation) for each standard deviation change in lipid. Data points shown in red indicate a significant association after adjusting for multiple comparisons. All regressions were adjusted for age and sex.



Supplementary Figure 13. Forest plots with results from regression analyses for triglycerides (TG). Each data point shows the beta with standard error, where beta represents the log-odds for categorical outcomes (lean vs obese controls, obese controls vs NAFLD, and NAFL vs NASH) or the unit change in histology score (NAS, fibrosis, or portal inflammation) for each standard deviation change in lipid. Data points shown in red indicate a significant association after adjusting for multiple comparisons. All regressions were adjusted for age and sex.



Supplementary Figure 14. Forest plots with results from regression analyses for triglyceride oxides (TG_oxid). Each data point shows the beta with standard error, where beta represents the log-odds for categorical outcomes (lean vs obese controls, obese controls vs NAFLD, and NAFL vs NASH) or the unit change in histology score (NAS, fibrosis, or portal inflammation) for each standard deviation change in lipid. Data points shown in red indicate a significant association after adjusting for multiple comparisons. All regressions were adjusted for age and sex.



Supplementary Figure 15. Plasma lipidomics does not improve upon clinical parameters for distinguishing NAFLD from control patients. Random forest analysis was used to build a predictive model using clinical and biochemical parameters in a training set (A), where ALT, BMI, and triglycerides were the most important variables (B). The process was replicated using lipidomics in a training set (C), with lipids from different pathways contributing to the model (D). These models were tested in separate validation sets (E), where the clinical model alone was superior to the lipidomics model (or both combined) in distinguishing NAFLD from control participants.

Supplementary Tables

Lipid Class	Lipid Chemistry
Cholesteryl ester	CE(18:0-d ₆)
Ceramide	C16-d ₃₁ Ceramide
Fatty acid	C15:0-d ₂₉ FA
Fatty acid	C17:0-d ₃₃ FA
Fatty acid	C20:0-d ₃₉ FA
lyso-Phosphatidylcholine	lysoPC(C14:0)-d ₄₂
Phosphatidic acid	PA(C16:0-d ₃₁ /C18:1) Na ⁺ salt
Phosphatidylcholine	PC(C16:0-d ₃₁ /C18:1)
Phosphatidylethanolamine	PE(C16:0-d ₃₁ /C18:1)
Phosphatidylglycerol	PG(C16:0-d ₃₁ /C18:1) Na ⁺ salt
Phosphatidylinositol	PI(C16:0-d ₃₁ /C18:1) NH ₄ ⁺ salt
Phosphatidylserine	PS(C16:0-d ₆₂) Na ⁺ salt
Sphingomyelin	SM(C16:0-d ₃₁)
Triglyceride	TG(45:0-d ₂₉)
Triglyceride	TG(48:0-d ₃₁)
Triglyceride	TG(54:0-d ₃₅)

Supplementary Table 1. List of internal standards included in analyses.

Supplementary table 2 [Excel spreadsheet]. Full list of significant associations with lipid species. Beta (SE, standard error) represents a unit change in each outcome (for histological outcomes) or log-odds (for categorical outcomes) per standard deviation change in lipid species. Beta (SE) was calculated using univariable linear regression for continuous outcomes (e.g. NAS) and logistic regression for dichotomous outcomes (e.g. NASH). Analyses were adjusted for age and sex, except where the outcome also includes ‘_BMI-HOMA-IR’ where the analysis was additionally adjusted for BMI and HOMA-IR. The critical p-value to assess significance is given in column N.

	All biopsied NAFLD	NAFL	NASH	P-value	Q-value
	N=120	N=34	N=86		
Female	59. (49)	18 (53)	41 (48)	0.6	0.7
Age	11.6 (2.4)	10.0 (2.8)	12.3 (1.8)	1.3x10 ⁻⁶	6.5x10 ⁻⁶
BMI z-score	2.1 (.6)	2.2 (.7)	2.1 (.6)	0.92	0.92
HOMA-IR	3.2 (1.7)	2.6 (1.2)	3.5 (1.8)	0.005	0.01
ALT	68.7 (41.8)	81.9 (50.4)	63.4 (36.8)	0.03	0.06
AST	47.8 (22.1)	50.8 (26.)	46.6 (20.3)	0.35	0.5
GGT	50.3 (16.4)	48.7 (25.1)	52. (6.1)	0.83	0.9
Total cholesterol	4.3 (.9)	4.1 (1.)	4.3 (.9)	0.33	0.5
HDL	1.1 (.2)	1.1 (.2)	1.1 (.2)	0.53	0.6
LDL	2.6 (.4)	2.6 (.5)	2.6 (.4)	0.49	0.6
Triglycerides	1.4 (.7)	1.3 (.8)	1.5 (.7)	0.19	0.3
<i>PNPLA3</i> rs738409-CC	45 (38)	8 (24)	37 (43)	.03	0.06
<i>PNPLA3</i> rs738409-CG	56 (47)	17 (52)	39 (45)		
<i>PNPLA3</i> rs738409-GG	18 (15)	8 (24)	10 (12)		
<i>TM6SF2</i> rs58542926-CC	103 (87)	27 (82)	76 (88)	.37	0.5
<i>TM6SF2</i> rs58542926-CT	16 (13)	6 (18)	10 (12)		
<i>MBOAT7</i> rs641738-CC	35 (29)	11 (33)	24 (28)	.27	0.5
<i>MBOAT7</i> rs641738-CT	57 (48)	17 (52)	40 (47)		
<i>MBOAT7</i> rs641738-TT	27 (23)	5 (15)	22 (26)		
Steatosis grade 1	34 (28)	16 (47)	18 (21)	5.4x10 ⁻⁴	0.002
Steatosis grade 2	49 (41)	15 (44)	34 (40)		
Steatosis grade 3	37 (31)	3 (9)	34 (40)		
Ballooning – 0	29 (25)	29 (94)	0	2.6x10 ⁻²¹	6.0x10 ⁻²⁰
Ballooning – 1	46 (40)	2 (6)	44 (53)		
Ballooning – 2	39 (34)	0	39 (47)		
Lobular inflammation – 0	15 (13)	15 (44)	0	1.3x10 ⁻⁸	8.7x10 ⁻⁸
Lobular inflammation – 1	64 (53)	16 (47)	48 (56)		
Lobular inflammation – 2	41 (34)	3 (09)	38 (44)		
NAFLD Activity Score	4.3 (1.7)	2.2 (1.1)	5.1 (1.2)	1.7x10 ⁻²⁰	1.7x10 ⁻¹⁹
Portal inflammation – 0	28 (25)	2 (6)	26 (31)	.03	0.06
Portal inflammation – 1	70 (61)	23 (74)	47 (57)		
Portal inflammation – 2	16 (14)	6 (19)	10 (12)		
Fibrosis stage 0	21 (18)	14 (41)	7 (8)	1.9x10 ⁻⁴	7.6x10 ⁻⁴
Fibrosis stage 1	71 (59)	18 (53)	53 (62)		
Fibrosis stage 2	23 (19)	2 (6)	21 (24)		
Fibrosis stage 3	5 (4)	0	5 (6)		

Supplementary Table 3. Participant demographics, biochemistry, and liver histology for children who had undergone liver biopsy. Continuous variables are given as mean (standard deviation). P-values and false-discovery rate corrected Q-values are given for comparison between NAFL and NASH patients. NAFL, non-alcoholic fatty liver; NASH, non-alcoholic steatohepatitis; BMI, body mass index; ALT, alanine aminotransferase; AST, aspartate aminotransferase; GGT, gamma glutamyl transferase; ALP, alkaline phosphatase; HDL, high density lipoprotein; LDL, low density lipoprotein; HOMA-IR, homeostatic model assessment of insulin resistance.

Histological feature (dependent variable)	Participant characteristic (independent variable)	Beta (SE)	P-value	Q-value
NASH	Age (years)	.08 (.02)	1.3x10 ⁻⁶	1.7x10 ⁻⁵
NASH	HOMA-IR	.07 (.02)	0.01	0.02
NAS	BMI z-score	-1.7 (.2)	9.9x10 ⁻¹⁴	3.2x10 ⁻¹²
NAS	Triglycerides	1.6 (.3)	5.2x10 ⁻⁸	8.3x10 ⁻⁷
NAS	AST	.04 (.01)	3.0x10 ⁻⁶	2.7x10 ⁻⁵
NAS	ALT	.02 (.005)	8.3x10 ⁻⁵	5.3x10 ⁻⁴
NAS	Age	-.2 (.07)	0.01	0.03
NAS	Total cholesterol	.6 (.2)	0.01	0.04
Portal inflammation	BMI z-score	-.5 (.06)	5.7x10 ⁻¹⁵	3.7x10 ⁻¹³
Portal inflammation	Age (years)	-.07 (.02)	1.1x10 ⁻⁴	6.2x10 ⁻⁴
Portal inflammation	ALT	.005 (.001)	1.2x10 ⁻⁴	6.2x10 ⁻⁴
Portal inflammation	AST	.01 (.002)	4.1x10 ⁻⁴	2.0x10 ⁻³
Portal inflammation	Triglycerides	.2 (.08)	0.01	0.04
Fibrosis	BMI z-score	-.5 (.07)	1.6x10 ⁻¹¹	3.3x10 ⁻¹⁰
Fibrosis	AST	.01 (.003)	2.8x10 ⁻⁶	2.7x10 ⁻⁵
Fibrosis	ALT	.01 (.001)	2.2x10 ⁻⁵	1.8x10 ⁻⁴
Fibrosis	Triglycerides	.3 (.08)	5.9x10 ⁻⁵	4.2x10 ⁻⁴

Supplementary table 4. Associations between participant characteristics and histological features in children with NAFLD (n=120). Beta (standard error (SE)) represents a unit change in the histological feature per standard deviation change in characteristic. Beta (SE) was calculated using univariable linear regression for continuous outcomes (e.g. NAS) and logistic regression for dichotomous outcomes (e.g. NASH). False-discovery rate corrected Q-values were calculated using all P-values derived from all regression analyses (including those not shown). NASH, non-alcoholic steatohepatitis; NAS, NAFLD Activity Score; BMI, body mass index; ALT, alanine aminotransferase; AST, aspartate aminotransferase; HOMA-IR, homeostatic model assessment of insulin resistance.

Outcome	Parameters	AUC	AUC 95% CI
Control vs. NAFLD	Clinical & biochemical	0.9845	0.9639-1
Control vs. NAFLD	Lipidomics	0.8142	0.7244-0.9039
Control vs. NAFLD	Clinical, biochemical, & lipidomics	0.8406	0.7418-0.9394
NAFL vs. NASH/Fibrosis	Clinical & biochemical	0.8494	0.7464-0.9524
NAFL vs. NASH/Fibrosis	Lipidomics	0.9322	0.8652-0.9993
NAFL vs. NASH/Fibrosis	Clinical, biochemical, & lipidomics	0.87	0.768-0.972

Supplementary table 5. Summary of results from random forest analysis used to distinguish control participants from those with NAFLD, and then those with NASH or fibrosis from those with NAFL alone. AUC, area under the receiver-operator characteristic curve; CI, confidence interval.

Code used in analyses

****STATA v16.1 - regression analyses for associations of lipids using logarithmically transformed lipid abundances, all named 'Lip*' and internal standards named 'IS'**

****normalise lipids**

```
foreach var of varlist IS* Lip* {  
egen std_`var' = std(`var')  
}
```

drop IS* Lip*

****demographic associations with traits**

```
foreach var of varlist Age Sex HOMAIR ALT AST GGT ALP Chol HDL LDL Trig BMIzscore WC  
HipCirc syst_bp dias_bp HTN HFF ELF TE CAP MBOAT7 TM6SF2 PNPLA3 {  
eststo: quietly logistic LnVsOb `var'  
}  
estout using Demog_LnVsOb_v6.txt, cells("b se ci_l ci_u p")  
estimates clear
```

```
foreach var of varlist Age Sex HOMAIR ALT AST GGT ALP Chol HDL LDL Trig BMIzscore WC  
HipCirc syst_bp dias_bp HTN HFF ELF TE CAP MBOAT7 TM6SF2 PNPLA3 {  
eststo: quietly logistic ObVsNAFLD `var'  
}  
estout using Demog_ObVsNAFLD_v6.txt, cells("b se ci_l ci_u p")  
estimates clear
```

```
foreach var of varlist Age Sex HOMAIR ALT AST Chol HDL LDL Trig BMIzscore HTN MBOAT7  
TM6SF2 PNPLA3 {  
eststo: quietly regress NASH `var'  
}  
estout using Demog_NASH_v6.txt, cells("b se ci_l ci_u p")  
estimates clear
```

```
foreach var of varlist Age Sex HOMAIR ALT AST Chol HDL LDL Trig BMIzscore HTN MBOAT7  
TM6SF2 PNPLA3 {  
eststo: quietly regress NAS `var'  
}  
estout using Demog_NAS_v6.txt, cells("b se ci_l ci_u p")  
estimates clear
```

```
foreach var of varlist Age Sex HOMAIR ALT AST Chol HDL LDL Trig BMIzscore HTN MBOAT7
TM6SF2 PNPLA3 {
eststo: quietly regress Fib `var'
}
estout using Demog_Fib_v6.txt, cells("b se ci_l ci_u p")
estimates clear
```

```
foreach var of varlist Age Sex HOMAIR ALT AST Chol HDL LDL Trig BMIzscore HTN MBOAT7
TM6SF2 PNPLA3 {
eststo: quietly regress Portal `var'
}
estout using Demog_Portal_v6.txt, cells("b se ci_l ci_u p")
estimates clear
```

****regressions for associations between lipids and traits, adjusted for age and sex only**

**** lean vs obese controls**

```
foreach var of varlist std_Lip1* std_Lip2* {
eststo: quietly logistic LnVsOb `var' Age i.Sex
}
estout using LnVsOb-1_v6.txt, cells("b se ci_l ci_u p")
estimates clear
```

```
foreach var of varlist std_Lip3* std_Lip4* std_Lip5* std_Lip6* std_Lip7* std_Lip8* std_Lip9*
{
eststo: quietly logistic LnVsOb `var' Age i.Sex
}
estout using LnVsOb-2_v6.txt, cells("b se ci_l ci_u p")
estimates clear
```

****obese controls vs all NAFLD**

```
foreach var of varlist std_Lip1* std_Lip2* {
eststo: quietly logistic ObVsNAFLD `var' Age i.Sex
}
estout using ObVsNAFLD-1_v6.txt, cells("b se ci_l ci_u p")
estimates clear
```

```
foreach var of varlist std_Lip3* std_Lip4* std_Lip5* std_Lip6* std_Lip7* std_Lip8* std_Lip9*
{
eststo: quietly logistic ObVsNAFLD `var' Age i.Sex
```

```
}  
estout using ObVsNAFLD-2_v6.txt, cells("b se ci_l ci_u p")  
estimates clear
```

```
**NAFL vs NASH (biopsy-proven only)
```

```
foreach var of varlist std_Lip1* std_Lip2* {  
  eststo: quietly logistic NASH `var' Age i.Sex  
}  
estout using NASH-1_v6.txt, cells("b se ci_l ci_u p")  
estimates clear
```

```
foreach var of varlist std_Lip3* std_Lip4* std_Lip5* std_Lip6* std_Lip7* std_Lip8* std_Lip9*  
{  
  eststo: quietly logistic NASH `var' Age i.Sex  
}  
estout using NASH-2_v6.txt, cells("b se ci_l ci_u p")  
estimates clear
```

```
**associations with each histological component
```

```
*fibrosis
```

```
foreach var of varlist std_Lip1* std_Lip2* {  
  eststo: quietly regress Fib `var' Age i.Sex  
}  
estout using Fib-1_v6.txt, cells("b se ci_l ci_u p")  
estimates clear
```

```
foreach var of varlist std_Lip3* std_Lip4* std_Lip5* std_Lip6* std_Lip7* std_Lip8* std_Lip9*  
{  
  eststo: quietly regress Fib `var' Age i.Sex  
}  
estout using Fib-2_v6.txt, cells("b se ci_l ci_u p")  
estimates clear
```

```
*portal inflammation
```

```
foreach var of varlist std_Lip1* std_Lip2* {  
  eststo: quietly regress Portal `var' Age i.Sex  
}  
estout using Portal-1_v6.txt, cells("b se ci_l ci_u p")  
estimates clear
```

```
foreach var of varlist std_Lip3* std_Lip4* std_Lip5* std_Lip6* std_Lip7* std_Lip8* std_Lip9*  
{
```

```
eststo: quietly regress Portal `var' Age i.Sex
}
estout using Portal-2_v6.txt, cells("b se ci_l ci_u p")
estimates clear
```

*NAS

```
foreach var of varlist std_Lip1* std_Lip2* {
eststo: quietly regress NAS `var' Age i.Sex
}
estout using NAS-1_v6.txt, cells("b se ci_l ci_u p")
estimates clear
```

```
foreach var of varlist std_Lip3* std_Lip4* std_Lip5* std_Lip6* std_Lip7* std_Lip8* std_Lip9*
{
eststo: quietly regress NAS `var' Age i.Sex
}
estout using NAS-2_v6.txt, cells("b se ci_l ci_u p")
estimates clear
```

**run associations, adjusted for insulin resistance (bmi & homa-ir), for significant lipids only

```
foreach var of varlist std_Lip94 std_Lip63 std_Lip8 std_Lip82 std_Lip92 std_Lip54 std_Lip3
std_Lip38 std_Lip6 std_Lip5 std_Lip53 std_Lip7 std_Lip27 std_Lip72 std_Lip45 std_Lip31
std_Lip93 std_Lip62 std_Lip43 std_Lip44 std_Lip21 std_Lip25 std_Lip52 std_Lip39 std_Lip50
std_Lip49 std_Lip81 std_Lip12 std_Lip1 std_Lip17 std_Lip91 std_Lip113 std_Lip61 std_Lip4
std_Lip32 std_Lip73 std_Lip26 std_Lip157 std_Lip132 std_Lip147 std_Lip168 std_Lip160
std_Lip273 std_Lip238 std_Lip237 std_Lip269 std_Lip260 std_Lip239 std_Lip264 std_Lip240
std_Lip267 std_Lip210 std_Lip268 std_Lip235 std_Lip258 std_Lip261 std_Lip259 std_Lip236
std_Lip374 std_Lip364 std_Lip365 std_Lip356 std_Lip373 std_Lip358 std_Lip367 std_Lip360
std_Lip96 std_Lip34 std_Lip47 std_Lip9 std_Lip30 std_Lip29 std_Lip37 std_Lip55 std_Lip40
std_Lip87 std_Lip10 std_Lip89 std_Lip42 std_Lip51 std_Lip35 std_Lip13 std_Lip23 std_Lip36
std_Lip57 std_Lip56 std_Lip28 std_Lip46 std_Lip64 std_Lip11 std_Lip95 std_Lip85 std_Lip58
std_Lip60 std_Lip175 std_Lip142 std_Lip133 std_Lip134 std_Lip174 std_Lip184 std_Lip180
std_Lip193 std_Lip182 std_Lip164 std_Lip169 std_Lip167 std_Lip189 std_Lip165 std_Lip177
std_Lip181 std_Lip126 std_Lip179 std_Lip298 std_Lip302 std_Lip301 std_Lip229 std_Lip225
std_Lip220 std_Lip242 std_Lip262 std_Lip257 std_Lip319 std_Lip200 std_Lip223 std_Lip245
std_Lip341 std_Lip243 std_Lip256 std_Lip346 std_Lip320 std_Lip303 std_Lip306 std_Lip375
std_Lip370 std_Lip357 std_Lip354 std_Lip368 std_Lip372 std_Lip369 std_Lip366 std_Lip371
std_Lip344 std_Lip71 std_Lip141 std_Lip151 std_Lip146 std_Lip191 std_Lip353 std_Lip382 {
eststo: quietly quietly regress Fib `var' Age i.Sex bmi HOMA
}
estout using Fib_bmi_ir_v1.txt, cells("b se ci_l ci_u p")
estimates clear
```



```
foreach var of varlist std_Lip94 std_Lip63 std_Lip8 std_Lip82 std_Lip92 std_Lip54 std_Lip3
std_Lip38 std_Lip6 std_Lip5 std_Lip53 std_Lip7 std_Lip27 std_Lip72 std_Lip45 std_Lip31
std_Lip93 std_Lip62 std_Lip43 std_Lip44 std_Lip21 std_Lip25 std_Lip52 std_Lip39 std_Lip50
std_Lip49 std_Lip81 std_Lip12 std_Lip1 std_Lip17 std_Lip91 std_Lip113 std_Lip61 std_Lip4
std_Lip32 std_Lip73 std_Lip26 std_Lip157 std_Lip132 std_Lip147 std_Lip168 std_Lip160
std_Lip273 std_Lip238 std_Lip237 std_Lip269 std_Lip260 std_Lip239 std_Lip264 std_Lip240
std_Lip267 std_Lip210 std_Lip268 std_Lip235 std_Lip258 std_Lip261 std_Lip259 std_Lip236
std_Lip374 std_Lip364 std_Lip365 std_Lip356 std_Lip373 std_Lip358 std_Lip367 std_Lip360
std_Lip96 std_Lip34 std_Lip47 std_Lip9 std_Lip30 std_Lip29 std_Lip37 std_Lip55 std_Lip40
std_Lip87 std_Lip10 std_Lip89 std_Lip42 std_Lip51 std_Lip35 std_Lip13 std_Lip23 std_Lip36
std_Lip57 std_Lip56 std_Lip28 std_Lip46 std_Lip64 std_Lip11 std_Lip95 std_Lip85 std_Lip58
std_Lip60 std_Lip175 std_Lip142 std_Lip133 std_Lip134 std_Lip174 std_Lip184 std_Lip180
std_Lip193 std_Lip182 std_Lip164 std_Lip169 std_Lip167 std_Lip189 std_Lip165 std_Lip177
std_Lip181 std_Lip126 std_Lip179 std_Lip298 std_Lip302 std_Lip301 std_Lip229 std_Lip225
std_Lip220 std_Lip242 std_Lip262 std_Lip257 std_Lip319 std_Lip200 std_Lip223 std_Lip245
std_Lip341 std_Lip243 std_Lip256 std_Lip346 std_Lip320 std_Lip303 std_Lip306 std_Lip375
std_Lip370 std_Lip357 std_Lip354 std_Lip368 std_Lip372 std_Lip369 std_Lip366 std_Lip371
std_Lip344 std_Lip71 std_Lip141 std_Lip151 std_Lip146 std_Lip191 std_Lip353 std_Lip382 {
eststo: quietly quietly regress Portal `var' Age i.Sex bmi HOMA
}
estout using Portal_bmi_ir_v1.txt, cells("b se ci_l ci_u p")
estimates clear
```

```
foreach var of varlist std_Lip94 std_Lip63 std_Lip8 std_Lip82 std_Lip92 std_Lip54 std_Lip3
std_Lip38 std_Lip6 std_Lip5 std_Lip53 std_Lip7 std_Lip27 std_Lip72 std_Lip45 std_Lip31
std_Lip93 std_Lip62 std_Lip43 std_Lip44 std_Lip21 std_Lip25 std_Lip52 std_Lip39 std_Lip50
std_Lip49 std_Lip81 std_Lip12 std_Lip1 std_Lip17 std_Lip91 std_Lip113 std_Lip61 std_Lip4
std_Lip32 std_Lip73 std_Lip26 std_Lip157 std_Lip132 std_Lip147 std_Lip168 std_Lip160
std_Lip273 std_Lip238 std_Lip237 std_Lip269 std_Lip260 std_Lip239 std_Lip264 std_Lip240
std_Lip267 std_Lip210 std_Lip268 std_Lip235 std_Lip258 std_Lip261 std_Lip259 std_Lip236
std_Lip374 std_Lip364 std_Lip365 std_Lip356 std_Lip373 std_Lip358 std_Lip367 std_Lip360
std_Lip96 std_Lip34 std_Lip47 std_Lip9 std_Lip30 std_Lip29 std_Lip37 std_Lip55 std_Lip40
std_Lip87 std_Lip10 std_Lip89 std_Lip42 std_Lip51 std_Lip35 std_Lip13 std_Lip23 std_Lip36
std_Lip57 std_Lip56 std_Lip28 std_Lip46 std_Lip64 std_Lip11 std_Lip95 std_Lip85 std_Lip58
std_Lip60 std_Lip175 std_Lip142 std_Lip133 std_Lip134 std_Lip174 std_Lip184 std_Lip180
std_Lip193 std_Lip182 std_Lip164 std_Lip169 std_Lip167 std_Lip189 std_Lip165 std_Lip177
std_Lip181 std_Lip126 std_Lip179 std_Lip298 std_Lip302 std_Lip301 std_Lip229 std_Lip225
std_Lip220 std_Lip242 std_Lip262 std_Lip257 std_Lip319 std_Lip200 std_Lip223 std_Lip245
std_Lip341 std_Lip243 std_Lip256 std_Lip346 std_Lip320 std_Lip303 std_Lip306 std_Lip375
std_Lip370 std_Lip357 std_Lip354 std_Lip368 std_Lip372 std_Lip369 std_Lip366 std_Lip371
std_Lip344 std_Lip71 std_Lip141 std_Lip151 std_Lip146 std_Lip191 std_Lip353 std_Lip382 {
eststo: quietly quietly regress NAS `var' Age i.Sex bmi HOMA
}
estout using NAS_bmi_ir_v1.txt, cells("b se ci_l ci_u p")
estimates clear
```

```

foreach var of varlist std_Lip94 std_Lip63 std_Lip8 std_Lip82 std_Lip92 std_Lip54 std_Lip3
std_Lip38 std_Lip6 std_Lip5 std_Lip53 std_Lip7 std_Lip27 std_Lip72 std_Lip45 std_Lip31
std_Lip93 std_Lip62 std_Lip43 std_Lip44 std_Lip21 std_Lip25 std_Lip52 std_Lip39 std_Lip50
std_Lip49 std_Lip81 std_Lip12 std_Lip1 std_Lip17 std_Lip91 std_Lip113 std_Lip61 std_Lip4
std_Lip32 std_Lip73 std_Lip26 std_Lip157 std_Lip132 std_Lip147 std_Lip168 std_Lip160
std_Lip273 std_Lip238 std_Lip237 std_Lip269 std_Lip260 std_Lip239 std_Lip264 std_Lip240
std_Lip267 std_Lip210 std_Lip268 std_Lip235 std_Lip258 std_Lip261 std_Lip259 std_Lip236
std_Lip374 std_Lip364 std_Lip365 std_Lip356 std_Lip373 std_Lip358 std_Lip367 std_Lip360
std_Lip96 std_Lip34 std_Lip47 std_Lip9 std_Lip30 std_Lip29 std_Lip37 std_Lip55 std_Lip40
std_Lip87 std_Lip10 std_Lip89 std_Lip42 std_Lip51 std_Lip35 std_Lip13 std_Lip23 std_Lip36
std_Lip57 std_Lip56 std_Lip28 std_Lip46 std_Lip64 std_Lip11 std_Lip95 std_Lip85 std_Lip58
std_Lip60 std_Lip175 std_Lip142 std_Lip133 std_Lip134 std_Lip174 std_Lip184 std_Lip180
std_Lip193 std_Lip182 std_Lip164 std_Lip169 std_Lip167 std_Lip189 std_Lip165 std_Lip177
std_Lip181 std_Lip126 std_Lip179 std_Lip298 std_Lip302 std_Lip301 std_Lip229 std_Lip225
std_Lip220 std_Lip242 std_Lip262 std_Lip257 std_Lip319 std_Lip200 std_Lip223 std_Lip245
std_Lip341 std_Lip243 std_Lip256 std_Lip346 std_Lip320 std_Lip303 std_Lip306 std_Lip375
std_Lip370 std_Lip357 std_Lip354 std_Lip368 std_Lip372 std_Lip369 std_Lip366 std_Lip371
std_Lip344 std_Lip71 std_Lip141 std_Lip151 std_Lip146 std_Lip191 std_Lip353 std_Lip382 {
  eststo: quietly quietly logistic NASH `var' Age i.Sex bmi HOMA
}
estout using NASH_bmi_ir_v1.txt, cells("b se ci_l ci_u p")
estimates clear

```

```

## R 3.6.1 for random forest analysis to test predictive capacity of lipidomics against
clinical/biochemical data

```

```

install.packages("MASS")
install.packages("randomForest")
install.packages("dplyr")
install.packages("pROC")

```

```

library(readxl)
library(MASS)
library(randomForest)
library(dplyr)
library(pROC)
library(ggplot2)
set.seed(123)

```

```

## control vs NAFLD for clinical parameters only

```

```

traindata <- read_excel("RF_CtrlNAFLD_clin_v1_train.xlsx")
testdata <- read_excel("RF_CtrlNAFLD_clin_v1_test.xlsx")

```

```
traindata$Dx = as.factor(traindata$Dx)
testdata$Dx = as.factor(testdata$Dx)
```

```
bestmtry<-tuneRF(traindata, traindata $Dx, ntreeTry=20, stepFactor = 1.2, improve = 0.01,
trace =T, plot=T)
```

```
bestmtry
```

```
## = 3
```

```
trainRF <-randomForest(Dx~.,data= traindata, mtry=3, ntree=100,
proximity=TRUE,keep.forest=FALSE)
```

```
pdf(file="MDSplot_CtrlNAFLD_Clin.pdf",width=7.5,height=7.5)
```

```
MDSplot(trainRF, traindata$Dx)
```

```
legend("topleft", pch=16, col = c("firebrick3","dodgerblue3"), legend = c("Control",
"NAFLD"))
```

```
dev.off()
```

```
sink("train_CtrlNAFLD_Clin.txt")
```

```
print(trainRF)
```

```
sink()
```

```
import_trainRF <- importance(trainRF)
```

```
sink("import_CtrlNAFLD_Clin.txt")
```

```
print(import_trainRF)
```

```
sink()
```

```
pdf(file="varImpPlot_CtrlNAFLD_Clin.pdf",width=5,height=5)
```

```
varImpPlot(trainRF)
```

```
dev.off()
```

```
trainRF <-randomForest(Dx~.,data= traindata, mtry=3, ntree=100,
proximity=TRUE,keep.forest=TRUE)
```

```
PredictionsWithClass<-predict(trainRF, testdata, type = 'class')
```

```
t<-table(predictions=PredictionsWithClass, actual = testdata$Dx)
```

```
sink("test_CtrlNAFLD_Clin.txt")
```

```
print(t)
```

```
sink()
```

```
library(pROC)
```

```
Prob_clin <-predict(trainRF, testdata, type = 'prob')
```

```
sink("Probs_CtrlNAFLD_Clin.txt")
```

```
print(Prob_clin)
```

```
sink()
```

```
auc_Clin<-auc(testdata$Dx,Prob_clin [,2])
```

```
sink("AUC_CtrlNAFLD_Clin.txt")
```

```

print(auc_Clin)
sink()

roc_clin <- roc(testdata$Dx, Probab_clin [,2])
ci.auc(roc_clin)

## control vs. NAFLD for lipidomics only

traindata_lip <- read_excel("RF_CtrlNAFLD_lip_v1_train.xlsx")
testdata_lip <- read_excel("RF_CtrlNAFLD_lip_v1_test.xlsx")

traindata_lip$Dx = as.factor(traindata_lip$Dx)
testdata_lip$Dx = as.factor(testdata_lip$Dx)

bestmtry <- tuneRF(traindata_lip, traindata_lip$Dx, ntreeTry=20, stepFactor = 1.2, improve =
0.01, trace =T, plot=T)
bestmtry
## = 15
trainRF_lip <- randomForest(Dx~., data= traindata_lip, mtry=15, ntree=100,
proximity=TRUE, keep.forest=FALSE)

pdf(file="MDSplot_CtrlNAFLD_lip.pdf", width=7.5, height=7.5)
MDSplot(trainRF_lip, traindata_lip$Dx)
legend("topleft", pch=16, col = c("firebrick3", "dodgerblue3"), legend = c("Control",
"NAFLD"))
dev.off()

sink("train_CtrlNAFLD_lip.txt")
print(trainRF_lip)
sink()

import_trainRF_lip <- importance(trainRF_lip)
sink("import_CtrlNAFLD_lip.txt")
print(import_trainRF_lip)
sink()

pdf(file="varImpPlot_CtrlNAFLD_lip.pdf", width=5, height=5)
varImpPlot(trainRF_lip)
dev.off()

trainRF_lip <- randomForest(Dx~., data= traindata_lip, mtry=15, ntree=100,
proximity=TRUE, keep.forest=TRUE)

PredictionsWithClass <- predict(trainRF_lip, testdata_lip, type = 'class')
t <- table(predictions=PredictionsWithClass, actual = testdata_lip$Dx)
sink("test_CtrlNAFLD_lip.txt")
print(t)

```

```
sink()
```

```
Prob_lip <- predict(trainRF_lip, testdata_lip, type = 'prob')  
sink("Probs_CtrlNAFLD_lip.txt")  
print(Prob_lip)  
sink()
```

```
auc_lip <- auc(testdata_lip$Dx, Prob_lip[,2])  
sink("AUC_CtrlNAFLD_lip.txt")  
print(auc_lip)  
sink()
```

```
roc_lip <- plot(roc(testdata_lip$Dx, Prob_lip[,2]))  
ci.auc(roc_lip)
```

```
## run with both clinical and lipidomics
```

```
traindata_both <- read_excel("RF_CtrlNAFLD_both_v1_train.xlsx")  
testdata_both <- read_excel("RF_CtrlNAFLD_both_v1_test.xlsx")
```

```
traindata_both$Dx = as.factor(traindata_both$Dx)  
testdata_both$Dx = as.factor(testdata_both$Dx)
```

```
bestmtry <- tuneRF(traindata_both, traindata_both$Dx, ntreeTry=20, stepFactor = 1.2,  
improve = 0.01, trace = T, plot = T)
```

```
bestmtry
```

```
## = 16
```

```
trainRF_both <- randomForest(Dx ~ ., data = traindata_both, mtry = 16, ntree = 100,  
proximity = TRUE, keep.forest = FALSE)
```

```
pdf(file = "MDSplot_CtrlNAFLD_both.pdf", width = 7.5, height = 7.5)
```

```
MDSplot(trainRF_both, traindata_both$Dx)
```

```
legend("topleft", pch = 16, col = c("firebrick3", "dodgerblue3"), legend = c("Control",  
"NAFLD"))
```

```
dev.off()
```

```
sink("train_CtrlNAFLD_both.txt")
```

```
print(trainRF_both)
```

```
sink()
```

```
import_trainRF_both <- importance(trainRF_both)
```

```
sink("import_CtrlNAFLD_both.txt")
```

```
print(import_trainRF_both)
```

```
sink()
```

```
pdf(file = "varImpPlot_CtrlNAFLD_both.pdf", width = 5, height = 5)
```

```
varImpPlot(trainRF_both)
dev.off()
```

```
trainRF_both <- randomForest(Dx~., data= traindata_both, mtry=16, ntree=100,
proximity=TRUE, keep.forest=TRUE)
```

```
PredictionsWithClass <- predict(trainRF_both, testdata_both, type = 'class')
t <- table(predictions=PredictionsWithClass, actual = testdata_both$Dx)
sink("test_CtrlNAFLD_both.txt")
print(t)
sink()
```

```
Prob_both <- predict(trainRF_both, testdata_both, type = 'prob')
sink("Probs_CtrlNAFLD_both.txt")
print(Prob_both)
sink()
```

```
auc_both <- auc(testdata_both$Dx, Prob_both [,2])
sink("AUC_CtrlNAFLD_both.txt")
print(auc_both)
sink()
```

```
roc_both <- plot(roc(testdata_both$Dx, Prob_both [,2]))
ci.auc(roc_both)
```

```
## combine all in ROC curve
```

```
pdf(file="roc_CtrlNAFLD_all.pdf", width=5, height=3)
g4 <- ggroc(list(roc_clin, roc_lip, roc_both), aes = c("linetype"))
g4 + scale_linetype_discrete(labels=c("Clinical", "Lipidomics", "Clinical + lipidomics")) +
theme_minimal() + ggtitle("Control vs NAFLD") + xlab("Specificity") + ylab("Sensitivity") +
geom_segment(aes(x = 1, xend = 0, y = 0, yend = 1), color="grey", linetype="dashed") +
theme(legend.title=element_blank())
dev.off()
```

```
### repeat for NAFL vs NASH-Fibrosis
```

```
traindata_clin <- read_excel("RF_NAFLNASH_clin_v3_train.xlsx")
testdata_clin <- read_excel("RF_NAFLNASH_clin_v3_test.xlsx")
```

```
traindata_clin$Dx = as.factor(traindata_clin$Dx)
testdata_clin$Dx = as.factor(testdata_clin$Dx)
traindata_clin$Sex = as.factor(traindata_clin$Sex)
testdata_clin$Sex = as.factor(testdata_clin$Sex)
```

```
bestmtry<-tuneRF(traindata_clin, traindata_clin$Dx, ntreeTry=20, stepFactor = 1.2, improve
= 0.01, trace =T, plot=T)
```

```
bestmtry
```

```
## = 3
```

```
trainRF_clin <-randomForest(Dx~.,data= traindata_clin, mtry=5, ntree=100,
proximity=TRUE,keep.forest=FALSE)
```

```
pdf(file="MDSplot_NAFLNASH_Clin.pdf",width=7.5,height=7.5)
```

```
MDSplot(trainRF_clin, traindata_clin$Dx)
```

```
legend("topleft", pch=16, col = c("firebrick3","dodgerblue3"), legend = c("NAFL", "NASH-
Fibrosis"))
```

```
dev.off()
```

```
sink("train_NAFLNASH_Clin.txt")
```

```
print(trainRF_clin)
```

```
sink()
```

```
import_trainRF_clin <- importance(trainRF_clin)
```

```
sink("import_NAFLNASH_Clin.txt")
```

```
print(import_trainRF_clin)
```

```
sink()
```

```
pdf(file="varImpPlot_NAFLNASH_Clin.pdf",width=5,height=5)
```

```
varImpPlot(trainRF_clin)
```

```
dev.off()
```

```
trainRF_clin <-randomForest(Dx~.,data= traindata_clin, mtry=5, ntree=100,
proximity=TRUE,keep.forest=TRUE)
```

```
PredictionsWithClass_clin <-predict(trainRF_clin, testdata_clin, type = 'class')
```

```
t<-table(predictions=PredictionsWithClass_clin, actual = testdata_clin$Dx)
```

```
sink("test_NAFLNASH_Clin.txt")
```

```
print(t)
```

```
sink()
```

```
library(pROC)
```

```
Prob_clin <-predict(trainRF_clin, testdata_clin, type = 'prob')
```

```
sink("Probs_NAFLNASH_Clin.txt")
```

```
print(Prob_clin)
```

```
sink()
```

```
auc_Clin<-auc(testdata_clin$Dx,Prob_clin [,2])
```

```
sink("AUC_NAFLNASH_Clin.txt")
```

```
print(auc_Clin)
```

```
sink()
```

```
roc_clin <- roc(testdata_clin$Dx,Prob_clin [,2])
```

```
ci.auc(roc_clin)
```

```
## NAFLvsNASHFib for lipidomics only
```

```
traindata_lip <- read_excel("RF_NAFLNASH_lip_v4_train.xlsx")
```

```
testdata_lip <- read_excel("RF_NAFLNASH_lip_v4_test.xlsx")
```

```
traindata_lip$Dx = as.factor(traindata_lip$Dx)
```

```
testdata_lip$Dx = as.factor(testdata_lip$Dx)
```

```
bestmtry <- tuneRF(traindata_lip, traindata_lip$Dx, ntreeTry=20, stepFactor = 1.2, improve =  
0.01, trace = T, plot = T)
```

```
bestmtry
```

```
## = 10
```

```
trainRF_lip <- randomForest(Dx~., data = traindata_lip, mtry=18, ntree=100,  
proximity=TRUE, keep.forest=FALSE)
```

```
pdf(file="MDSplot_NAFLNASH_lip.pdf", width=7.5, height=7.5)
```

```
MDSplot(trainRF_lip, traindata_lip$Dx)
```

```
legend("topleft", pch=16, col = c("firebrick3", "dodgerblue3"), legend = c("NAFL",  
"NASH/Fibrosis"))
```

```
dev.off()
```

```
sink("train_NAFLNASH_lip.txt")
```

```
print(trainRF_lip)
```

```
sink()
```

```
import_trainRF_lip <- importance(trainRF_lip)
```

```
sink("import_NAFLNASH_lip.txt")
```

```
print(import_trainRF_lip)
```

```
sink()
```

```
pdf(file="varImpPlot_NAFLNASH_lip.pdf", width=5, height=5)
```

```
varImpPlot(trainRF_lip)
```

```
dev.off()
```

```
trainRF_lip <- randomForest(Dx~., data = traindata_lip, mtry=18, ntree=100,  
proximity=TRUE, keep.forest=TRUE)
```

```
PredictionsWithClass <- predict(trainRF_lip, testdata_lip, type = 'class')
```

```
t <- table(predictions = PredictionsWithClass, actual = testdata_lip$Dx)
```

```
sink("test_NAFLNASH_lip.txt")
```

```
print(t)
```

```
sink()
```

```
Prob_lip <- predict(trainRF_lip, testdata_lip, type = 'prob')
```



```
sink("Probs_NAFLNASH_lip.txt")
print(Prob_lip)
sink()
```

```
auc_lip<-auc(testdata_lip$Dx,Prob_lip [,2])
sink("AUC_NAFLNASH_lip.txt")
print(auc_lip)
sink()
```

```
roc_lip <- roc(testdata_lip$Dx,Prob_lip [,2])
ci.auc(roc_lip)
```

```
## run with both clinical and lipidomics
```

```
traindata_both <- read_excel("RF_NAFLNASH_both_v3_train.xlsx")
testdata_both <- read_excel("RF_NAFLNASH_both_v3_test.xlsx")
```

```
traindata_both$Dx = as.factor(traindata_both$Dx)
testdata_both$Dx = as.factor(testdata_both$Dx)
traindata_both$Sex = as.factor(traindata_both$Sex)
testdata_both$Sex = as.factor(testdata_both$Sex)
```

```
bestmtry<-tuneRF(traindata_both, traindata_both$Dx, ntreeTry=20, stepFactor = 1.2,
improve = 0.01, trace =T, plot=T)
```

```
bestmtry
```

```
## = 16
```

```
trainRF_both <-randomForest(Dx~.,data= traindata_both, mtry=16, ntree=100,
proximity=TRUE,keep.forest=FALSE)
```

```
pdf(file="MDSplot_NAFLNASH_both.pdf",width=7.5,height=7.5)
```

```
MDSplot(trainRF_both, traindata_both$Dx)
```

```
legend("topleft", pch=16, col = c("firebrick3","dodgerblue3"), legend = c("NAFL",
"NASH/Fibrosis"))
```

```
dev.off()
```

```
sink("train_NAFLNASH_both.txt")
```

```
print(trainRF_both)
```

```
sink()
```

```
import_trainRF_both <- importance(trainRF_both)
```

```
sink("import_NAFLNASH_both.txt")
```

```
print(import_trainRF_both)
```

```
sink()
```

```
pdf(file="varImpPlot_NAFLNASH_both.pdf",width=5,height=5)
```

```
varImpPlot(trainRF_both)
dev.off()
```

```
trainRF_both <- randomForest(Dx~., data= traindata_both, mtry=13, ntree=100,
proximity=TRUE, keep.forest=TRUE)
```

```
PredictionsWithClass <- predict(trainRF_both, testdata_both, type = 'class')
t <- table(predictions=PredictionsWithClass, actual = testdata_both$Dx)
sink("test_NAFLNASH_both.txt")
print(t)
sink()
```

```
Prob_both <- predict(trainRF_both, testdata_both, type = 'prob')
sink("Probs_NAFLNASH_both.txt")
print(Prob_both)
sink()
```

```
auc_both <- auc(testdata_both$Dx, Prob_both [,2])
sink("AUC_NAFLNASH_both.txt")
print(auc_both)
sink()
```

```
roc_both <- roc(testdata_both$Dx, Prob_both [,2])
ci.auc(roc_both)
```

```
## combine all in ROC curve
```

```
library(ggplot2)
```

```
pdf(file="roc_NAFLNASH_all.pdf", width=5, height=3)
g4 <- ggroc(list(roc_clin, roc_lip), aes = c("linetype"))
g4 + scale_linetype_discrete(labels=c("Clinical", "Lipidomics")) + theme_minimal() +
ggtitle("NAFL vs NASH/Fibrosis") + xlab("Specificity") + ylab("Sensitivity") +
geom_segment(aes(x = 1, xend = 0, y = 0, yend = 1), color="grey", linetype="dashed") +
theme(legend.title=element_blank())
dev.off()
```