Table of Contents

Extended Materials and Methods 2
- General Study Design 2
- Reference Cigarettes and potential MRTPs 2
- Mice and Inhalation Exposures 3
- Lipidomics 5
- Proteomics 7
- Transcriptomics 9
- Additional data analysis methods 10

References 11

Supplementary Figures 12

Supplementary Tables 14

Supplementary Information

Effects of cigarette smoke, cessation and switching to a reduced risk tobacco product on the liver in Apoe$^{-/-}$ mice – a systems toxicology analysis

Extended Materials and Methods

General Study Design

Apoe$^{-/-}$ study. Female Apoe−/− mice were randomized into five groups: (i) sham (exposed to air), (ii) 3R4F (exposed to CS from the reference cigarette 3R4F), (iii) THS2.2 (exposed to mainstream aerosol from THS2.2 at nicotine levels matched to those of 3R4F), (iv) smoking cessation, and (v) switching to THS2.2. Mice from the sham, 3R4F, and THS2.2 groups were exposed to fresh air, CS from 3R4F, or THS2.2 aerosol, respectively, for up to 8 months. To model effects of smoking cessation and switching to THS2.2, mice from the cessation and switching groups were first exposed to 3R4F for 2 months and then switched to air or THS2.2 aerosol, respectively, for up to 6 additional months (Fig. 1A). Female mice were chosen because of their proposed increased susceptibility to emphysema (Bartalesi et al., 2005).

Reference Cigarettes and potential MRTPs

3R4F cigarettes were purchased from the University of Kentucky (Lexington, KY, USA, http://www.ca.uky.edu/refcig). The candidate MRTP, THS2.2, consists of a stick containing a tobacco plug inserted into a holder that electrically heats the tobacco in a controlled way to ensure combustion temperatures are not reached. This process generates an aerosol containing mainly water, glycerin, nicotine, and tobacco flavors. THS2.2 sticks were produced at Philip Morris International (PMI; Neuchâtel, Switzerland) in three batches. In the smoking system, the THS2.2 stick was inserted into a cigarette holder that heats the tobacco plug. The cigarette holder included a battery, electronics for controlling, a heating element, and the cigarette extractor. Cigarette holders were provided by PMI.

Analytical specifications of the THS2.2 aerosol and 3R4F CS yields are given in (Phillips et al., 2016), which shows the quantification of 56 HPHCs plus water and glycerol (as a humectant) in a total of 58 analytes from 3R4F CS and THS2.2 aerosol. CS from 3R4F cigarettes was generated on 30-port rotary smoking machines (type PMRL-G, SM2000) as described previously (Phillips et al., 2015). Aerosols from THS2.2 sticks were generated on modified 30-port rotary smoking machines equipped with the appropriate holders. Two modified smoking machines per chamber were required to achieve the target THS2.2 aerosol concentration.

3R4F cigarettes and THS2.2 sticks were smoked according to the Health Canada Intensive Smoking Protocol based on ISO standard 3308 (revised in 2000), with the exception of the puff volume (55 ml) and puff frequency (one puff every 30 s) as described previously (Phillips et al., 2015). Several additional minor deviations from ISO standard 3308 were necessary for technical reasons (Phillips et al., 2016).

Mice and Inhalation Exposures

All procedures involving animals were performed in an Association for Assessment and Accreditation of Laboratory Animal Care International-accredited, Agri-Food & Veterinary Authority of Singapore-licensed facility with approval from an Institutional Animal Care and Use Committee (IACUC protocol #15015). This was in compliance with the National Advisory Committee for Laboratory Animal Research Guidelines on the Care and Use of Animals for Scientific Purposes (NACLAR, 2004).

Female B6.129P2-Apoe (tm1Unc)N11 ($Apoe^{-/-}$) mice bred under specific pathogen-free conditions were obtained from Taconic Biosciences (Germantown, NY, USA). The mice were approximately 6–8 wk old on arrival and 8–10 wk old at the start of the exposure. Their health status on arrival was verified using the health check certificate provided by the breeder. Additional health checks on tissue samples collected in Singapore were performed at Harlan Laboratories (Derby, UK). Mice were individually identified by subcutaneous transponders and were housed and whole-body exposed in the animal laboratory under specific pathogen-free conditions. Random assignment of mice to

experimental groups was conducted prior to exposure using a randomization sequence stratified by body weight. A maximum of eight mice were housed per cage. Cage-enrichment (Igloo™, Biosy, Malaysia, and Nylabone™, Neptune City, NJ, USA) was provided in each cage during the non-exposure periods. The bedding material (Lignocel®BK 8-15, J. Rettenmaier & Soehne, GmbH & Co KG., Rosenberg, Germany) was composed of autoclaved softwood (fur and spruce) granulate. A gamma-irradiated pellet diet (T2914C rodent diet, Harlan Laboratories) was provided. Filtered tap water was supplied ad libitum and changed daily. Diet was unavailable only during the exposure periods, but the animals had constant access to drinking water. Additional details of animal housing, randomization, and acclimatization are described in our previous reports (Phillips et al., 2016; Phillips et al., 2015).

The Apoe$^{-/-}$ mice were whole body-exposed to diluted mainstream smoke from 3R4F (600 mg TPM/m3, equivalent to 29.9 µg nicotine/l), THS2.2 aerosol (nicotine-matched to 3R4F, 29.9 mg/ m3) or filtered air for 3 h per day, 5 days per wk, for up to 8 months, with intermittent daily exposure to fresh filtered air for 30 min after the first h of smoke exposure and for 60 min after the second h of exposure. This was done to avoid a build-up of excessive carboxyhemoglobin (COHb) concentrations in the 3R4F group. For the sham group, mice were exposed only to air.

The atmosphere in the aerosol exposure chambers was monitored as previously described (Phillips et al., 2015). Briefly, flow rate, temperature, relative humidity and carbon monoxide (CO) were monitored continuously; TPM (gravimetric evaluation of TPM collected on Cambridge filters), nicotine (LC evaluation of nicotine trapped on sulphuric acid acidified Extrelut® 3NT column) and static puff volume measurements were taken daily; formaldehyde, acetaldehyde, and acrolein (LC evaluation of aldehydes collected and converted to hydrazine derivatives in 2,4 DNPH trapping solution) measurements were taken at least weekly; and particle size distribution was determined at least once per mo.

Animals were observed on a daily basis, body weight progression was monitored weekly, and exposure parameters (COHb in blood and nicotine metabolites in urine) were measured three times during the study. For a more detailed description of the procedures, see (Phillips et al., 2015).

Lipidomics

Samples preparation. Lipidomics samples were analyzed after 2, 3, and 8 months exposure for liver and after 8 months exposure for plasma samples. For each exposure group, tissue samples from eight mice, providing eight biological replicates, were analyzed. Liver samples were pulverized with a CP02 CryoPrep Dry Pulverization System (Covaris, Woburn, MA,USA) and samples, at a concentration of 100 mg/ml, were homogenized in ice-cold 70% methanol-H_2O containing 0.1% butyl-hydroxy-toluene (BHT). Homogenized samples were stored at -80°C prior to lipid extraction and analysis.

Lipid extraction. Lipids (ceramides, cholesterol, glycerolipids, and glycerophospholipids) were extracted using a modified Folch lipid extraction (Jung et al., 2011) performed on a Hamilton Microlab Star robot. Samples were spiked with known amounts of non-endogeneous synthetic internal standards. After lipid extraction, samples were reconstituted in chloroform:methanol (1:2, v/v) and synthetic external standards were post-extract spiked to the extracts. The extracts were stored at -20°C prior to MS analysis. For quantification of free cholesterol, an aliquot of each lipid extract was treated with acetyl chloride to derivatize FC to modified cholesteryl ester species.

Mass spectrometry analyses. In Shotgun Lipidomics, lipid extracts were analyzed using a hybrid triple quadrupole/linear ion trap mass spectrometer (QTRAP 5500) equipped with a robotic nanoflow ion source (NanoMate TriVersa) according to Ståhlman and colleagues (Ståhlman et al., 2009). Molecular lipids were analyzed in both positive and negative modes using Multiple Precursor Ion Scanning (MPIS) based methods (Ekroos et al., 2002; Ekroos et al., 2003). Triacylglycerols (TAG) were analyzed using Precursor Ion Scanning (PIS) and Neutral Loss scanning (NL) based methods (McAnoy et al., 2005). The molecular lipid species were identified and quantified in semi-absolute or absolute amounts (Ståhlman et al., 2009) according to the corresponding LMSs. Targeted molecular

lipids were analyzed on a hybrid triple quadrupole/linear ion trap mass spectrometer (5500 QTRAP) equipped with an ultra-high pressure liquid chromatography (UHPLC) system (Eksigent 100-XL) using multiple reaction monitoring (MRM) –based method in negative ion mode according to the corresponding LMS.

Data processing. Software used in the data analysis: Processing of QTRAP: LipidViewTM (Version 1.2); MultiQuantTM (Version 3.0). Analytical software: Analyst (Version 1.6.2). Data analysis: SAS (Version 9.3); Tableau desktop (Version 8.2).

MS data. The mass spectrometry data obtained from MS instruments were exported as .wiff or .txt files that represent the basic raw files of lipidomic analysis. These files contain information on masses of identified molecules and their counts (intensities and areas). Masses and counts of detected peaks were converted into a list of corresponding lipid names and concentrations. Calibration lines were generated to determine the dynamic quantification range for each lipid monitored, e.g., the quantification limits. As the internal standards used behave in the same way as endogenous lipids they were used for quantifying endogenous lipids. The MS data files were processed using LipidViewTM 1.2 or MultiQuant™ 2.0 for producing a list of lipid names and peak areas. Lipids were normalized to their respective internal standard (the peak area of the endogenous lipid was divided by the peak area of the corresponding internal standard) and sample volume. The concentrations of molecular lipids are presented as pmol/µl or pmol/µg tissue wet weight for plasma or tissue samples, respectively.

Quality control. The ratio of synthetic Internal Standards (IS) to corresponding post-extract spiked External Standards (ES), and MS analysis of extracted matrix and solvents served as quality controls (QC) of the analysis. In addition, extracted reference plasma/tissue samples were analyzed for monitoring the instruments' performance.

Quantification range. Calibration lines using synthetic standards were obtained prior to sample analysis. The calibration lines consist of a minimum of four accepted standard points covering the linear quantification range. The calibration lines were used to determine the quantification limits of the method.

Analysis acceptance criteria. The analysis was accepted based on the linearity of the calibration lines. The linear regression must exceed 0.95 based on at least four out of six non-zero standards. The analysis is accepted based on the identification of sample specific IS.

Data analysis and differential abundance. Outlier samples with low total concentrations of all measured lipids were excluded. Samples with total lipid concentrations 1.5 interquartile ranges (IQRs) below the first quartile were defined as outliers. The "species fractions" represent the relative molar contribution of a given lipid species to the total measured concentration of its class, e.g. the contribution of PC 16:0/16:0 to all PC lipids. These "species fractions" were calculated as the median concentration of the lipid species divided by the sum of all median concentrations of its lipid class. The differential abundance analysis was conducted separately for each exposure time point. In each comparison between a given exposure group and the respective sham group only lipid species detected in at least 50% of the samples of each compared group were considered. The log2 fold-change relative to corresponding levels in the sham exposure group was calculated. A t-test was used to estimate the statistical significance and within each comparison the obtained p-values were corrected for multiple hypothesis testing using the Benjamini-Hochberg (BH) approach. Abundance differences of lipid species with a BH-adjusted p-value < 0.05 were considered as statistically significant.

Proteomics

Sample preparation for LC-MS based proteomics. Tissue samples from livers of eight mice, providing eight biological replicates, were analyzed for each exposure condition and time point. Liver tissue samples from months 1, 2, 3, 6, and 8 for 3R4F and THS 2.2 exposures and months 3, 6, and 8 for cessation and switching (to THS 2.2) were available for quantitative proteomic analysis. All samples were processed in random order. Frozen liver samples were pulverized with a Cryoprep Impactor (Covaris, Woburn, MA, USA) and lyzed in lysis buffer (BioRad, Hercules, CA, USA) with a bead-assisted procedure in Tissue Lyser II. After lysis the proteins were precipitated with acetone (Sigma-Aldrich). The precipitate was resuspended in 0.5 M triethylammonium bicarbonate (TEAB, Sigma-

Aldrich), 1M urea (Sigma-Aldrich) and 0.1% SDS (Sigma-Aldrich). Aliquots of 50 µg were processed for the TMT 6-plex labeling procedure according to the manufacturer's instructions (Thermo Scientific, USA). Trypsin (Promega) was added to samples at a 1:10 trypsin to protein ratio (w/w) followed by overnight digestion at 37°C. Trypsin-digested samples were labeled with TMT reporter-ion tags. In addition, a common reference mix containing 50 µg of all protein extracts from each time point was prepared and labeled with a TMT reporter-ion tag. For each month of exposure eight replicate TMT 6-plex sets with the five different exposure groups and the reference mix were defined. Within each month, replicates were randomly assigned to the TMT sets and the sample-to-reporter mapping was randomized within each set. All labeled samples of one TMT set were pooled and dried in a SpeedVac. Samples were purified from SDS and remaining salts with 0.5 ml bed volume detergent removal columns (Pierce) and 1 cc C18 reversed phase SepPak columns (Waters) according to the manufacturers' instructions, followed by drying in a Speedvac and resuspension in 20 µL of nanoLC buffer A (5% acetonitrile (Sigma-Aldrich), 0.2% formic acid (Sigma-Aldrich).

Mass-spectrometry for identification and quantification. Samples were analyzed in random order using an Easy nanoLC 1000 instrument (Thermo Scientific) connected online to a Q-Exactive (Thermo Scientific) mass-analyzer. Peptides were fractionated on a 50 cm C18RP RSLC Easyspray column (2 µm particle size; Thermo Scientific) at a flow rate of 200 nl/min with a 200 min gradient from nanoLC buffer A (5% acetonitrile, 0.2% formic acid) to 40% nanoLC buffer B (90 % acetonitrile, 0.2% formic acid). Each sample was injected twice with 2 different analysis methods: fast and a sensitive methods as described by Kelstrup et al. (Kelstrup et al., 2012) on the same column. Both mass-spectrometry runs were mass-list against the mouse reference proteome set (Uniprot, version July 2014, canonical isoforms only) using Proteome Discoverer vers. 1.4.0.288 software (Thermo Scientific). Mascot (v.2.4.1, Matrixscience) and SequestHT (implemented in Proteome Discoverer) were used as search tools and resulting protein lists were merged and TMT-reporter ion intensities were determined from the Proteome Discoverer software. The Percolator node of the Proteome Discoverer software was used to estimate peptide-level adjusted p-values (q-values) and the peptides were filtered for a q-value < 0.05 (that is, the false discovery rate was controlled at the 5%

level). TMT peptide-level quantification data was exported and further processed in the R statistical environment (R Development Core Team, 2007). The quantification data were filtered for "unique" quantification results as defined by the Proteome Discoverer software, for example, to remove redundant quantification results from multiple search engines. A global variance stabilizing normalization (VSN) was performed with the respective Bioconductor package in R (Huber et al., 2002; Karp et al., 2010). Each TMT reporter ion set was normalized to its median and protein expression values were calculated as the median of these normalized peptide-level quantification values (Herbrich et al., 2013). For the detection of differentially expressed proteins, a linear model was fit for each exposure condition and its respective sham group and p-values from a moderated t-statistics were calculated with the empirical Bayes approach (Gentleman et al., 2004). The Benjamini-Hochberg False Discovery Rate (FDR) method was then used to correct for multiple testing effects. Proteins with an adjusted p-value < 0.05 were considered as differentially expressed.

Transcriptomics.

Total RNA was isolated from tissues using the miRNeasy Mini Kit (Qiagen), resuspended in an appropriate volume of RNAse-free water, quantified using the Nanodrop 8000 Spectrophotometer (Thermo Scientific, Waltham, MA, USA) and quality checked using the Agilent 2100 Bioanalyzer (Agilent Technologies). The Gene expression workflow was performed using the Affymetrix HT IVT express kit (Affymetrix, Santa Clara, CA):Total RNA (100 ng) was reverse-transcribed into cRNA, amplified and labeled (using an enzymatic attachment of nucleotides coupled to biotin). cRNA was then purified using magnetic beads to remove unincorporated nucleotide triphosphates, salts, enzymes, and inorganic phosphates. Purified cRNA was quantified, normalized. Fragmented, labeled cRNA (33 µL) was added to 217 µL of Hybridization Cocktail Master Mix (Affymetrix, Santa Clara, CA), then denatured for 5 min at 99°C and 5 min at 45°C, followed by centrifugation at Vmax for 1 min. A total of 200 µL of the cRNA cocktail was then hybridized on GeneChip® Mouse Genome 430 2.0 Array (Affymetrix). The arrays were incubated in the GeneChip® Hybridization Oven 645 for 18 ± 2 h at 45°C with a rotation speed of 60 rpm. After hybridization, the arrays were washed and stained

on a Fluidics Station FS450DX using Affymetrix GeneChip® Command Console™ Software (AGCC software version 3.2) with protocol FS450_0001, then scanned using the GeneChip® Scanner 3000 7G. Raw images from the scanner were saved as DAT files. Using the AGCC Viewer software application, each image was checked for artifacts, overall intensity distribution, checkerboards at the corners, a central cross to ensure adequate grid alignment, and readability of the array name. The AGCC software automatically gridded the DAT file image and extracted probe cell intensities into a CEL file. These files were further processed (MAS5.0) with Affymetrix Expression Console™ software (version Build 1.3.1.187) for a first quality check of the data.

Additional data analysis methods

Protein set analysis was supported by the Piano package in the R statistical environment (Väremo et al., 2013). Lipid pathway maps were obtained from the KEGG database (Kanehisa et al., 2014). The $\log_2$ fold-change was used as the protein statistic, the mean as the set statistic and sample permutation within the group comparisons and p-value adjustment with the Benjamini-Hochberg procedure were used to estimate statistical significances. The sparse partial least squares (sPLS) approach was conducted in the canonical mode with the mixOmics package in R (Lê Cao et al., 2009).

References

Ekroos, K., et al., 2002. Quantitative profiling of phospholipids by multiple precursor ion scanning on a hybrid quadrupole time-of-flight mass spectrometer. Analytical chemistry. 74**,** 941-949.

Ekroos, K., et al., 2003. Charting molecular composition of phosphatidylcholines by fatty acid scanning and ion trap MS3 fragmentation. Journal of lipid research. 44**,** 2181-2192.

Gentleman, R. C., et al., 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome biology. 5**,** R80.

Herbrich, S. M., et al., 2013. Statistical Inference from Multiple iTRAQ Experiments without Using Common Reference Standards. Journal of proteome research. 12**,** 594-604.

Huber, W., et al., 2002. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics. 18**,** S96-S104.

Jung, H. R., et al., 2011. High throughput quantitative molecular lipidomics. Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids. 1811**,** 925-934.

Kanehisa, M., et al., 2014. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic acids research. 42**,** D199-D205.

Karp, N. A., et al., 2010. Addressing accuracy and precision issues in iTRAQ quantitation. Molecular & Cellular Proteomics. 9**,** 1885-1897.

Kelstrup, C. D., et al., 2012. Optimized fast and sensitive acquisition methods for shotgun proteomics on a quadrupole orbitrap mass spectrometer. Journal of proteome research. 11**,** 3487-3497.

Lê Cao, K.-A., et al., 2009. integrOmics: an R package to unravel relationships between two omics datasets. Bioinformatics. 25**,** 2855-2856.

McAnoy, A. M., et al., 2005. Direct qualitative analysis of triacylglycerols by electrospray mass spectrometry using a linear ion trap. Journal of the American Society for Mass Spectrometry. 16**,** 1498-1509.

Phillips, B., et al., 2016. An 8-Month Systems Toxicology Inhalation/Cessation Study in Apoe-/- Mice to Investigate Cardiovascular and Respiratory Exposure Effects of a Candidate Modified Risk Tobacco Product, THS 2.2, Compared with Conventional Cigarettes. Toxicol Sci. 149**,** 411-432.

Phillips, B., et al., 2015. A 7-month cigarette smoke inhalation study in C57BL/6 mice demonstrates reduced lung inflammation and emphysema following smoking cessation or aerosol exposure from a prototypic modified risk tobacco product. Food Chem Toxicol. 80**,** 328-45.

Ståhlman, M., et al., 2009. High-throughput shotgun lipidomics by quadrupole time-of-flight mass spectrometry. Journal of Chromatography B. 877**,** 2664-2672.

Väremo, L., et al., 2013. Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. Nucleic acids research. 41**,** 4378-4391.

Supplementary Figures

Supplementary Figure 1. Assessment of the effects on liver by histopathology. Evaluation results for the histopathology endpoints. Data are means ± SEM, n = 8. Statistics: * $p < 0.05$ compared with sham at same time point; # $p < 0.05$ compared with month 1 for the same exposure group.

Supplementary Figure 2. Principal component and partial-least-squares (PLS) analyses. (A) Principal component analysis (PCA) of transcriptomics data. Pair-wise score plots shown for components 1-4. Lines connect the samples of each exposure group. (B) Principal component analysis (PCA) of proteomics data. (C) Sparse canonical partial-least-squares (sPLS-can) analysis. Pair-wise score plots shown for components 1-3. PLS-can is conceptually similar to a principal component analysis, but is designed to identify shared correlated components among data sets. Exposure groups are color-coded and the numbers indicate the time point in months. Note that, of the identified components, component 3 shows the highest correlation between the mRNA and protein data (correlation coefficients were 0.89, 0.80, and 0.92 for the three identified components).

Supplementary Figure 3. Set enrichment results for pathways from the KEGG database. The fold-change was used as the protein-level metric and the mean as the protein set statistic (Väremo et al., 2013); pathways were obtained from the KEGG database (http://www.genome.jp/kegg/). Significantly affected pathways (BH-adjusted p-value < 0.05) in any of the comparisons are shown. The enrichment score is color coded (see key) and the statistical significance of an enrichment is indicated with "*" (BH-adjusted p-value < 0.05).

Supplementary Figure 4. Plasma lipidomics. (A) Lipid classes in plasma after 8 months of exposure. The mean concentration ± SEM is shown. * = p-value < 0.05 (compared to Sham group). (B) Lipidomics results for individual lipid species. Differential abundance profiles for lipid species with significant differential abundance in any exposure to sham comparison (BH-adjusted p-value < 0.05). Each row shows one lipid species, each column an exposure condition compared to the sham

group at the respective time point. The fold-change compared to sham exposure is color-coded (see color key). Significantly abundant differences are indicated (x = BH-adjusted p-value <0.05, * = BH-adjusted p-value < 0.01). In addition, the molar contribution of significantly abundant lipid species to their lipid class is indicated by the shade of the significance indicator ("species fractions"). (C) Lipid-species and sample level lipidomics results for TAG class. The measured lipid concentration is represented (see color key) for each TAG lipid species (rows) and each sample (columns). The groups are indicated at the top and the identifiers of individual mice at the bottom of the heatmap.

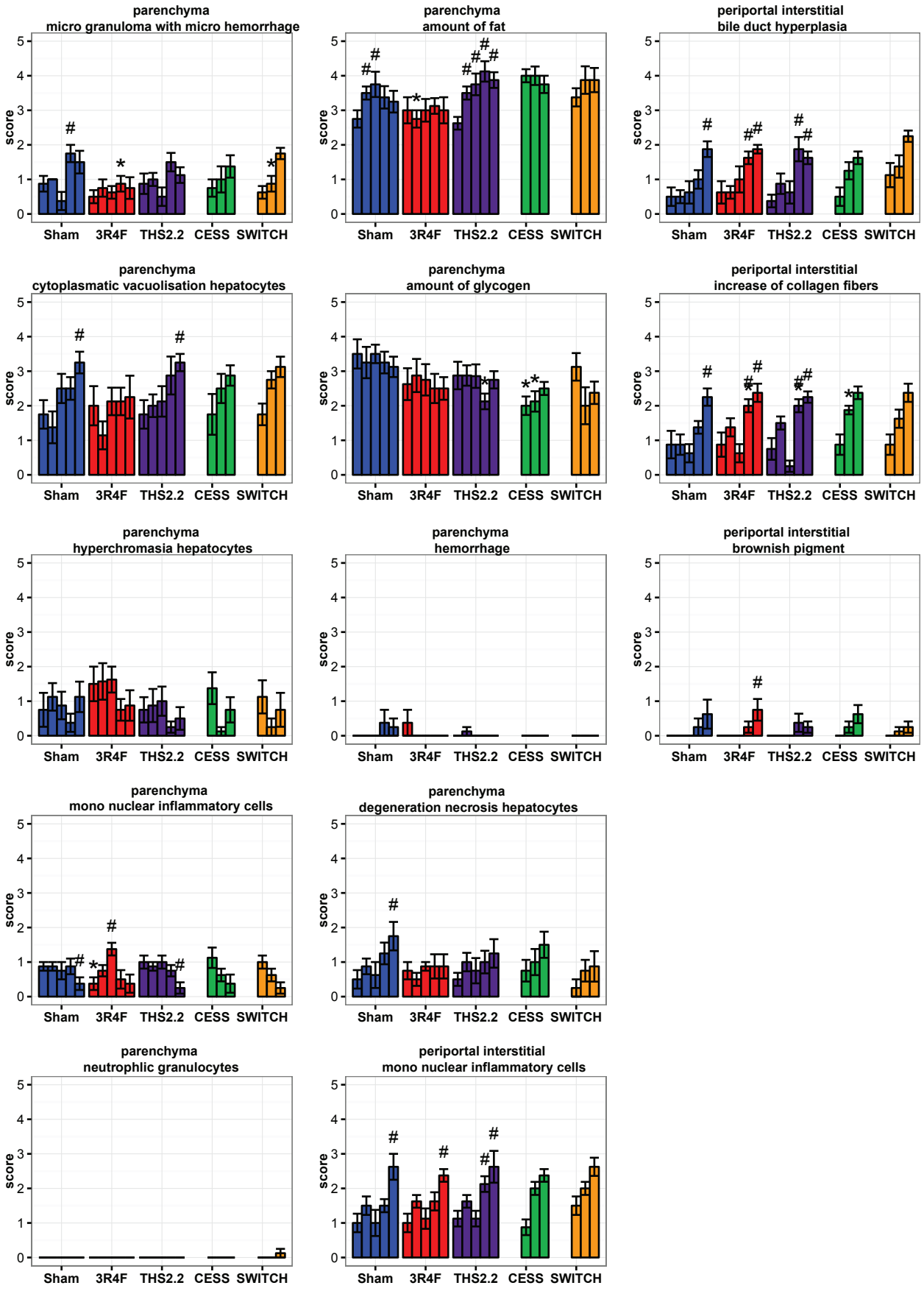
Supplementary Figure 5. Liver lipidomics. Lipidomics results for individual lipid species. Differential abundance profiles for lipid species with significant differential abundance in any exposure to sham comparison (BH-adjusted p-value < 0.05). Each row shows one lipid species, each column an exposure condition compared to the sham group at the respective time point. The fold-change compared to sham exposure is color-coded (see color key). Significantly abundant differences are indicated (x = BH-adjusted p-value <0.05, * = BH-adjusted p-value < 0.01). In addition, the molar contribution of significantly abundant lipid species to their lipid class is indicated by the shade of the significance indicator ("species fractions").

Supplementary Tables

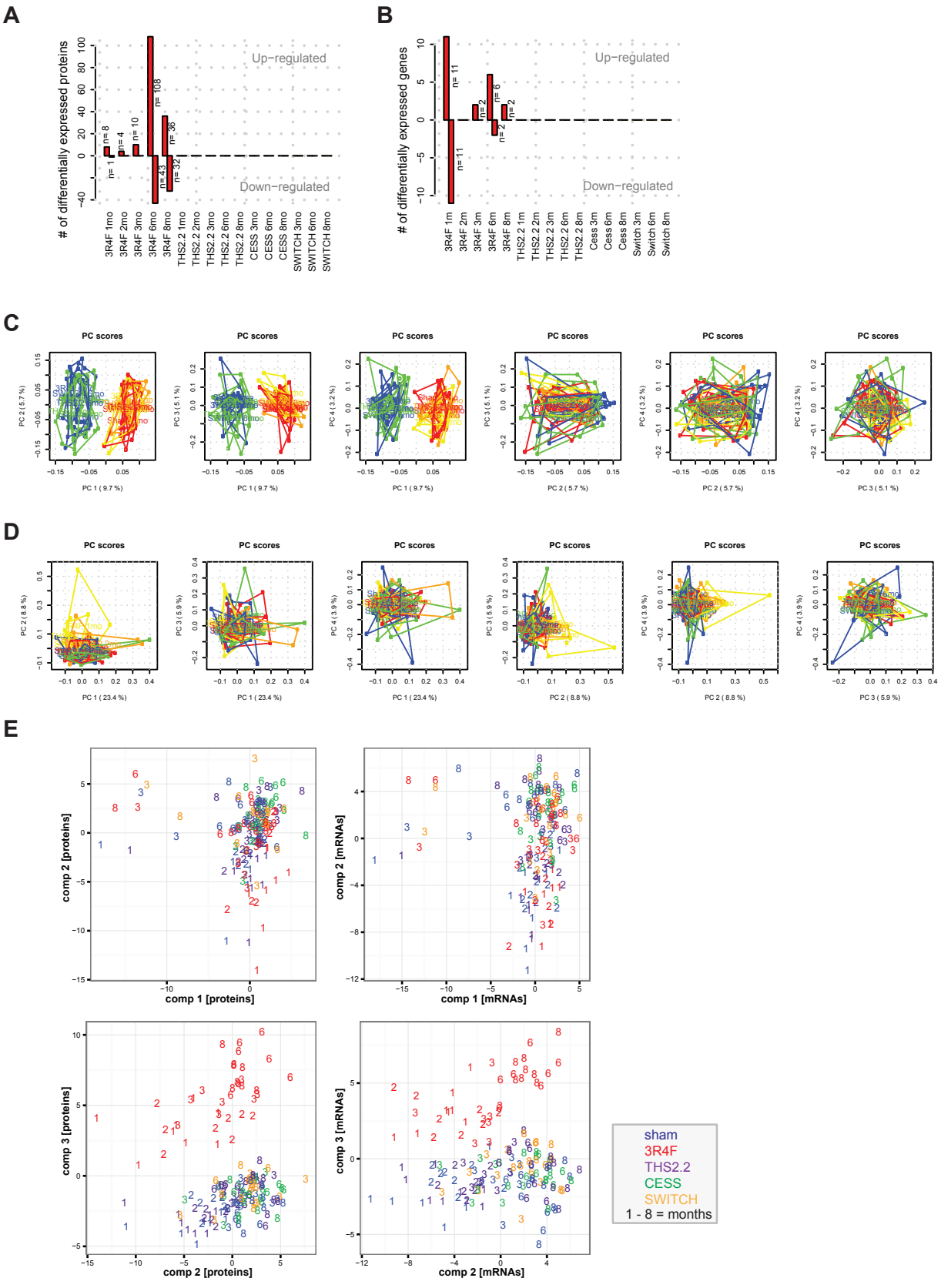
Supplementary Table 1. GSEA results for the 1st principal component of the transcriptomics data.

Supplementary Table 2. Lipidomics data. Data table of lipidomics analyses.

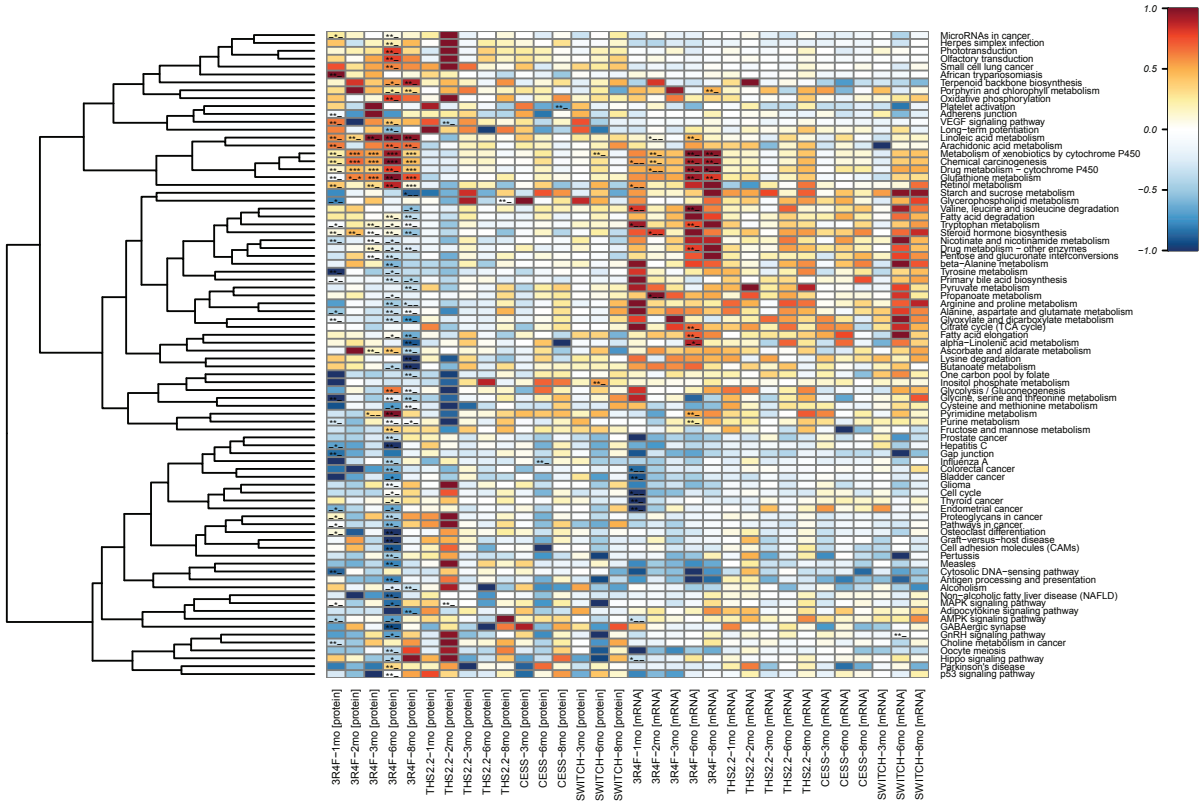
Supplementary Figure 1



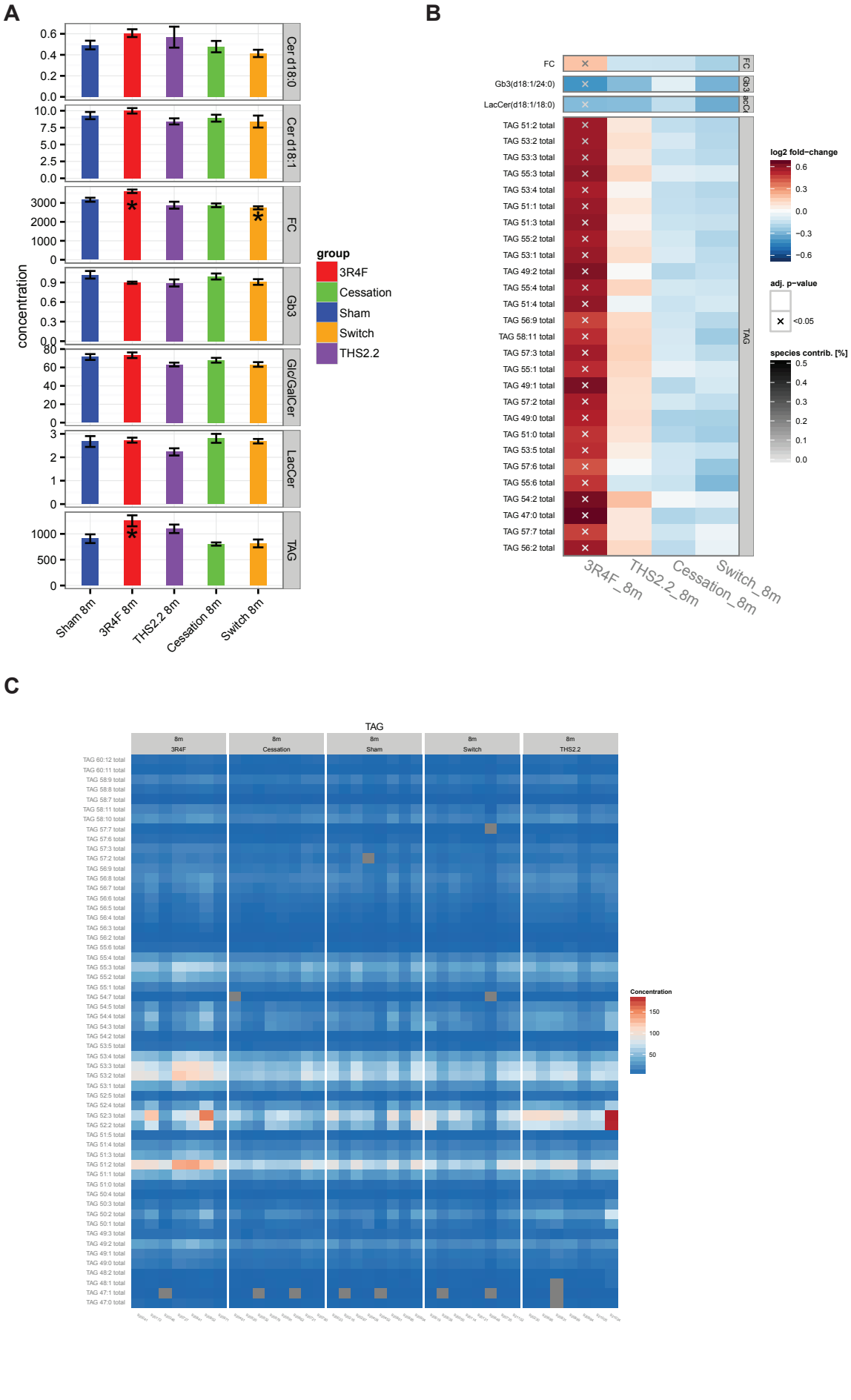
Supplementary Figure 2



Supplementary Figure 3



Supplementary Figure 4



Supplementary Figure 5

