An unbiased ranking of murine dietary models based on their proximity to human metabolic dysfunction-associated steatotic liver disease (MASLD)

A list of authors and their affiliations appears at the end of the paper

Metabolic dysfunction-associated steatotic liver disease (MASLD), previously known as non-alcoholic fatty liver disease, encompasses steatosis and metabolic dysfunction-associated steatohepatitis (MASH), leading to cirrhosis and hepatocellular carcinoma. Preclinical MASLD research is mainly performed in rodents; however, the model that best recapitulates human disease is yet to be defined. We conducted a wide-ranging retrospective review (metabolic phenotype, liver histopathology, transcriptome benchmarked against humans) of murine models (mostly male) and ranked them using an unbiased MASLD 'human proximity score' to define their metabolic relevance and ability to induce MASH-fibrosis. Here, we show that Western diets align closely with human MASH; high cholesterol content, extended study duration and/or genetic manipulation of disease-promoting pathways are required to intensify liver damage and accelerate significant (F2+) fibrosis development. Choline-deficient models rapidly induce MASH-fibrosis while showing relatively poor translatability. Our ranking of commonly used MASLD models, based on their proximity to human MASLD, helps with the selection of appropriate in vivo models to accelerate preclinical research.

Non-alcoholic fatty liver disease (NAFLD) is the hepatic manifestation of the metabolic syndrome[1]. It clusters with obesity, insulin resistance, diabetes, dyslipidemia, atherosclerosis and cancer[2,3]. Given its metabolic context, a multi-society consensus statement on fatty liver disease proposed the new nomenclature, metabolic dysfunction-associated steatotic liver disease (MASLD)[4].

The MASLD spectrum ranges from simple steatosis to steatohepatitis (NASH/MASH) and MASH-fibrosis[5,6]. Under the pressure of environmental factors (lifestyle, nutrition, microbiome) and genetic predisposition (for example, the rs738409 C>G polymorphism in the *PNPLA3* gene), the disease can progress to MASH, promoted by lipotoxic insults driving hepatocyte injury, inflammation and chronic activation of wound-healing responses[7,8]. Progressive MASH places patients at risk of cirrhosis and hepatocellular carcinoma, which may result in liver-related mortality and the need for liver transplantation[1,9].

The lack of a standard translationally relevant preclinical model has hindered the field's ability to study the chronic and complex pathophysiology of MASH. Furthermore, overinflated preclinical efficacy data generated from models in which human pathophysiology is not accurately replicated probably contribute to negative clinical trial results in the MASH field. Many diets differing in macronutrient

✉e-mail: michele.vacca@uniba.it; Dina.Tiniakos@newcastle.ac.uk; Perfield_james_w@lilly.com; petsalaki@ebi.ac.uk; pkdavidsen@gmail.com; ajv22@medschl.cam.ac.uk
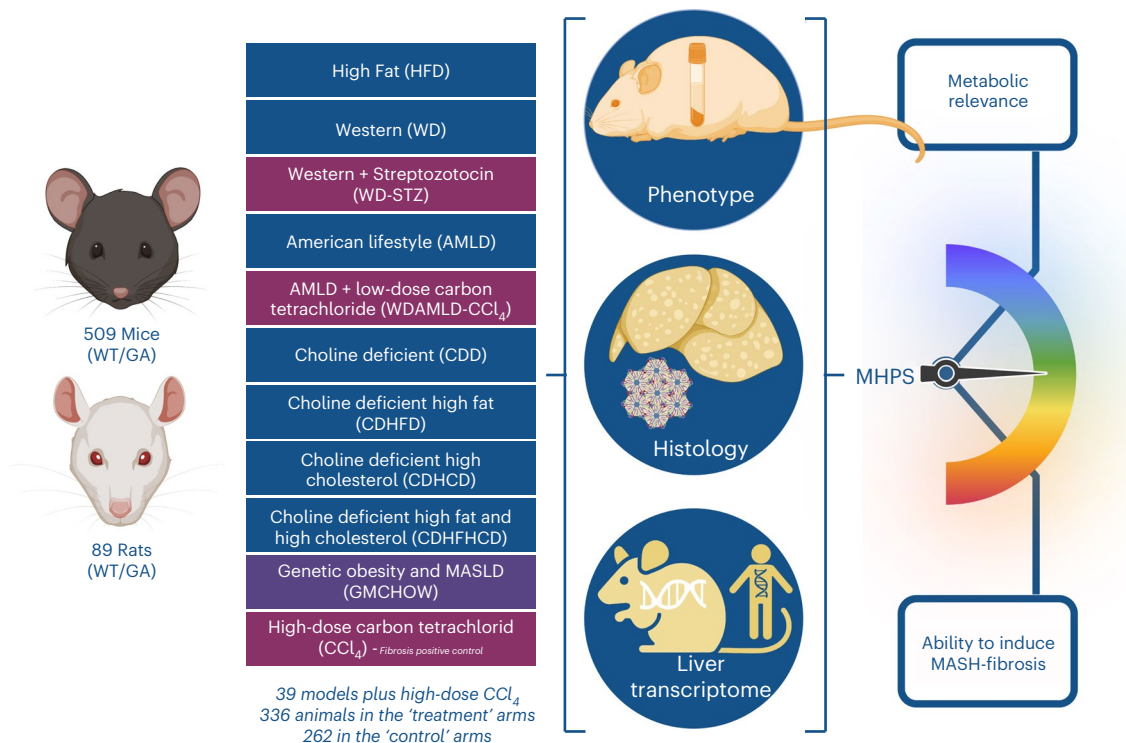
Fig. 1 | Study design. In this study, we collected retrospective information from 598 animals (509 WT/GA mice, 89 WT/GA rats): 336 animals subjected to treatment (MASLD-inducing conditions: 315 animals; or CCl₄: 21 animals) and 262 animals as controls for MASLD-inducing conditions (247 animals) or CCl₄ (15 animals), returning 39 models (that is, study designs) aimed at modeling MASLD, and two time points for CCl₄ (positive controls of MASLD-independent fibrosis). Details of the study designs (numerosity, species, background, genetic manipulation, diet, time point and room temperature) are provided in Supplementary Table 1. For all the studies, phenotypic information (Supplementary Table 5), centralized histopathology assessment (Supplementary Table 6) and liver transcriptomics (Supplementary Table 4) were available. These data were integrated into an unbiased binary score (MHPS) ranking the models in terms of their metabolic relevance and ability to induce MASH-fibrosis. Created with BioRender (agreement number GG26BHMS6Y).

composition have been tested in a wide range of rodent models aiming to feature the whole spectrum of metabolic disease and/or hepatic damage[10–12]. Disappointingly, within each diet macro-category, relatively subtle differences in micronutrient and macronutrient composition (for example, amount of cholesterol or choline) and study designs are sufficient to introduce variability in MASLD disease endpoints[12–16]. Several stressors have been used to accelerate disease progression and homogenize phenotypes, including genetically modified mice and rats (featuring obesity[17] or hypercholesterolemia[18,19]), chemically induced reduction of insulin secretory capacity[20] and/or the addition of toxic chemicals (for example, carbon tetrachloride, CCl₄ (ref. 21)), and the use of genetically modified mice that spontaneously develop progressive steatohepatitis (for example, NEMO, PTEN knockout (KO) mice[16,22]). Moreover, some rodent models seem to recapitulate the metabolic aspects of MASLD, whereas others are better at developing its fibro-inflammatory features[12,13,23]. Given the vast proliferation of models, variability and lack of standardization[12–16], a systematic comparison validating metabolic features, histology and transcriptomics against human disease is warranted but currently lacking.

To resolve the uncertainty about the most relevant preclinical mouse models, we have performed a wide-ranging retrospective analysis of the most common murine models used in academia and the pharmaceutical industry that were available to our consortium and collaborators, benchmarked against human MASLD and ranked according to the following characteristics: (1) obesity and/or metabolic syndrome; (2) development of steatohepatitis with progressive fibrosis (following hard outcomes of clinical trials defined for an amelioration, or at least a non-worsening, of fibrosis[24]); and (3) similarity of the histological features and molecular events to human MASH.

We developed a bioinformatic pipeline integrating metabolic phenotype data, liver histology (with centralized staining and assessment) and liver transcriptome benchmarked against human disease to create a MASLD 'human proximity score' (MHPS; Fig. 1). The MHPS was invaluable in generating a dual ranking of models based on their metabolic relevance and/or ability to induce MASH-fibrosis, highlighting the models that more closely resembled the metabolic and/or the fibrotic features of the disease. Our approach identified murine MASLD models showing (phenotypic and/or histologic and/or transcriptomic) profiles relevant to human MASH, which are suitable for most preclinical experimental applications.

Results

Main phenotypic and histologic attributes characterizing human MASLD in murine models

We collated retrospective data and samples from 39 commonly used murine genetic or dietary MASLD models (treatment: 315 animals; control diet: 247 animals; see Fig. 2 and Supplementary Tables 1 and 2 for more details) available to the consortium and/or collaborators, and clustered them according to macro-categories (diet and/or genetic background) as follows: (1) genetically modified models of obesity or MASLD; (2) high-fat diet (HFD); (3) Western (that is, atherogenic) diet (WD; HFD enriched in refined carbohydrates and cholesterol); these models were subclustered according to cholesterol concentration (0–2%) and/or the use of chemicals (streptozotocin, STZ); (4) American lifestyle diet (AMLD) including HFD (HFDAMLD) or WD (WDAMLD) supplemented with refined sugars in the drinking water, with or without the use of low-dose CCl₄; and (5) choline-deficient dietary models including 'canonical' choline-deficient diets (CDD) and choline-deficient diets supplemented with fat (CDHFD), cholesterol (CDHCD; 1% or 2%)
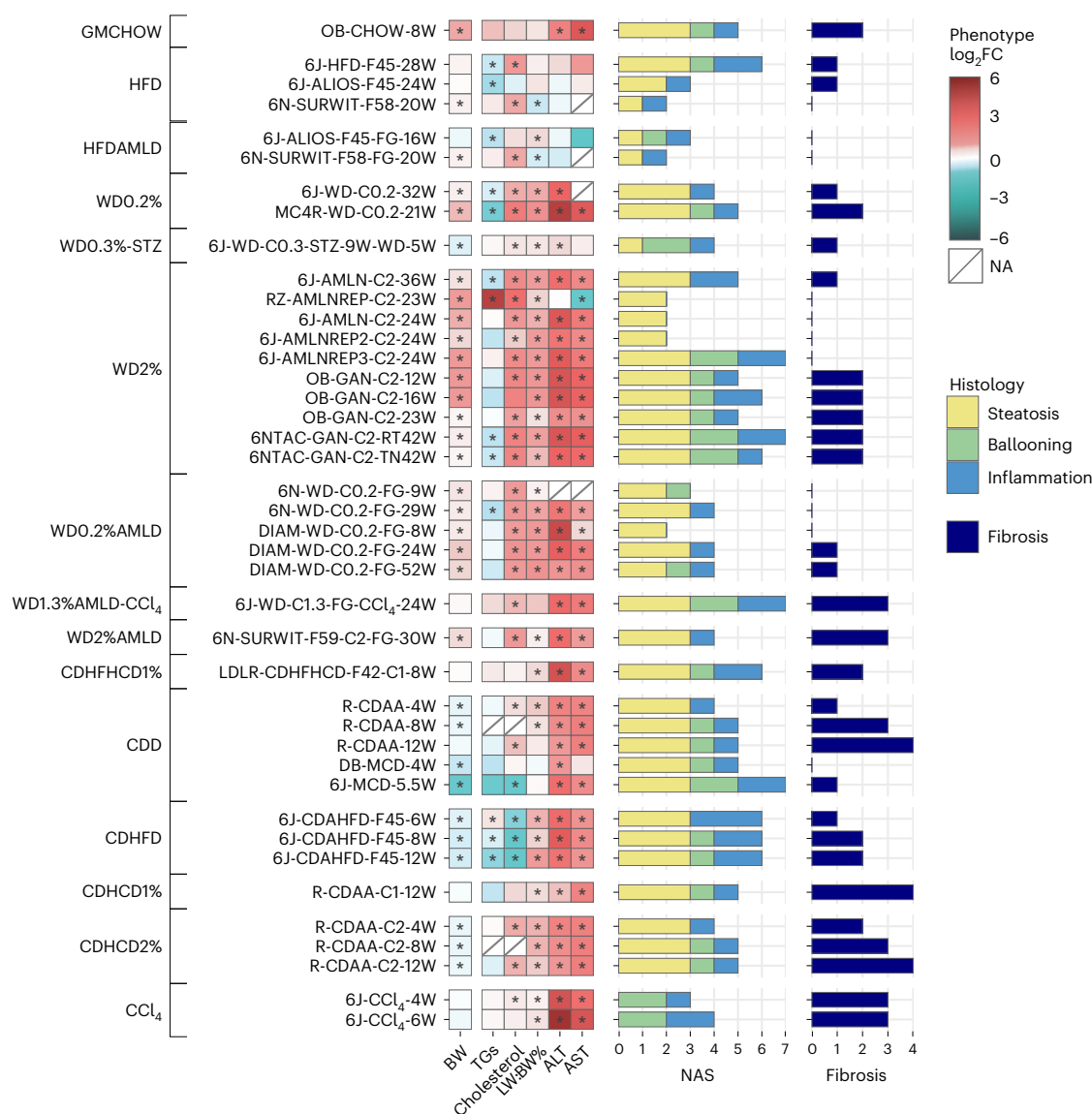
Fig. 2 | Phenotypic and histologic characterization of the models. Phenotypic changes observed in the MASLD models compared to their matched controls were profiled as the log$_2$ fold change (log$_2$FC) across measures of BW, blood triglycerides (TGs) and cholesterol, LW:BW% ratio, and ALT and AST. The red–blue color gradient indicates the level of increase–decrease of the measure in the MASLD models compared to their controls, while an asterisk indicates a significant change at $P < 0.05$ (two-sided Mann–Whitney U-test). The two panels of horizontal bars give an overview of the complete histological profiles, in which the total length indicates the activity score (CRN NAS) and fibrosis[25]. In addition, NAS components (steatosis, ballooning and inflammation) are represented by the stacked bar (yellow, green and blue, respectively) lengths. All models are grouped according to their macro-categories (detailed by the leftmost annotations).

or both (CDHFHCD). (6) As positive controls for isolated fibrosis, we used CCl$_4$-treated mice (CCl$_4$ treatment: 21 animals; control treatment: 15 animals; two time points).

The anticipated outcome of a MASLD murine model exhibiting body weight (BW) gain was readily achieved in genetically modified (leptin-deficient (ob/ob)) mice, WD and HFD models. Notably, the weight gain was less prominent in HFD models, and in WD or WDAMLD supplemented with chemicals (CCl$_4$ or STZ). By contrast, choline-deficient models generally reduced BW with the exception of CDHFHCD1% in low-density lipoprotein receptor (LDLR) KO mice. Most models did not significantly elevate circulating triglyceride levels, except for ZSF1 (diabetic) rats on a WD2% diet (RZ-AMLNREP-C2-23W) and a CDHFD mouse model. As expected, most HFD models and all diets with increased cholesterol content (0.2–2%) (including WD, WDAMLD,

CDHCD and CDHFHCD) as well as wild-type rats treated with CDAA exhibited hypercholesterolemia.

Apart from HFDs (and HFDAMLD), most experimental designs resulted in notable liver enlargement (that is, the ratio of liver weight to BW (LW:BW%)) and elevated aspartate aminotransferase (AST) and alanine aminotransferase (ALT) levels, especially at the final time point studied. Data on glucose metabolism was limited to a subset of our models (Supplementary Table 1). However, most HFD and WD diets examined have been documented in the literature to decrease insulin tolerance and impair glucose metabolism; conversely, choline-deficient models are not as well-characterized in this context (see Supplementary Table 2 for details and references).

To assess the histological characteristics of MASLD, the LIT-MUS ('Liver Investigation: Testing Marker Utility in Steatohepatitis'
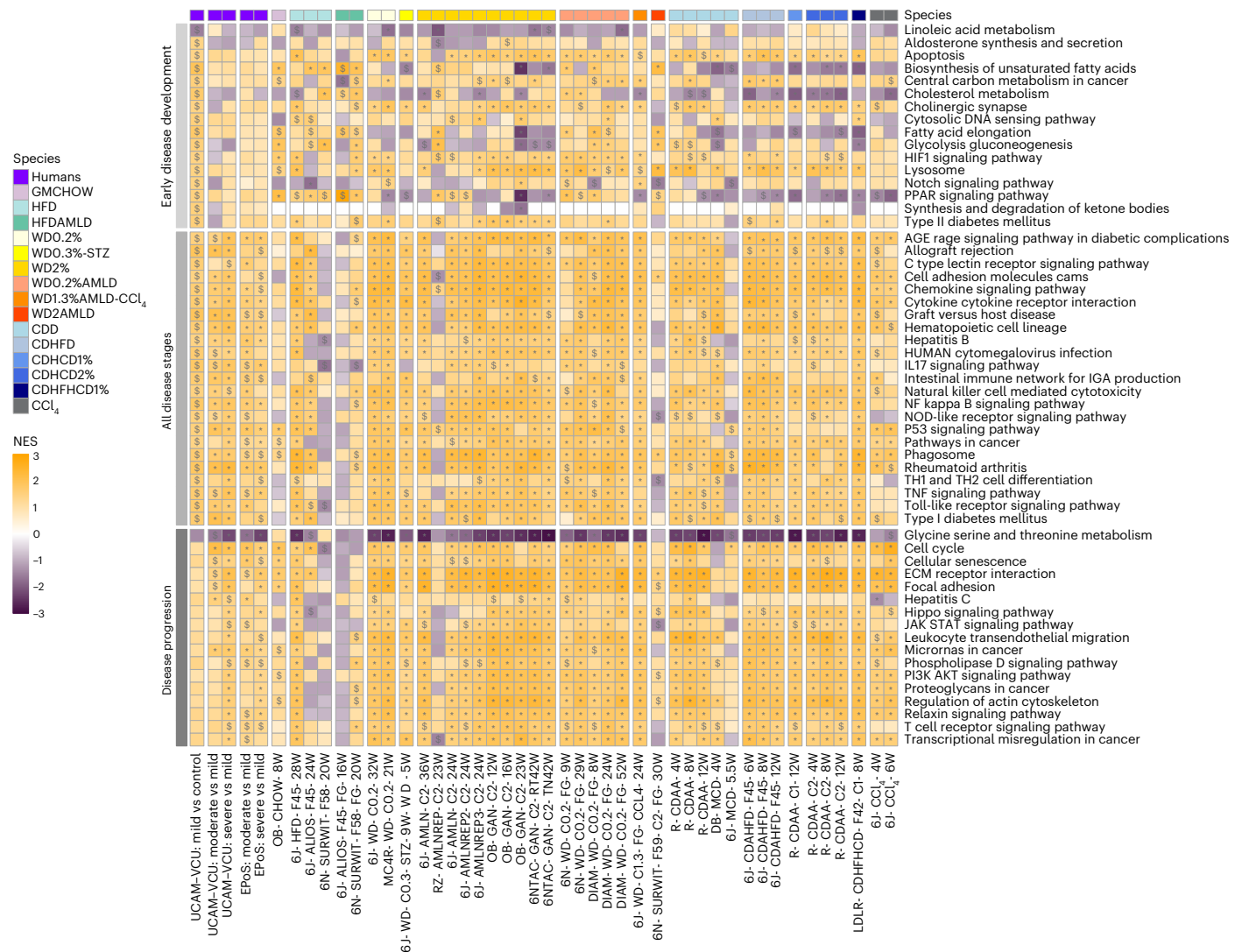
Fig. 3 | DRPs in human and murine MASLD. Selection of KEGG-affected pathways characterizing the 'early disease development', 'all disease stages' and 'disease progression' groups. The 'early disease development' group includes statistically significant ($P < 0.05$) modulated pathways in 'mild vs controls' but not in 'moderate–severe vs mild' human disease stages comparisons. The 'all disease stages' group includes statistically significant and homogeneously modulated pathways at all human disease stages ('mild vs controls' and 'moderate–severe vs mild' in both UCAM–VCU and EPoS); the 'disease progression' group includes homogeneously modulated and statistically significant pathways in 'moderate–severe vs mild' comparisons in both UCAM–VCU and EPoS but not in 'mild vs controls' comparisons. FGSEA calculated a normalized enrichment score (NES; the enrichment score normalized to mean enrichment of random samples of the same size) for each pathway and determined statistical significance using permutation testing (two-sided), adjusting for multiple comparisons to control the false discovery rate (FDR). The human and murine datasets are represented in a color-scale matrix showing the NES. * and $ symbols denote statistical significance ($, $P < 0.05$; *, FDR < 0.05). All models are grouped according to their macro-categories, as indicated by the panel on the top of the heatmap.

Consortium) Histopathology Group evaluated liver pathology using the Clinical Research Network (CRN)[25], and Steatosis, Activity and Fibrosis (SAF)[26] grading and staging systems. This evaluation was conducted on tissue sections centrally stained with hematoxylin and eosin (H&E) and Sirius Red. The NAFLD activity score (NAS) grading confirmed that most models induced moderate to severe steatosis (score 2–3) and mild to moderate lobular inflammation (score 1–2). Analysis of the HFD and HFDAMLD models showed relatively mild MASH activity (characterized by lobular inflammation and hepatocyte ballooning) and fibrosis, resulting in mild steatohepatitis[12,13]. Ballooning (score 1–2) was a prominent feature in choline-deficient models (CDD, CDHFD, CDHCD and CDHFHCD) and some WD models, especially those containing 2% cholesterol (WD2%); for instance, Gubra Amylin Diet (GAN), Amylin Liver NASH (AMLN-C2) or AMLN Replacement Diet 3 (AMLNREP3-C2). Cholesterol at relatively low concentrations (WD0.2%) induced ballooning

primarily on genetically modified obese mice (for example, MC4R KO mice), when supplemented with sugar water (WD0.2%AMLD) in both C57BL/6N and Diamond mice, or in those models additionally challenged with chemicals (STZ or CCl_4).

The highest NAS (score 6–7) was attained in models consuming diets containing 40–45% fat. This was observed across various macro-categories, including HFD-F45, AMLNREP3-C2, GAN-C2, CDAHFD-F45, the WD1.3%AMLD-CCl_4 model and CDHFHCD1% administered to LDLR KO mice, as well as in the methionine- and choline-deficient diet (MCD) model.

Notably, significant (F2 or higher) fibrosis was prevalent in choline-deficient dietary models (CDD, CDHFD, CDHCD and CDHFHCD1%). By contrast, fibrosis was absent in short-term MCD, indicating that a minimum duration of 8 weeks is necessary for MCD to induce MASH-associated fibrosis. Most animals fed with standard chow,
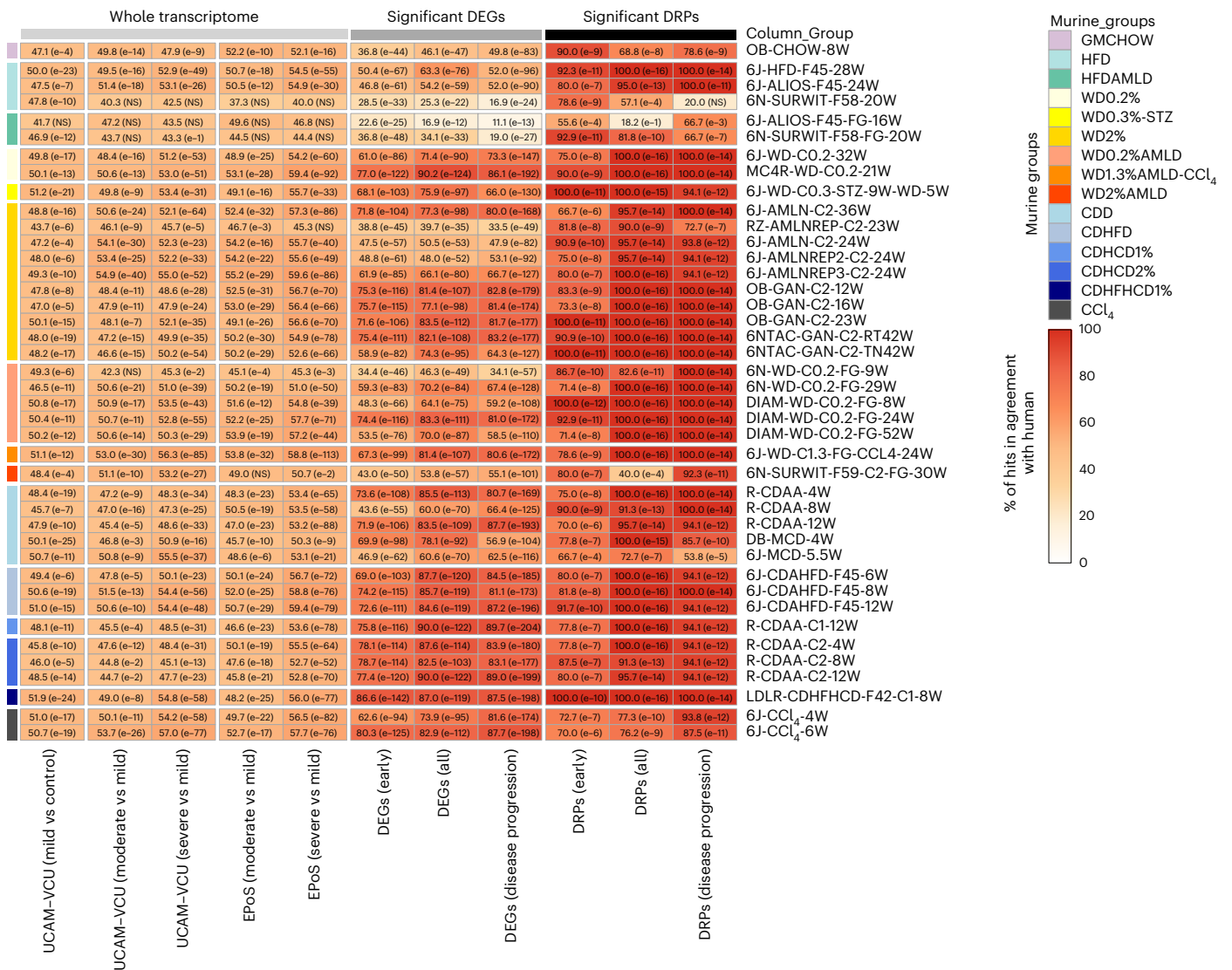
Fig. 4 | Agreement of murine DEGs and DRPs with human MASLD. Heatmap showing the agreement between murine MASLD models and human data based on the list of significant DEGs, DRPs, or the whole transcriptome. The percentage (%) of agreement between murine and human datasets defines the proportion of DEGs and DRPs statistically modulated and in the same direction compared to the human reference datasets (defined for 'early disease development', 'all disease stages' and 'disease progression' comparisons) or the proportion of DEGs statistically modulated and in the same direction compared to the human disease stage comparisons (defined for the whole transcriptome). All models are grouped according to their macro-categories as indicated in the graphic legend of the figure. Data are represented in a color-scale matrix showing the percentage of agreement and refer to DEGs (Supplementary Table 4; whole dataset or genes defined for 'early disease development', 'all disease stages' and 'disease progression', respectively) and DRPs (Fig. 3). In parenthesis, we show the results of the hypergeometric test (one-sided) performed on the same comparison groups, indicating the statistical significance (NS, non-significant (P > 0.05)).

Whole transcriptome columns: UCAM-VCU (mild vs control), UCAM-VCU (moderate vs mild), UCAM-VCU (severe vs mild), EPoS (moderate vs mild), EPoS (severe vs mild). Significant DEGs columns: DEGs (early), DEGs (all), DEGs (disease progression). Significant DRPs columns: DRPs (early), DRPs (all), DRPs (disease progression).

UCAM-VCU (mild vs control)	UCAM-VCU (moderate vs mild)	UCAM-VCU (severe vs mild)	EPoS (moderate vs mild)	EPoS (severe vs mild)	DEGs (early)	DEGs (all)	DEGs (disease progression)	DRPs (early)	DRPs (all)	DRPs (disease progression)	Column_Group
47.1 (e-4)	49.8 (e-14)	47.9 (e-9)	52.2 (e-10)	52.1 (e-16)	36.8 (e-44)	46.1 (e-47)	49.8 (e-83)	90.0 (e-9)	68.8 (e-8)	78.6 (e-9)	OB-CHOW-8W
50.0 (e-23)	49.5 (e-16)	52.9 (e-49)	50.7 (e-18)	54.5 (e-55)	50.4 (e-67)	63.3 (e-76)	52.0 (e-96)	92.3 (e-11)	100.0 (e-16)	100.0 (e-14)	6J-HFD-F45-28W
47.5 (e-7)	51.4 (e-18)	53.1 (e-26)	50.5 (e-12)	54.9 (e-30)	46.8 (e-61)	54.2 (e-59)	52.0 (e-90)	80.0 (e-7)	95.0 (e-13)	100.0 (e-11)	6J-ALIOS-F45-24W
47.8 (e-10)	40.3 (NS)	42.5 (NS)	37.3 (NS)	40.0 (NS)	28.5 (e-33)	25.3 (e-22)	16.9 (e-24)	78.6 (e-9)	57.1 (e-4)	20.0 (NS)	6N-SURWIT-F58-20W
41.7 (NS)	47.2 (NS)	43.5 (NS)	49.6 (NS)	46.8 (NS)	22.6 (e-25)	16.9 (e-12)	11.1 (e-13)	55.6 (e-4)	18.2 (e-1)	66.7 (e-3)	6J-ALIOS-F45-FG-16W
46.9 (e-12)	43.7 (NS)	43.3 (e-1)	44.5 (NS)	44.4 (NS)	36.8 (e-48)	34.1 (e-33)	19.0 (e-27)	92.9 (e-11)	81.8 (e-10)	66.7 (e-7)	6N-SURWIT-F58-FG-20W
49.8 (e-17)	48.4 (e-16)	51.2 (e-53)	48.9 (e-25)	54.2 (e-60)	61.0 (e-86)	71.4 (e-90)	73.3 (e-147)	75.0 (e-8)	100.0 (e-16)	100.0 (e-14)	6J-WD-C0.2-32W
50.1 (e-13)	50.6 (e-13)	53.0 (e-51)	53.1 (e-28)	59.4 (e-92)	77.0 (e-122)	90.2 (e-124)	86.1 (e-192)	90.0 (e-14)	100.0 (e-16)	100.0 (e-14)	MC4R-WD-C0.2-21W
51.2 (e-21)	49.8 (e-9)	53.4 (e-31)	49.1 (e-16)	55.7 (e-33)	68.1 (e-103)	75.9 (e-97)	66.0 (e-130)	100.0 (e-11)	100.0 (e-15)	94.1 (e-12)	6J-WD-C0.3-STZ-9W-WD-5W
48.8 (e-6)	50.6 (e-24)	52.1 (e-64)	52.4 (e-32)	57.3 (e-86)	71.8 (e-104)	77.3 (e-98)	80.0 (e-168)	66.7 (e-6)	95.7 (e-14)	100.0 (e-14)	6J-AMLN-C2-36W
43.7 (e-6)	46.1 (e-9)	45.7 (e-5)	46.7 (e-3)	45.3 (NS)	38.8 (e-35)	39.7 (e-35)	33.5 (e-49)	81.8 (e-8)	90.0 (e-7)	72.7 (e-7)	RZ-AMLNREP-C2-23W
47.2 (e-4)	54.1 (e-30)	52.3 (e-23)	54.2 (e-41)	55.7 (e-40)	47.5 (e-57)	50.5 (e-53)	47.9 (e-82)	90.9 (e-10)	95.7 (e-14)	93.8 (e-12)	6J-AMLN-C2-24W
48.0 (e-6)	53.4 (e-25)	52.2 (e-33)	54.2 (e-22)	55.6 (e-49)	48.8 (e-61)	48.0 (e-52)	53.1 (e-92)	75.0 (e-8)	95.7 (e-14)	94.1 (e-12)	6J-AMLNREP2-C2-24W
49.3 (e-10)	54.9 (e-40)	55.0 (e-52)	55.2 (e-29)	59.6 (e-86)	61.9 (e-85)	66.1 (e-80)	66.7 (e-127)	80.0 (e-7)	100.0 (e-16)	94.1 (e-12)	6J-AMLNREP3-C2-24W
47.8 (e-8)	48.4 (e-11)	48.6 (e-28)	52.5 (e-31)	56.7 (e-70)	75.3 (e-116)	81.4 (e-107)	82.8 (e-179)	83.3 (e-9)	100.0 (e-16)	100.0 (e-14)	OB-GAN-C2-12W
47.0 (e-5)	47.9 (e-11)	47.9 (e-24)	53.0 (e-29)	56.4 (e-66)	75.7 (e-115)	77.1 (e-98)	81.4 (e-174)	73.3 (e-8)	100.0 (e-16)	100.0 (e-14)	OB-GAN-C2-16W
50.1 (e-15)	48.1 (e-7)	52.1 (e-35)	49.1 (e-26)	56.6 (e-70)	71.6 (e-106)	83.5 (e-112)	81.7 (e-177)	100.0 (e-11)	100.0 (e-16)	100.0 (e-14)	OB-GAN-C2-23W
48.0 (e-19)	47.2 (e-15)	49.9 (e-54)	50.2 (e-30)	54.9 (e-78)	75.4 (e-115)	82.1 (e-108)	83.2 (e-177)	90.9 (e-10)	100.0 (e-16)	100.0 (e-14)	6NTAC-GAN-C2-RT42W
48.2 (e-17)	46.6 (e-15)	50.2 (e-54)	50.2 (e-29)	52.6 (e-66)	58.9 (e-82)	74.3 (e-95)	64.3 (e-127)	100.0 (e-11)	100.0 (e-16)	100.0 (e-14)	6NTAC-GAN-C2-TN42W
49.3 (e-6)	42.3 (NS)	45.3 (e-2)	45.1 (e-4)	45.3 (e-3)	34.4 (e-46)	46.3 (e-49)	34.1 (e-57)	86.7 (e-10)	82.6 (e-11)	100.0 (e-14)	6N-WD-C0.2-FG-9W
46.5 (e-6)	50.6 (e-21)	51.0 (e-9)	50.2 (e-19)	51.0 (e-50)	59.3 (e-83)	70.2 (e-84)	67.4 (e-128)	71.4 (e-8)	100.0 (e-16)	100.0 (e-14)	6N-WD-C0.2-FG-29W
50.8 (e-17)	50.9 (e-17)	53.5 (e-43)	51.6 (e-12)	54.8 (e-39)	48.3 (e-66)	64.1 (e-75)	59.2 (e-108)	100.0 (e-12)	100.0 (e-16)	100.0 (e-14)	DIAM-WD-C0.2-FG-8W
50.4 (e-11)	50.7 (e-11)	52.8 (e-55)	52.2 (e-25)	57.7 (e-71)	74.4 (e-116)	83.3 (e-111)	81.0 (e-172)	92.9 (e-11)	100.0 (e-16)	100.0 (e-14)	DIAM-WD-C0.2-FG-24W
50.2 (e-12)	50.6 (e-14)	50.3 (e-29)	53.9 (e-19)	57.2 (e-44)	53.5 (e-76)	70.0 (e-87)	58.5 (e-110)	71.4 (e-8)	100.0 (e-16)	100.0 (e-14)	DIAM-WD-C0.2-FG-52W
51.1 (e-12)	53.0 (e-30)	56.3 (e-85)	53.8 (e-32)	58.8 (e-113)	67.3 (e-99)	81.4 (e-107)	80.6 (e-172)	78.6 (e-9)	100.0 (e-16)	100.0 (e-14)	6J-WD-C1.3-FG-CCL4-24W
48.4 (e-4)	51.1 (e-10)	53.2 (e-27)	49.0 (NS)	50.7 (e-2)	43.0 (e-50)	53.8 (e-57)	55.1 (e-101)	80.0 (e-7)	40.0 (e-4)	92.3 (e-11)	6N-SURWIT-F59-C2-FG-30W
48.4 (e-19)	47.2 (e-9)	48.3 (e-34)	48.3 (e-23)	53.4 (e-65)	73.6 (e-108)	85.5 (e-113)	80.7 (e-169)	75.0 (e-8)	100.0 (e-16)	100.0 (e-14)	R-CDAA-4W
45.7 (e-7)	47.0 (e-16)	47.3 (e-25)	50.5 (e-19)	53.5 (e-58)	43.6 (e-55)	60.0 (e-70)	66.4 (e-125)	90.0 (e-9)	91.3 (e-13)	100.0 (e-14)	R-CDAA-8W
47.9 (e-10)	45.4 (e-5)	48.6 (e-33)	47.0 (e-23)	53.2 (e-88)	71.9 (e-104)	83.5 (e-109)	87.7 (e-193)	70.0 (e-6)	95.7 (e-14)	94.1 (e-12)	R-CDAA-12W
50.1 (e-25)	46.8 (e-3)	50.9 (e-16)	45.7 (e-10)	50.3 (e-9)	69.9 (e-98)	78.1 (e-92)	56.9 (e-104)	77.8 (e-7)	100.0 (e-15)	85.7 (e-10)	DB-MCD-4W
50.7 (e-11)	50.8 (e-9)	55.5 (e-37)	48.6 (e-6)	46.9 (e-62)	60.6 (e-70)	60.6 (e-70)	62.5 (e-116)	66.7 (e-4)	72.7 (e-7)	53.8 (e-5)	6J-MCD-5.5W
49.4 (e-6)	47.8 (e-5)	50.1 (e-24)	50.1 (e-24)	56.7 (e-72)	69.0 (e-104)	87.7 (e-120)	84.5 (e-185)	80.0 (e-7)	100.0 (e-16)	94.1 (e-12)	6J-CDAHFD-F45-6W
50.6 (e-19)	51.5 (e-13)	54.4 (e-56)	52.0 (e-25)	58.8 (e-76)	74.2 (e-115)	85.7 (e-119)	81.1 (e-173)	81.8 (e-8)	100.0 (e-16)	100.0 (e-14)	6J-CDAHFD-F45-8W
51.0 (e-15)	50.6 (e-10)	54.4 (e-48)	50.7 (e-29)	59.4 (e-79)	72.6 (e-111)	84.6 (e-119)	87.2 (e-196)	91.7 (e-10)	100.0 (e-16)	94.1 (e-12)	6J-CDAHFD-F45-12W
48.1 (e-11)	45.5 (e-4)	48.5 (e-31)	46.6 (e-23)	53.6 (e-78)	75.8 (e-116)	90.0 (e-122)	89.7 (e-204)	77.8 (e-7)	100.0 (e-16)	94.1 (e-12)	R-CDAA-C1-12W
45.8 (e-10)	47.6 (e-12)	48.4 (e-31)	50.1 (e-19)	55.5 (e-64)	78.1 (e-114)	87.6 (e-114)	83.9 (e-180)	77.8 (e-7)	100.0 (e-16)	94.1 (e-12)	R-CDAA-C2-4W
46.0 (e-5)	44.8 (e-2)	45.1 (e-13)	47.6 (e-18)	52.7 (e-52)	78.7 (e-114)	82.5 (e-103)	83.1 (e-177)	87.5 (e-7)	91.3 (e-13)	94.1 (e-12)	R-CDAA-C2-8W
48.5 (e-4)	44.7 (e-2)	47.7 (e-23)	45.8 (e-21)	52.8 (e-70)	77.4 (e-120)	90.0 (e-122)	89.0 (e-199)	80.0 (e-7)	95.7 (e-14)	94.1 (e-12)	R-CDAA-C2-12W
51.9 (e-24)	49.0 (e-8)	54.8 (e-58)	48.2 (e-25)	56.0 (e-77)	86.6 (e-142)	87.0 (e-119)	87.5 (e-198)	100.0 (e-10)	100.0 (e-16)	100.0 (e-14)	LDLR-CDHFHCD-F42-C1-8W
51.0 (e-17)	50.1 (e-11)	54.2 (e-58)	49.7 (e-22)	56.5 (e-82)	62.6 (e-94)	73.9 (e-95)	81.6 (e-174)	72.7 (e-7)	77.3 (e-10)	93.8 (e-12)	6J-CCl$_4$-4W
50.7 (e-19)	53.7 (e-26)	57.0 (e-77)	52.7 (e-17)	57.7 (e-76)	80.3 (e-125)	82.9 (e-112)	87.7 (e-198)	70.0 (e-6)	76.2 (e-9)	87.5 (e-11)	6J-CCl$_4$-6W

Murine_groups legend: GMCHOW, HFD, HFDAMLD, WD0.2%, WD0.3%-STZ, WD2%, WD0.2%AMLD, WD1.3%AMLD-CCl$_4$, WD2%AMLD, CDD, CDHFD, CDHCD1%, CDHCD2%, CDHFHCD1%, CCl$_4$.

% of hits in agreement with human: 100, 80, 60, 40, 20, 0.

HFD, HFDAMLD, WD or WDAMLD did not develop significant (F2 or higher) fibrosis even after prolonged challenges (lasting 36–52 weeks). However, significant fibrosis was observed in C57BL6/N mice on a WD2%AMLD diet (6N-SURWIT-F59-C2-FG-30W), C57BL/6NTac or ob/ob mice fed the GAN-C2 diet, WD1.3%AMLD-CCl$_4$ and MC4R KO mice on a WD0.2% diet.

Our findings suggest that no single model currently replicates all the phenotypic and histological characteristics of human MASLD among the models we evaluated. The HFD, WD and AMLD models, with or without chemical supplementation, are broadly effective in simulating the metabolic aspects of MASLD. However, these diets typically result in a milder histological phenotype, with some notable exceptions. Conversely, the choline-deficient dietary models (CDD, CDHFD, CDHCD and CDHFHCD) rapidly (within 12 weeks or less) lead to the development of MASH with significant fibrosis. However, they fall short in accurately modeling the metabolic burden of MASLD owing to BW loss and improved dyslipidemia. Exceptions to this trend include CDHFHCD-fed LDLR KO mice and certain CDHCD-fed rat models.

Murine MASLD transcriptomes are close to humans but do not predict fibrosis efficiently

Using publicly accessible human NAFLD/MASLD transcriptomes described by our teams[27–29], we collated two datasets comprising 136 (University of Cambridge (UCAM) and Virginia Commonwealth University (VCU)) and 168 (Newcastle University (EPoS)) patients, as detailed in Supplementary Table 3. We aimed to identify differentially expressed genes (DEGs) and differentially regulated pathways (DRPs) that define MASLD. We identified specific pathways (DRPs; illustrated in Fig. 3) and DEGs (listed in Supplementary Table 4) that are prevalent in the early stages of the disease (that is, mild MASLD vs control), throughout all stages of the disease (that is, across all comparisons within the two datasets) and during the progression of fibrosis (that is, moderate–severe vs

Fig. 5 | Highly performing genes in human and murine MASLD. Selection of statistically significant DEGs (complete list in Supplementary Table 4) enriching statistically significant modulated pathways (complete list in Fig. 3), thus being highly biologically relevant genes associated with the different stages of MASLD development or progression. To assess the statistical significance between the compared groups, a Wald test statistic (two-sided hypothesis testing) was deployed to compare the coefficients of explanatory variables in a regression model, representing the gene expression differences among the compared groups. The human and murine datasets are represented in a color-scale matrix showing the log2FC. * and $ symbols denote statistical significance ($, $P < 0.05$; *, adjusted $P < 0.05$ (Benjamini–Hochberg correction)). As the graphic legend indicates, all models were grouped according to their macro-categories.

mild MASLD). These human datasets served as benchmarks for comparing changes in murine transcriptomics, with the replicability of these findings across various datasets detailed in Supplementary Fig. 1. In our analysis, we observed that diets inducing MASLD formed distinct clusters from control diets when analyzed through the first principal component in mice (Extended Data Figure 1) and in rats (Extended Data Figure 2). Additionally, genetically obese mouse models (ob/ob; leptin receptor-deficient (db/db)) were differentiated from other data on the second principal component, as depicted in Extended Data Figure 1.

The murine models were compared to human NAFLD/MASLD transcriptomic data to evaluate their alignment in terms of DEGs and DRPs, as illustrated in Fig. 4. At the whole transcriptome level, the comparison of DEGs between murine models and human data showed relatively modest alignment rates (approximately 50%) when considering only fold changes and the direction of regulation for all models. However, when focusing on statistically significant findings in humans (DEGs and DRPs) characterizing 'early disease development', 'all disease stages' and 'disease progression' (as detailed in the Methods), we observed an improved and statistically significant agreement for the majority of murine models. It is important to note that although variable responses were at the individual transcript level among different models, this variability was not mirrored in the DRPs analysis. The DRPs (shown in Fig. 3) demonstrated a high level of agreement with human data, with only a few exceptions, as detailed in Fig. 4.

This analysis indicates that despite variations in specific single gene expression, the biological pathways implicated in the disease were consistently regulated across both the animal models and human cases. Generally, the pathways that characterize 'all disease stages' (which include wound healing pathways such as inflammation and cell proliferation) and those that define MASH 'disease progression' (which include wound healing pathways such as stellate cell activation, fibrosis, and carcinogenesis) were regulated similarly in humans and in most murine models. Of relevance, significant transcriptional changes associated with 'disease progression' in humans were also noted in murine models that paradoxically only exhibited mild (F0–F1) fibrosis. These findings underscore the critical importance of conducting liver histopathology in rodents. Relying solely on transcriptomic data is inadequate, as it does not fully capture the complex pathophysiology of fibrosis or accurately predict extracellular matrix deposition.

Similar to human cases, most murine models displayed significant regulation in the pathways associated with 'early disease development', which are involved in the pathophysiology of type 2 diabetes, hypoxia (HIF1 signaling) and lysosome function. However, there were notable species-specific differences in regulating cholesterol and lipid metabolism pathways among different macro-clusters, as detailed in Figs. 3 and 4. HFD and WD models aligned more with human data than choline-deficient models, particularly in linoleic acid metabolism, unsaturated fatty acid biosynthesis, cholesterol metabolism and PPAR signaling.

The addition of sugar water to HFDs or WDs, with or without CCl_4, led to profiles more closely aligned with humans regarding glucose metabolism dysregulation (glycolysis or gluconeogenesis) and
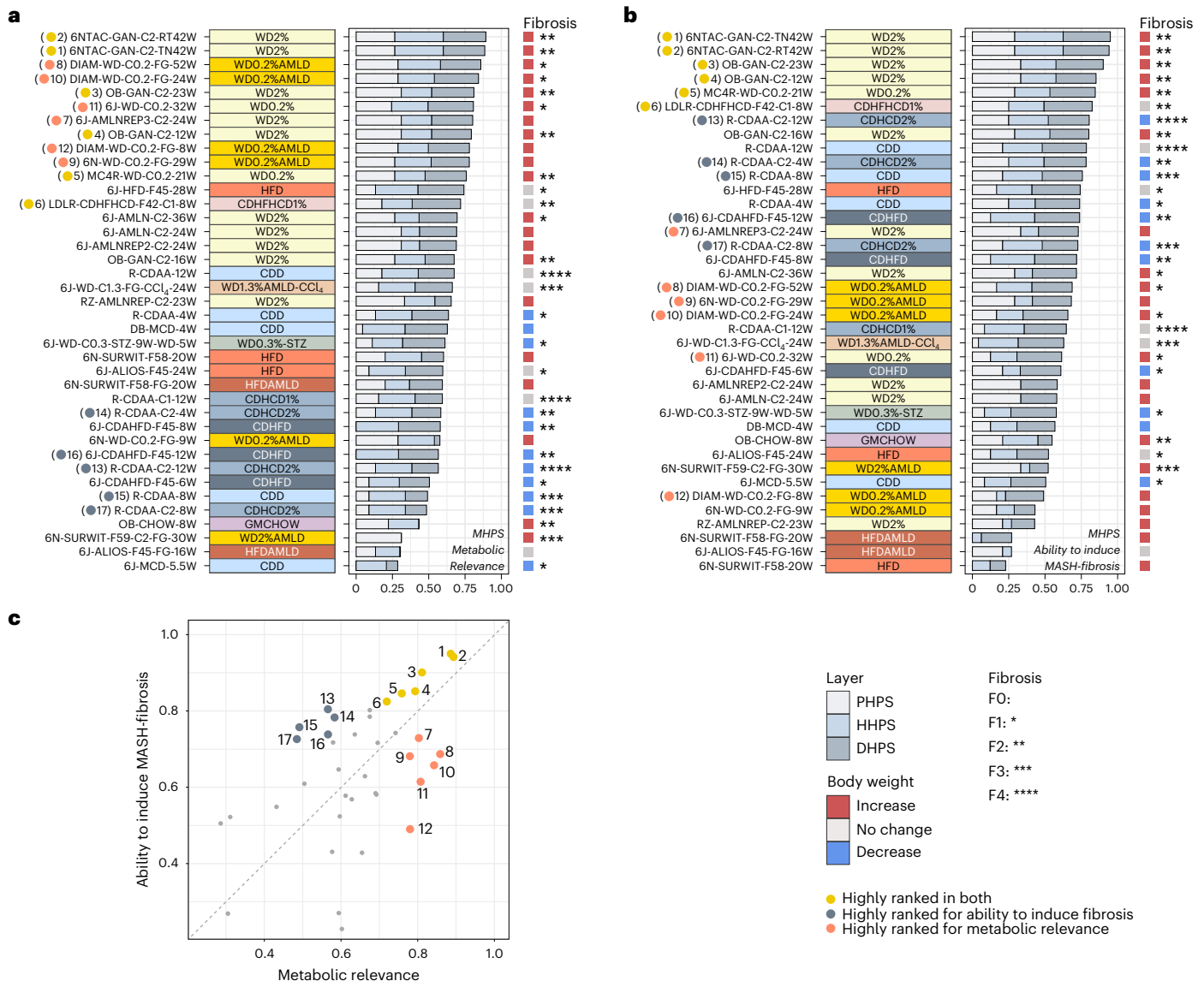
Fig. 6 | The MHPS–metabolic relevance and progressive MASLD.
a,**b**, Comparison of the MASLD models, performed based on the MHPS that incorporates the PHPS (details in Supplementary Table 5), the HHPS (details in Supplementary Table 6) and the DHPS (see Supplementary Fig. 4). The average of these normalized scores (MHPS) ranks the murine models (from high to low) based on their metabolic relevance (**a**) or their ability to induce MASH-fibrosis (**b**). A detailed description of the different components is provided in Extended Data Figs. 3 and 4, respectively. For both **a** and **b**, the total length of the horizontal bars indicates the MHPS, while the length of the stacks within each bar indicates the relative contribution from the three evidence layers: PHPS, HHPS and

DHPS. A reference panel to the right indicates BW (significant increases in red; decreases in blue) and fibrosis score (*). Macro-categories are indicated by the color panel to the left of the plots. **c**, Correlation among the two MHPS outputs ('metabolic relevance' vs 'ability to induce MASH-fibrosis'). Specific models are highlighted based on their performance within the two rankings. Yellow dots represent models that score high with both rankings and represent the best approximation to human MASH. Red dots are models that score highly for metabolic relevance but are less relevant for MASH-fibrosis. Grey dots are models that score highly for MASH-fibrosis but have less metabolic relevance. Panels **a** and **b** provide a specific reference to the position in the scatter.

lipid remodeling (fatty acid elongation). These molecular changes are probably attributable to the sugar water promoting the ChREBP/SREBP1-mediated de novo lipogenesis pathway[7,27,30]. Despite these similarities, very few models (mainly HFD, WD and HFD/WD2%-AMLD and LDLR-CDHFHCD) closely recapitulated (with >90% agreement) the burden of 'early disease development' (metabolic) DRP changes observed in human subjects.

Applying more restrictive filtering to focus on significant DEGs enriched in significant pathways (see Fig. 5), we identified a subset closely linked to various stages of MASLD development and progression. This selected list of biologically relevant genes includes enzymes, cytochromes, transporters, receptors, signal transducers, adaptor proteins and proteins involved in inflammation, extracellular

matrix remodeling and cell proliferation or differentiation. It demonstrates that WD, WDAMLD and most choline-deficient dietary models achieve the highest congruence with human MASLD at the transcriptomic level.

These transcriptomics data suggest that most models (apart from GMCHOW, HFD and HFDAMLD, and some choline-deficient models) generally replicate the gene expression dysregulation observed in progressive human MASH. However, human fibro-inflammatory-related transcriptional patterns did not consistently predict murine histological fibrosis, and metabolic pathways did not entirely align with human MASLD. Notably, the analysis of diet macro-clusters and specific diet formulations within the same cluster revealed significant differences in their 'early disease' (metabolic) transcriptional responses.
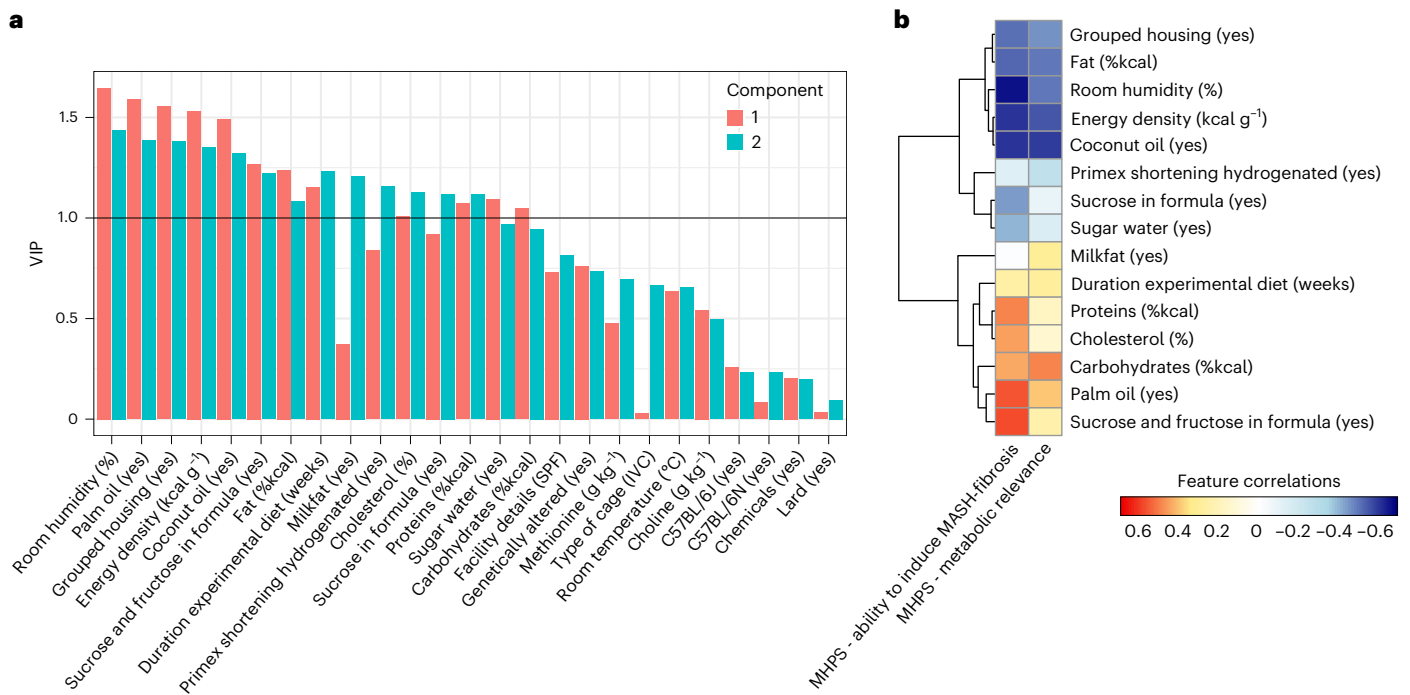
Fig. 7 | Key experimental components contributing to MHPS 'metabolic relevance' and 'ability to induce MASH-fibrosis' outputs. Relationship between study design parameters and the MHPS of metabolic relevance and ability to induce MASH-fibrosis evaluated by PLSR. **a**, VIP for each study parameter (among those detailed in Supplementary Table 1) for the two components of the model. Parameters contributing the most to the PLSR model are characterized by having VIPs > 1 (black solid line) in any of the two components of the model. **b**, Clustered heatmap of the most influential study parameters (VIP > 1), indicating their correlation with the MHPS metabolic relevance and ability to induce MASH-fibrosis (color-scale matrix showing a positive correlation represented in red; negative correlation represented in blue).

This underscores the significance of dietary composition, especially in experiments targeting specific metabolic processes.

The MHPS is a tool to rank murine models mimicking human MASLD

The typical subjective method of choosing a murine model for MASLD research is often influenced by standard laboratory practices and constrained by available resources. To provide a more reliable approach for selecting the most suitable models for preclinical research, we developed the MHPS. This data-driven bioinformatics pipeline integrates transcriptomic (drug set enrichment analysis (DSEA) human proximity score, DHPS; referenced in Supplementary Table 4), phenotypic (phenotype human proximity score, PHPS; detailed in Supplementary Table 5) and histopathologic (histopathology human proximity score, HHPS; outlined in Supplementary Table 6) comparative analyses of murine models with human disease. Consequently, the MHPS system ranks murine models based on their congruence with human MASLD in terms of metabolic (illustrated in Fig. 6a; details in Extended Data Figure 3) and/or fibro-inflammatory (shown in Fig. 6b; details in Extended Data Figure 4) profiles.

The MHPS identified the WD0.2%, WD2% and WD0.2%AMLD diets as the most effective in mirroring the metabolic burden of human MASLD (Fig. 6a and Extended Data Figure 3). These diets were highly scored for their role in inducing obesity and dyslipidemia, liver damage (LW:BW% ratio, increased AST and ALT), histological activity (development of steatosis, lobular inflammation and/or ballooning; Fig. 2) and gene expression (Fig. 4 and Supplementary Table 4; under 'early disease development' and 'all disease stages'). The GAN-C2 (ref. 17) (ambient and thermoneutral conditions) and the Diamond WD0.2%AMLD models ranked highest. Genetically modified animals (ob/ob, MC4R KO) fed with WDs and the WD1.3%AMLD-CCl₄ model also scored well, with the additional benefit of a shorter induction period (21–24 weeks) required

to develop significant (F2+) fibrosis. Except for CDHFHCD1%-fed LDLR KO mice (8 weeks) and CDAA-fed rats (12 weeks), most choline-deficient dietary models had a lower metabolic ranking, indicating their limited capability to fully replicate the metabolic burden of human MASH.

We then applied the MHPS to evaluate the progression of fibrosing MASH (Fig. 6b and Extended Data Figure 4). This assessment included factors such as liver damage (LW:BW% ratio and increased AST and ALT), histology (MASH with significant fibrosis; Fig. 2) and gene expression patterns outlined in Fig. 4 and Supplementary Table 4, specifically focusing on 'all disease stages' and 'disease progression'. The MHPS highly ranked models like the GAN-C2 in C57BL6/NTac and ob/ob mice, MC4R KO mice on a WD0.2% diet, CDHFHCD1%-fed LDLR KO mice (over 8 weeks) and CDHCD2% fed rats (over 12 weeks); in terms of liver histopathology outcomes, each of these models developed significant (F2–F4) fibrosis. The only diet inducing F4 fibrosis within a 12-week period was CDAA (with or without 1–2% cholesterol) in rats. Despite its high performance in individual MHPS components such as HHPS (developing F3 fibrosis) and DHPS, WD1.3%AMLD-CCl₄ (ref. 21) was penalized in the overall MHPS because of variability in mouse phenotypes reflected in the PHPS. Most other HFD, HFDAMLD, WD and WDAMLD models scored lower in this ranking, partly due to limited fibrosis. Notably, most choline-deficient models (CDD, CDHCD, CDHFD) did not rank highly despite inducing significant (F2–F4) fibrosis when considering all MHPS components together.

Subsequently, we compared the two sets of rankings to identify study designs that most effectively model the complete spectrum of MASLD features, as shown in Fig. 6c. Our comparison confirmed that WDs predominantly replicated the metabolic aspects of the disease, while choline-deficient diets better featured the pro-fibrotic components of MASH. Interestingly, several specific murine models scored highly in both rankings, making them the best choices for replicating both metabolic and fibro-inflammatory characteristics of human
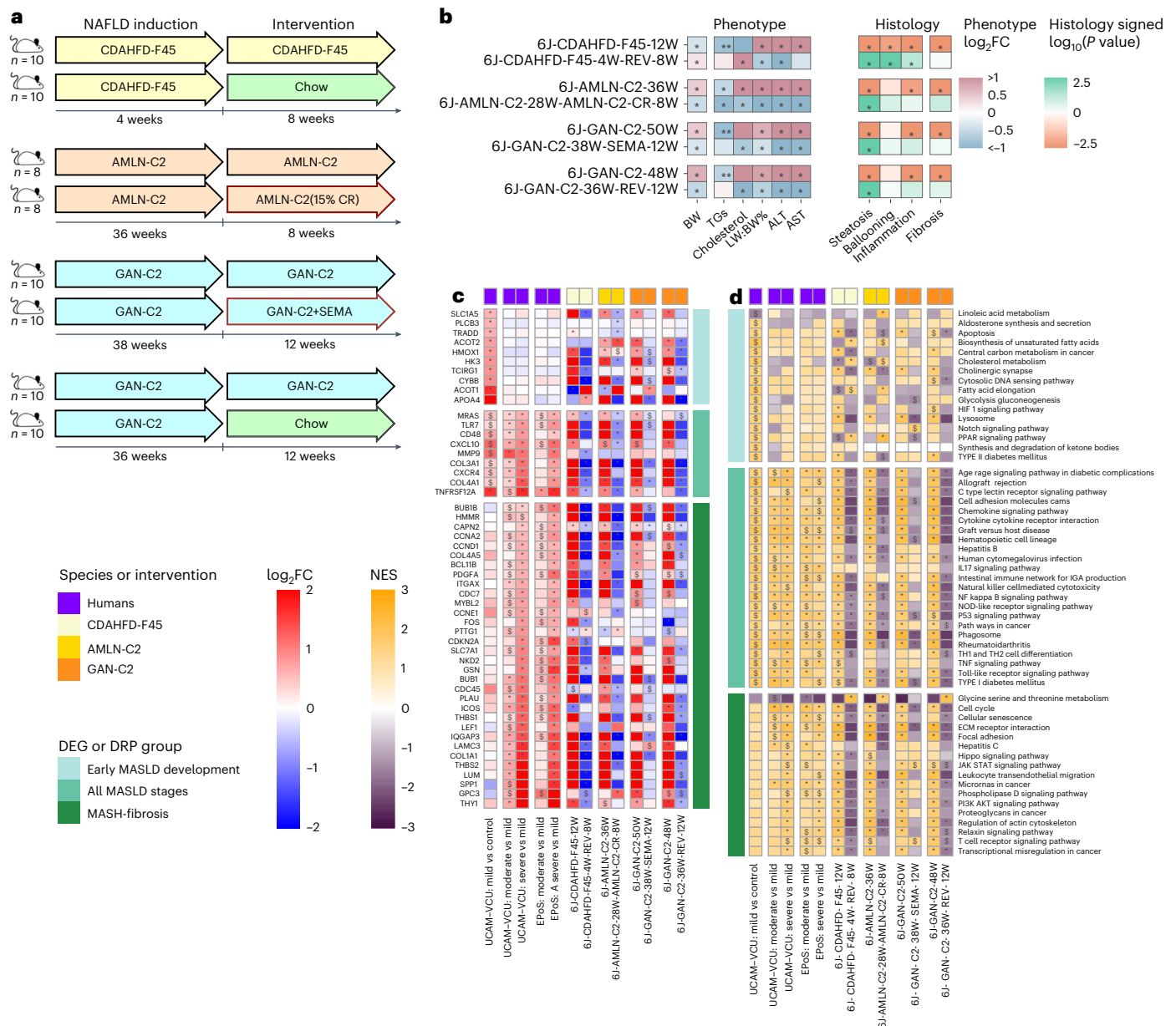
Fig. 8 | Effects of treatments in a selection of WD and choline-deficient models. Response of the best-ranked diet (GAN-C2), another WD (AMLN-C2) and a CDHFD (CDAHFD-F45) to treatments mimicking lifestyle intervention (CR, caloric restriction; REV, chow reversal) and semaglutide pharmacological treatment (SEMA, 30 nmol per kg per day). **a**, Simplified study designs (details in Supplementary Table 7). **b**, Effect on the phenotype and histology. Phenotypic changes observed in the treatment or dietary models compared to their matched dietary models or controls, respectively, were profiled as log2FC across measures of BW, blood TGs and cholesterol, LW:BW%, and ALT and AST. The red–blue color gradient indicates the level of increase–decrease of the measure in the models compared to their respective controls; * indicates statistical significance ($P < 0.05$; two-sided Mann–Whitney U-test). For the histological changes,

the Mann–Whitney U-test was used to calculate P values for the differences in the ordinal scores ($P < 0.05$ are shown with *). The color scale indicates the signed P value: $-\log_{10}(P\text{ value})$ for up-regulation and $+\log_{10}(P\text{ value})$ for down-regulation. **c**, Effect of treatments on a selection of biologically relevant DEGs as described in Fig. 5. The human and murine datasets are represented in a color-scale matrix showing the log2FC. * and $ symbols denote statistical significance (two-sided Wald test statistic and adjustment for multiple testing using the Benjamini–Hochberg correction; $: $P < 0.05$; *: adjusted $P < 0.05$). **d**, Effect of treatments on pathways as described in Fig. 3. FGSEA calculated a NES for each pathway and determined statistical significance using permutation testing (two-sided), adjusting for multiple comparisons to control the FDR. The human and murine datasets are represented in a color-scale matrix showing the NES. $, $P < 0.05$; *, FDR < 0.05.

MASLD. These standout models included C57BL6/NTac or ob/ob mice on a GAN-C2 diet, MC4R KO mice on a WD0.2% diet and LDLR KO mice undergoing a CDHFHCD1% diet challenge. Representative histopathology images of these highly ranked models are provided in Extended Data Figs. 5–10.

In conclusion, although no single model perfectly matched human MASLD, our analysis identified several study designs that closely approximate various aspects of the disease burden.

Key experimental components contributing to MHPS 'metabolic relevance' and 'ability to induce MASH-fibrosis' outputs

To gain a deeper understanding of the specific parameters of the study designs (outlined in Supplementary Table 1) that influenced the rankings of the murine models, we used a partial least squares regression (PLSR) analysis. This analysis explored the connection between the study designs and the MHPS outputs 'metabolic relevance' and

'ability to induce MASH-fibrosis'. Our focus was on mouse models fed choline-sufficient diets, as the available data for rat models and choline-deficient dietary models lacked the necessary detail and granularity required for a robust PLSR analysis.

The variable importance in projection (VIP) score was used to identify the most influential parameters in the PLSR model. We set a VIP threshold of at least 1 in one component (Fig. 7a) to determine these key parameters. These were then further examined for their correlation with the MHPS outputs 'metabolic relevance' and 'ability to induce MASH-fibrosis' to understand their impact on study outcomes, depicted in Fig. 7b. The analysis revealed that some study outcomes negatively correlated with certain housing conditions, such as room humidity and grouped housing. Additionally, aspects of the dietary formula like energy density, percentage of fat and the use of coconut oil as a fat source also showed negative correlations. Conversely, the duration of the dietary challenge and specific nutritional formula features were positively correlated with MHPS. Noteworthy among these positively correlated variables were the relative nutrient composition (%kcal) in terms of proteins, cholesterol, carbohydrates, the use of palm oil as a fat source, and the combination of sucrose and fructose as carbohydrate source. Interestingly, the use of sucrose alone and sugar water (in the AMLD approaches) exhibited a negative correlation with the MHPS output for the 'ability to induce MASH-fibrosis'.

These results highlight critical features of the study designs that drive disease outcomes and provide insights into how additional model optimization could occur.

WDs allow the study of therapeutic interventions for BW loss and MASH amelioration

To gain further insights into the significance of the models that the MHPS ranked highly, we used the same phenotyping approach to assess their responsiveness to dietary and therapeutic interventions to promote weight loss. These interventions included chow reversal (12 weeks of chow reversal following 38 weeks on the GAN-C2 diet, or 8 weeks of chow reversal after 4 weeks on the CDHFD45% diet), calorie restriction (8 weeks of calorie restriction following 28 weeks of AMLN-C2 diet) and treatment with the GLP-1 receptor agonist semaglutide (12 weeks of semaglutide following 38 weeks of the GAN-C2 diet)[31–34]. Additional information regarding these study designs is provided in Supplementary Table 7 and illustrated in Fig. 8a.

Figure 8 and Supplementary Tables 7 and 8 demonstrate that all interventions (chow reversal, calorie restriction, semaglutide) substantially improved liver damage (evidenced by LW:BW% ratio, transaminases, steatosis) and reversed many effects of the MASLD-inducing diets on the DEGs and DRPs modulated in human MASH (as described in the sections above). However, it is essential to note that significant improvements in hepatocyte ballooning and inflammation were not observed in mice on AMLN-C2 and GAN-C2 diets. This suggests that longer treatment durations may be necessary to achieve substantial therapeutic effects. Interestingly, none of the treatments resulted in a significant improvement in fibrosis. Weight and dyslipidemia improved only in MASLD induced by AMLN-C2 and GAN-C2 diets. Notably, the GAN-C2 model effectively mimicked the lack of fibrosis improvement (despite an improved activity score) recently reported for semaglutide in human trials[35]. By contrast, chow reversal following the CDHFD diet paradoxically increased BW and cholesterol levels. This finding highlights the superiority of WD2% models like GAN-C2 for treatments targeting BW and/or metabolic pathways.

Discussion

Generating a murine model that fully recapitulates MASLD pathophysiology and mimics human disease is the holy grail of preclinical research for which there are too many candidates and no consensus[16]. Although various preclinical models have been suggested to replicate human MASLD, a thorough comparative study that aligns these animal models with human clinical outcomes and assesses their closeness to human disease pathophysiology through 'omics' approaches is still missing. Therefore, we performed a 3Rs (replacement, reduction and refinement)-compliant retrospective review of commonly used murine MASLD models available to the LITMUS Consortium and collaborators. These models were evaluated and ranked based on their alignment with three critical features of MASLD in humans: clinical metabolic phenotype, liver histopathology and liver transcriptome benchmarked against human transcriptomic changes.

The evaluation criteria for these models were established in advance and quantified using human proximity scores tailored to assess phenotype (PHPS), histopathology (HHPS) and transcriptome (DHPS). These distinct scores were subsequently integrated to create the MHPS. This innovative and insightful tool ranks murine models in terms of their resemblance to human cohorts with biopsy-confirmed NAFLD/MASLD[25,26].

Our study provides several important outcomes and unique insights.

A key strength of our study lies in the uniformity of methodology applied across all models for staining and scoring liver histopathology. Additionally, we have developed an original approach for interpreting murine histology: our proposed HHPS is grounded in the canonical parameters (such as steatosis, hepatocyte ballooning, lobular inflammation and fibrosis) used by established human NAFLD/MASLD scoring systems (CRN and SAF)[25,26]. However, the HHPS introduces a unique qualitative scoring approach, focusing on how the morphological features of liver tissue in mice mirror the pathology observed in humans, as detailed in Supplementary Table 6. This approach is designed to complement, but not replace, the canonical scores, which are more typically used to assess the impact of specific treatments on histology. Our study also details the development of hepatocyte ballooning in murine models, addressing previous debates and uncertainties surrounding this histological feature[16].

Our extensive comparison and ranking of preclinical models identified a short list of those that most closely align with human disease progression; however, none are able to fully replicate all aspects of human MASLD. This finding may reflect the inherent complexity and diversity of human disease and the fundamental physiological differences between humans and murine models; these differences encompass lipid metabolism, energy expenditure, circadian rhythms and eating patterns. Furthermore, how animals react to various diets, manifesting specific metabolic traits, exhibiting incomplete histological damage (for instance, hepatocyte ballooning) or responding to medications (including unique reactions or BW changes that do not mirror human responses), can influence the direct comparability between murine models and human disease[12–14,35–38].

We have characterized molecular transcriptomic changes in each murine model, introducing DHPS to evaluate their resemblance to human NAFLD/MASLD transcriptomic alterations. This molecular characterization is crucial for selecting the most appropriate preclinical models for specific mechanistic or pharmacological studies, such as targeting a particular gene or pathway. Moreover, our findings highlight that relying solely on transcriptomics does not accurately predict disease progression or fibrosis. Although transcript changes in HFD and most WD and AMLD models align with differentially modulated pathways and genes in MASH progression, like those markers indicating stellate cell activation or collagen deposition, they might not correspond with significant histologic fibrosis. This discrepancy is vital to consider in disease progression studies and therapeutic interventions whereby conclusions are based only on gene expression changes. For instance, the pharmacological treatment with semaglutide in the GAN-C2 model did not improve fibrosis histologically despite enhancing these pathways at the transcriptomic level, as seen in human trials[39].

Most importantly, our research has identified a particularly valuable subset of murine models accurately reflecting human MASLD in

terms of metabolic and fibro-inflammatory aspects. We found that dietary formulas rich in fat, refined carbohydrates (notably the combination of glucose and fructose) and cholesterol effectively replicate many facets of human MASLD, including fibrosis. A key observation is that diets with high cholesterol levels (up to 2%) accelerate disease progression and guarantee significant fibrosis.

Upon examining diet compositions more closely, WD and WDAMLD variations, with or without a low dose of CCl_4, emerge as the most well-rounded options for most preclinical studies. These diets induce metabolic disturbances and lead to progressive MASH, mirroring the progression in human MASLD. Significant fibrosis (F2 or higher) can be induced by combining a high cholesterol content (2%) with extended duration (over 40 weeks) and/or genetically altered (obese) backgrounds. In our evaluation of various study designs, the most effective performance was observed with the GAN-C2 diet fed to C57BL6/NTac or ob/ob mice. Regarding clinical translatability, the GAN-C2 model responded well to therapeutic interventions, showing improvements in metabolism, liver damage and gene expression. By contrast, choline-deficient diets (like CDD, CDHFD and CDHCD) achieved F2+ fibrosis in shorter time frames (12 weeks or less) but lacked a significant metabolic phenotype. Compared to the WD, these diets did not offer a substantial advantage in closely replicating human MASH-fibrosis regarding histopathology and gene expression, as illustrated in Fig. 6b and Extended Data Figure 4. However, they may benefit specific research objectives, such as testing anti-fibrotic treatments within reasonable time frames. The performance of the WD1.3%AMLD-CCl_4 (ref. 21) diet lies somewhere between WDs and choline-deficient dietary models. The translational potential of choline-deficient dietary models can be partly enhanced by using rat models and/or altering cholesterol metabolism. In our study, the choline-deficient dietary model with relatively better metabolic relevance was CDHFHCD1% given to LDLR KO mice for 8 weeks.

Our research also indicates the potential for further optimization of murine models. Although most model clusters effectively modulate key pathways integral to MASH progression, their alignment with human metabolic pathways is only partial. Building on the models we have identified as closely mirroring human metabolic characteristics and disease progression, a unique opportunity exists to further refine study designs to enhance the models' relevance. We observed that even minor variations in dietary formulas significantly impacted the regulation of metabolic pathways. PLSR and correlational analyses with the MHPS (Fig. 7) suggest that a combination of high cholesterol content, palm oil as a fat source and a carbohydrate mix of fructose plus glucose could be the basis for further dietary formula optimization. This approach could lead to more accurate models replicating the complex metabolic interactions seen in human MASLD.

Adjustments can be considered with regards to (1) modifying the energy density of MASH-inducing diets: incorporating sugar water (the AMLD approach) and the high-energy density of diets were found to correlate negatively with fibrosis endpoints. This suggests that rodents' mechanisms of adapting to nutrient intake[40] might delay disease progression. (2) Improving nutrient composition of MASLD-inducing diets: the study outcomes correlated with the relative nutrient composition suggest that a reduction in fat content (to or below 40%) and a more balanced inclusion of proteins and refined sugars could enhance both metabolic outcomes and fibrosis, potentially leading to a more physiological modulation of lipid and cholesterol pathways in agreement to human MASLD. (3) Considering housing parameters: grouped housing and room humidity also negatively correlated with MHPS. Thermoneutrality reduces energy expenditure, impairs systemic metabolism and might promote MASLD in chow-fed animals while accelerating the development of liver damage with a HFD[41–45]. By contrast, the phenotype of GAN-C2, which induces a more severe MASH phenotype than HFD at ambient temperature, did not differ significantly when evaluated at thermoneutrality (Figs. 2, 3 and 6), questioning the practicality

of its routine use due to organizational complexity[16,41–45]. (4) Duration of diet, but not the use of genetically modified animals, positively correlated with disease outcomes. Genetically modified models of obesity (for example, ob/ob, MC4R KO mice, ZSF1 rats) and dyslipidemia (for example, LDLR KO mice) accelerated liver histopathology damage and intensified some metabolic features. Despite the convenience of these genetic backgrounds for certain studies (for example, to study MASLD association with cardiovascular disease or obesity) and as quick models of MASH progression, caution is recommended, given that they do not adequately recapitulate human disease (given the low frequency of such mutations in human MASLD), they can feature an impairment in key pathophysiological pathways occurring in wild-type animals (for example, in Extended Data Figure 1, the transcriptomic profiles of ob/ob and db/db mice are different from those of wild-type animals) and their genetic perturbations may influence the response of the mouse to specific treatments (including confounding 'off-target' effects). However, genetically altered animal models can help to study the impact of genetic obesity or to explore the association of MASLD with atherosclerosis. (5) The addition of chemicals (for example, STZ or low-dose CCl_4) shortens the duration of the study and improves specific disease outcomes (like diabetes or fibrosis) but does not significantly enhance the overall relevance of the study designs to human conditions.

Other variables not evaluated in the present study that could provide an opportunity to improve model performance include investigating the impact of gender and/or microbiota on study outcomes. Additionally, exploring the effectiveness of these designs in other species, such as rats or larger mammals, could provide valuable insights. There is also scope for testing these models in genetic backgrounds more prone to obesity and liver damage. The use of humanized mice is another approach worth exploring; however, it is essential to acknowledge that high costs and technical and/or biological limitations may restrict their widespread use in the immediate future[11,46–48]. These future refinements could be crucial in enhancing the applicability and relevance of currently studied murine models in MASLD research.

Finally, our study significantly contributes to the field by proposing a standard for consensus in selecting murine models and offering an extensive dataset to facilitate future research. To address the challenges of standardization and consistency[16] and to prevent the proliferation of new models, we recommend that any claims regarding the superiority of proposed preclinical models over established 'gold standards' should be rigorously tested. This approach is essential to maintain clarity and avoid further confusion in the scientific literature.

This study, while comprehensive, does face certain limitations. (1) The retrospective design constrained our capabilities, for instance, in harmonizing study designs regarding fasting duration, sample sizes and statistical power. (2) Most of the studies were conducted on male rodents, preventing us from exploring gender differences in hepatic metabolism and MASLD, as well as the impact of diets in female animals; this highlights the need for work in female rodents as an important area for future research. (3) Available funds and resources limited our model selection: expanding our screening to include the vast array of study designs documented in the literature and female animals was technically and financially impractical[15]. (4) We could not examine the relevance of the murine models included to heterogeneous subtypes of human MASLD, such as pediatric or lean MASLD, or those involving genetic risk variants impacting phospholipid metabolism or lipoprotein secretion. The murine models tested also did not extend to advanced disease stages like cirrhosis, portal hypertension or hepatocellular carcinoma. (5) Of the various compounds currently undergoing human trials, only data and samples from mice treated with semaglutide were available for our study. (6) Insufficient information on glucose metabolism in most models precluded a systematic investigation of insulin resistance. (7) The impact of thermoneutrality has been explored only in the GAN-C2 model; therefore, its impact on disease outcomes cannot be generalized to all dietary models.

Despite these limitations, the thoroughness of our study and the systematic approach used in ranking the murine models render it a unique and valuable resource for the MASLD research community. This work lays the groundwork for future research to enhance preclinical murine models and achieve a consensus in the field.

In summary, our study reveals the extent to which the phenotype, liver histology and liver transcriptome of commonly used murine models of MASLD resemble human disease. Moreover, this study provides a comprehensive resource to explore how murine models recapitulate essential features of MASLD pathophysiology (for example, metabolic pathways, stellate cell activation, fibrosis development and so on), enabling optimal model selection based on the specific needs of a study. The MHPS rankings show that WD models emerge as the most balanced in terms of metabolic, histologic and transcriptomic similarities to human disease. However, only a few models exhibit MASH with significant fibrosis (F2 or higher), which typically necessitates extended durations, high cholesterol content and/or the use of genetically altered backgrounds.

In conclusion, our research provides a crucial foundation for the field, delineating the strengths and weaknesses of various prevalent MASLD models. This knowledge enables researchers to make informed decisions when choosing an experimental model that best aligns with human disease state and the specific requirements of their projects. Future studies should build on this work to refine MASLD murine models, enhancing their translational relevance and applicability to the diverse pathophysiology observed in human MASH.

Methods

Animal experiments
Training cohort. In this retrospective study, we compiled data and liver formalin-fixed paraffin-embedded tissues from 509 mice and 89 rats (39 MASLD models and CCl_4 as positive control for isolated fibrosis). These experimental designs represent a combination of diet, species, genetic background, wild-type or genetically altered animals, time points, sex and housing room temperature compared to their own control diets. A detailed description of each study protocol is described in Supplementary Table 1. Only models with centralized histology assessment (detailed below), transcriptomics data (detailed below) and phenotypic information were included in the analyses. Standard metabolic biochemistry (serum triglycerides, cholesterol, ALT, AST) and liver function tests were performed. The data included anthropometric measures (BW; LW; LW:BW%), breeder, strain, and housing and feeding design (for example, room temperature, type of facility and cages, health status of the animal facility).

For the 6J-WD-C1.3-FG-CCl₄-24W model, known as the FAT (fibrosis and tumors)-MASH rodent model (established in the Friedman laboratory[21]), 6-week-old male C57BL/6J mice were purchased from Jackson Laboratories. Five mice per cage were housed in a *Helicobacter*-free room in a 12 h light, 12 h dark cycle and weighed once weekly. CCl_4 was purchased from Sigma-Aldrich. CCl_4 was freshly dissolved in corn oil at a final concentration of 5% before injection. The final dose of pure CCl_4 was 0.2 µl g⁻¹ of BW, delivered intraperitoneally once per week, starting from initiation of the WDAMLD feeding, and continued for 24 weeks.

For the 6J-WD-C0.3-STZ-9W-WD-5W model, male C57BL/6J mice were purchased from Shanghai Lingchang Laboratory Animal Co. LTD. Two days after birth, these mice were administered a single dose (subcutaneous injection) of 200 µg STZ (Sigma, St. Louis, MO, USA). Upon reaching the age of 4 weeks, the mice were subjected to WD for 5 weeks. Details in Supplementary Table 1.

As a positive control for fibrosis, we used experiments in which peri-central fibrosis was induced by repeated CCl_4 injury. Mice were given CCl_4 (Sigma) dissolved in olive oil (Sigma) intraperitoneally at a dose of 0.75 ml kg⁻¹ of BW three times a week for 6 weeks. The age-matched control group received olive oil as a vehicle. At the end of the study, mice were humanely killed with carbon dioxide and liver

tissue was fixed in 10% buffered formalin for histopathology analysis and snap-frozen in liquid nitrogen for RNA analysis.

Treatments and intervention cohorts
The 'Treatments' cohort consisted of a total of 108 male mice challenged either with AMLN-C2 (36 weeks) or GAN-C2 (36 or 38 weeks) or CDAHFD-F45 (4 weeks) diets and followed by treatment, meaning dietary intervention (that is, calorie restriction (36 weeks) or chow reversal (8 or 12 weeks)) or pharmacological treatment (semaglutide (12 weeks)).

The data quality was comparable to the training cohort; a detailed description of each study protocol is provided in Supplementary Table 7.

Ethical approvals for all animal experiments
Relevant animal welfare authorities approved all the animal experiments that complied with national and international guidelines (details in Supplementary Tables 1 and 7).

Human next-generation sequencing datasets
The human cohorts were established according to the recruiting criteria of the old definition (NAFLD/NASH).

The UCAM–VCU super-cohort consisted of two publicly available (E-MTAB-9815, GSE130970) datasets (total of 136 patients) previously described by our teams[27,29]. All patients had a clinical diagnosis of NAFLD and histology scores according to the CRN scoring system[25]. Patients were divided into control ($n = 4$) and NAFLD ($n = 132$) subclustered against fibrosis (mild (F0), $n = 52$; moderate (F1–F2), $n = 50$; severe (F3–F4), $n = 30$).

The EPoS database orginates from a large cohort of NAFLD patients recruited in different European Union institutions, with publicly available next-generation sequencing data (GSE135251); this cohort was previously described by our team[28]. All patients had a clinical diagnosis of NAFLD/MASLD; histology was centrally scored according to the CRN scoring system as previously described[25]. A total of 38 patients from the initial cohort were removed, as they overlapped with the UCAM–VCU dataset. The remaining 168 patients with NAFLD were subclustered against fibrosis (mild (F0), $n = 47$; moderate (F1–F2), $n = 64$; severe (F3–F4), $n = 57$).

A description of the two cohorts is shown in Supplementary Table 3.

The relevant ethics committees (UCAM–VCU Cohort, East of England Research Ethics Committee and Virginia Commonwealth University; EPoS Cohort, multiple ethical committees in the participating countries) approved these studies as detailed in the original publications[27–29]. All patients gave informed consent to use biochemistry, clinical history and samples for research purposes. The principles of the Declaration of Helsinki were followed.

Murine histology
All murine liver samples were centrally stained and scanned by the Integrated Biobank of Luxembourg (IBBL). Each LITMUS partner shipped unstained formalin-fixed paraffin-embedded tissue slides. The slides were incubated for 30 min at 65 °C and then processed using a Varistain Gemini Slide Stainer. Following deparaffinization in xylene (×2, 5 min each) and rehydration in decreasing alcohols (100%, 95%, 70%; 2 min each) and tap water (1 min), one slide per case was stained as follows. H&E staining: immersion in hematoxylin (4 min), water wash (1 min), 1% acid alcohol (15 s), tap water wash (1 min), bluing reagent for counterstaining (1 min), tap water wash (1 min), 95% alcohol (1 min), alcoholic eosin (20 s), three alcohols (100%; 2 min, 2 min and 3 min, respectively), clearing in xylene substitute (two sets of 1 min and 2 min, respectively) and mounting. Hematoxylin 560 staining solution (Leica, 3801570 each four per case or 3801570 each); alcoholic eosin Y 515 (Leica, 3801616 each four per case or 3801615 each), bluing reagent (Thermo Scientific,

Epredia, 6769001), hydrochloric acid 37% (Sigma-Aldrich, 320331). Sirius Red staining: the slides were immersed in picrosirius red solution (VWR K640745, 500 ml) for 30 min, then rinsed in tap water, immersed in three alcohols (100%; 2 min, 2 min and 3 min, respectively), cleared in xylene substitute and mounted.

All stained slides were scanned by IBBL using the Nanozoomer 2.0 HT Slide Scanner; digital images were made available to the LITMUS Histopathology Group in a blinded manner using the CaloPix digital slide platform (TRIBVN Healthcare) for central histological scoring. LITMUS Histopathology Group members were all expert liver pathologists and assessed one H&E-stained and one Sirius Red-stained digital slide per case using the CRN and SAF scoring systems, as previously described[25,26]. Liver pathologists were harmonized to score human MASLD/MASH biopsies in LITMUS; their κ-score regarding interobserver agreement for hepatocyte ballooning has been shown to be substantial (κ = 0.8)[6]. Before scoring the animal model slides, a further round of harmonization took place. All liver pathologists had prior experience in histologically scoring animal MASLD model samples. A single pathologist scored each model.

The LITMUS Histopathology Group also developed an ad hoc scoring system (the HHPS; see Supplementary Table 6) that complements canonical scoring systems, providing a metric of how much the histological lesions in the murine models approximate human MASLD pathology.

Murine transcriptomics datasets

The murine models/datasets were derived from different LITMUS partners and facilities, and/or collaborators. As summarized in Supplementary Table 1, all models had RNA sequencing (RNA-seq) data, apart from two models profiled using Affymetrix microarray technology (6N-WD-C0.2-FG-9W, 6N-WD-C0.2-FG-29W). For RNA extraction and sample processing, different protocols and kits were used (STAT-60, RNeasy, Trizol), while the sequencing was performed mainly using Illumina's platforms (HiSeq3000/4000, NextSeq500/550, NOVASeq6000), using single- or paired-end reads (details in Supplementary Tables 1 and 7). The diversity of our dataset is explained by the retrospective design of the experiment. The experimental units have provided different models with samples obtained at different time points, leading to a merged non-standardized dataset regarding libraries, sequence methodologies and protocols.

Processing RNA-seq and microarray data

The models with microarray data were analyzed with R packages oligo (v.1.46)[49] and limma (v.3.38.2)[50], using RMA normalization and PCA analysis evaluating technical metadata. Technical outliers with only one or two samples processed for a given scan date were removed, retaining 13–15 samples for each time point and diet. Differential expression analysis was performed between the control and experimental diets, using linear model fit for each time point (Supplementary Fig. 2).

For all RNA-seq data, FastQC (v.0.11.9; https://github.com/s-andrews/FastQC) was used to test fastq files, and HISAT2 (v.2.1.0) mapped the reads to the human GRCh38, mouse GRCm38 or rat Rnor_6.0 genomes, using default parameters[51]. To proceed to subsequent analysis steps, all samples passed the following criteria: GC% content was approximately 50%, more than ten million reads passed the quality filtering and more than 80% of the reads per sample were mapped to the reference genome. HTSeq[52] (v.0.11.1) was used for gene counting and the R package biomaRt (v.2.54.0) mapped the ensemble gene IDs to HGNC symbols (Supplementary Fig. 2).

Given that the rodent datasets were produced by different facilities, in different batches and using different sequencing methodologies, we had to apply a batch-effect correction strategy to merge all datasets. First, quantile normalization was applied and then the Bioconductor function COMBAT[53] from the R package sva (v.3.38.0) removed the batch effect caused by data derived from different units.

For the human datasets, batch effects were addressed inside DESeq2 (ref. 54) using batch (dataset and gender) as a vector in the design formula (design = ~batch + condition). Differential gene expression analysis was performed with DESeq2 (ref. 54) (v.1.26.0) in the rodent models (treatment vs control) and the human comparison groups (mild vs control, moderate vs mild, severe vs mild).

Finally, we selected a standard set of genes expressed in all the models for input in the enrichment analysis of the different rodent models. Applying $\log_2$(transformed copies per million) ($\log_2$CPM) normalization (cpm from the R package edgeR v.3.32.1), the abundance of each transcript was normalized against the total number of reads in a sample, leading to high reproducibility of the average expression of housekeeping genes among the different batches. The selection of the standard set of genes expressed in the murine liver was based on the median CPMs derived from the livers of 50 healthy mice downloaded from Expression Atlas (https://www.ebi.ac.uk/gxa/download) and the position where the bimodal distribution for the high and low expressed genes cross (Supplementary Fig. 2; $x_{cross} = -2$). All the genes with $\log_2$CPM expression less than −2 in more than 10% of the models were excluded to ensure no substantial differences in the enrichment in the different models.

Statistical analysis and reproducibility

Each murine experiment was repeated once. Before calculating the statistical significance in R, $\log_2$ transformation was applied to the phenotypic data. The raw P values (each group vs its own control) were calculated for both phenotypic and histological data using the Mann–Whitney U-test (wilcox.test, R stats package v.3.6). The Benjamini–Hochberg method was applied to the results of the differential expression analysis to adjust the raw P values controlling the false discovery rate[55]. The batch-effect-corrected normalized counts corresponding to each animal were used to produce the principal components (PCs) in the mice and rats PCA plots, respectively (prcomp, R stats package v.4.0.3); the mean of the points corresponding to animals belonging to the same model (for each principal component) was then calculated to visualize each model as a separate point (the first two principal components are shown in Extended Data Figs. 1 and 2).

A hypergeometric test (base R package) was applied to evaluate the agreement between murine and human MASLD DEGs and DRPs. The test was performed on the results derived from the whole transcriptome as well as on those that were statistically significant and in the same direction of regulation as the human reference dataset hits (DEGs and DRPs; characterizing the 'early disease development', 'all disease stages' and 'disease progression groups'). As a background, we used the whole transcriptome or the union of all the DEGs or DRPs (derived from these three groups; Fig. 4). For the human MASLD clinical features, an ANOVA test (base R package) was performed on the different MASLD groups for all the continuous variables (age, steatosis, inflammation, ballooning, fibrosis and NAS score). A chi-squared test was implemented to characterize the categorical variables (N, T2DM). Tukey's test was then implemented to assess the post-hoc analysis significance by comparing the different MASLD groups (Supplementary Table 3).

A range of study design parameters from Supplementary Table 1 was evaluated in a PLSR model to examine their relationship with the MHPS ('ability to induce MASH-fibrosis' and 'metabolic relevance'). Study parameters included both categorical (introduced as binary variables as follows: strain/background as C57BL/6J (yes or no) and C57BL/N (yes or no); main source of fat as lard (yes or no), Primex shortening hydrogenated (yes or no), coconut oil (yes or no), milkfat (yes or no) and palm oil (yes or no); WT/GA as genetically altered (yes or no); housing details as grouped housing (yes or no); sugar water? (yes or no); type of cage (IVC or standard); STZ and CCl_4 together as chemicals (yes or no); details of the facility (SPF or standard); refined carbohydrates as sucrose in formula (yes or no) and sucrose and fructose in formula (yes or no)) and continuous variables (energy density (kcal g^{-1}),

carbohydrates (%kcal), fat (%kcal), proteins (%kcal), cholesterol (%), methionine (g kg^{-1}), choline (g kg^{-1}), duration experimental diet (weeks), room humidity and room temperature). The analysis focused on the larger subset of mouse models including HFD, HFDAMLD, WD, WD-STZ, WDAMLD or WDAMLD-CCl$_4$ to enhance interpretability. The R package mixOmics v.2.8.0 (http://mixomics.org) was used to build the PLSR model and evaluate the optimal number of components. Relevant parameters were selected by considering the VIP (score threshold of 1). A visually modified version of the clustered image map from mixOmics was used to evaluate the correlations between the selected study design parameters and the MHPS.

Being a retrospective analysis, no statistical methods were used a priori to pre-determine sample sizes that depended on data availability from previous experiments. Therefore, some comparisons with low sample sizes might be exposed to type II error. Analyses including log$_2$-transformed data met by default the normality assumptions. All other analyses were non-parametric and therefore normality testing was not a pre-requirement.

Enrichment analysis
The FGSEA (https://github.com/ctlab/fgsea) package was used for fast pre-ranked gene set enrichment analysis of human and preclinical model analysis results. The analysis was applied to pathways from the KEGG database (v.2019)[56].

MHPS
We developed a scoring system termed MHPS that favors murine models that induce clinical features of MASLD (obesity, dyslipidemia, increased ALT and AST levels) and mimics the histopathological and pathophysiological characteristics of human MASLD. It provides a dual ranking score based on the metabolic significance of the models or their ability to induce MASH-fibrosis. Each of the two 'arms' of the MHPS scoring system is composed of three different layers (PHPS, HHPS and DHPS). Each layer provides a score that is added to generate the two arms (metabolic-relevant or fibrotic-relevant) of the final MHPS ranking (Fig. 1), determining the resemblance of the preclinical models to the human MASLD.

The PHPS comprises a seven-point scoring system, ranking the models against the human phenotypic outcomes based on BW, triglyceride and cholesterol levels, LW:BW ratio, and AST and ALT levels. Preclinical models closely mimicking human systemic metabolic disease and MASH characteristics are prioritized, while those showing a lower resemblance are penalized. Details of the PHPS scoring system are shown in Supplementary Table 5. Additionally, ALT and AST levels separating the animals into either no MASLD or MASH F2+ were evaluated by a receiver operating characteristic (ROC) curve analysis using Youden's index to define the optimal threshold (Supplementary Fig. 3a,b). The final cutoff applied in the PHPS (Supplementary Table 5) calculation was defined as the combination of the optimal thresholds for ALT and AST (Supplementary Fig. 3c).

The HHPS assesses whether the histology samples mimic human MASLD/MASH (NAFLD/NASH) pathology and focuses on the key outputs, ranking the animals based on their 'metabolic relevance' or 'ability to induce MASH-fibrosis' as the main feature. HHPS includes some qualitative measures of human MASLD features that are expected in murine models to histologically mimic human MASH. Full details are shown in Supplementary Table 6.

The DHPS constitutes an adaptation of the DSEA and 'gene2drug' methods[57,58], ranking the preclinical models against the human RNA-seq outcomes. DSEA provides an enrichment score that ranks the proximity of a specific preclinical model to a standard human reference dataset. This reference was constructed by focusing on reproducible transcriptional and pathway changes in human MASH based on the UCAM–VCU and EPoS datasets. This conservative approach addresses the limitations of both studies (for example, UCAM–VCU contains a limited

control sample size and EPoS lacks lean controls) and provides confidence that the trend of regulation is similar between datasets, independently from possible power issues or biological differences in the cohorts. The DSEA method was applied in DEGs and differentially regulated KEGG pathways, after removing pathways not relevant for the liver and/or with redundant genes leading to their enrichment. To address the dual ranking system based on the metabolic or MASH-fibrosis relevance, human next-generation sequencing genes and pathways were divided into three groups: (1) all comparisons: hits homogeneously modulated at all disease stages (mild vs control, moderate vs mild, severe vs mild); (2) mild vs control: hits defining early disease stages (mild vs controls, but not moderate–severe vs mild) and (3) moderate–severe vs mild: hits defining progressive MASH (but not mild vs control). Hits from groups (1) and (2) or (1) and (3) were used as a reference for the metabolic-related or fibrotic-related ranking, respectively. This reference dataset in two layers has been used to rank the murine models based on their proximity to the expected outcome (how close their transcriptome changes are to human MASLD). The same layers of the analyses (genes or KEGG) performed in each murine model have been imputed in DSEA to complete the ranking. The outcome is an enrichment score and a P value that signifies how enriched the human disease signature is in the given model. A closer model to the expected phenotype has a higher enrichment score. To avoid bias in the interpretation of the results of DSEA, for all the non-statistically significant hits, the enrichment score is directly converted to zero, while enrichment scores from downregulated hits are multiplied by −1. The enrichment scores from the two different components of DSEA are then converted into a normalized enrichment score and averaged to generate the final DHPS. The basic principles of DSEA analysis that we have adapted to the needs of this study are described in Supplementary Fig. 4.

To merge the three different ranking methods described above (PHPS, DHPS, HHPS) into a final combined MHPS, the three components were scaled between 0 and 1 using the following formula:

$$Normalised\ scores = \frac{RowScore - \min(RowScore)}{\max(RowScore) - \min(RowScore)}$$

Each of these normalized scores contributed to one-third of the total score to generate the final MHPS score, which is directly translated into the final ranking.

Visualization of results
Data tables were generated using Excel (Microsoft Office 2019). All analyses were performed in R (v.4.0.3). The heatmaps were produced with the package Pheatmap (v.1.0.12; https://rdrr.io/cran/pheatmap). For the visualization of the Sankey diagram (signal flow from the pathways towards the genes; Fig. 5), we used the function sankeyNetwork from the package networkD3 (v.0.4; https://rdocumentation.org/packages/networkD3/versions/0.4/topics/sankeyNetwork).

The functions ggscatter (https://rdocumentation.org/packages/ggpubr/versions/0.5.0/topics/ggscatter) and plot_grid (https://rdocumentation.org/packages/cowplot/versions/1.1.1/topics/plot_grid) from the packages ggpubr (v.0.4.0) and cowplot (v.1.1.1) were used for the scatterplots in Supplementary Fig. 1, adding the regression lines and calculating the Pearson correlation scores and the corresponding P values. Figs. 2, 6 and 7a and Extended Data Figs. 3 and 4 were produced using packages cowplot and ggplot2 (https://rdocumentation.org/packages/ggplot2/versions/3.4.0).

The receiver operating characteristic curves in Supplementary Fig. 3 were produced using the roc.curve function from the PRROC package (https://rdocumentation.org/packages/PRROC/versions/1.3.1/topics/roc.curve); sensitivity and specificity were calculated using the caret package (https://rdocumentation.org/packages/caret/versions/6.0-93). For visualization, the ggplot2 and ggpubr packages were applied.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All the murine data in this manuscript are original and unpublished except for 6J-WD-C0.2-32W (GSE110404)[59], R-CDAA (GSE134715)[60] and GAN-C2 REV/SEMA (GSE196908)[34], which have been previously published and the raw data reused. Murine gene expression datasets have been deposited in the Array Express database (next-generation sequencing (NGS) accession number E-MTAB-12808; microarrays accession number E-MTAB-12817). All processed data used in or produced by this analysis have been deposited in Biostudies (accession number S-BSST1361; https://www.ebi.ac.uk/biostudies/studies/S-BSST1361), along with all murine metadata necessary for the interpretation, validation and expansion of the findings presented in this study. For those animals with available expression data, all metadata have also been deposited to the Array Express database. Human gene expression datasets and some metadata are publicly available (E-MTAB-9815, GSE130970, GSE135251); additional metadata are available upon request from the authors that originally published these datasets. Source data are provided with this paper.

Code availability

R code and scripts are available upon request.

References

1. Younossi, Z. et al. Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. *Nat. Rev. Gastroenterol. Hepatol.* **15**, 11–20 (2018).
2. Poirier, P. et al. Obesity and cardiovascular disease: pathophysiology, evaluation, and effect of weight loss: an update of the 1997 American Heart Association Scientific Statement on Obesity and Heart Disease from the Obesity Committee of the Council on Nutrition, Physical Activity, and Metabolism. *Circulation* **113**, 898–918 (2006).
3. Flegal, K. M., Kit, B. K., Orpana, H. & Graubard, B. I. Association of all-cause mortality with overweight and obesity using standard body mass index categories: a systematic review and meta-analysis. *JAMA* **309**, 71–82 (2013).
4. Rinella, M. E. et al. A multisociety Delphi consensus statement on new fatty liver disease nomenclature. *Hepatology* **78**, 1966–1986 (2023).
5. Tiniakos, D. G., Vos, M. B. & Brunt, E. M. Nonalcoholic fatty liver disease: pathology and pathogenesis. *Annu. Rev. Pathol.* **5**, 145–171 (2010).
6. Bedossa, P. & Consortium, F. P. Utility and appropriateness of the fatty liver inhibition of progression (FLIP) algorithm and steatosis, activity, and fibrosis (SAF) score in the evaluation of biopsies of nonalcoholic fatty liver disease. *Hepatology* **60**, 565–575 (2014).
7. 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).
8. Hardy, T., Oakley, F., Anstee, Q. M. & Day, C. P. Nonalcoholic fatty liver disease: pathogenesis and disease spectrum. *Annu. Rev. Pathol.* **11**, 451–496 (2016).
9. Anstee, Q. M., Reeves, H. L., Kotsiliti, E., Govaere, O. & Heikenwalder, M. From NASH to HCC: current concepts and future challenges. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 411–428 (2019).
10. Asgharpour, A. et al. A diet-induced animal model of non-alcoholic fatty liver disease and hepatocellular cancer. *J. Hepatol.* **65**, 579–588 (2016).
11. Bachmann, A. M. et al. Genetic background and sex control the outcome of high-fat diet feeding in mice. *iScience* **25**, 104468 (2022).
12. Farrell, G. et al. Mouse models of nonalcoholic steatohepatitis: toward optimization of their relevance to human nonalcoholic steatohepatitis. *Hepatology* **69**, 2241–2257 (2019).
13. Anstee, Q. M. Animal models in nonalcoholic steatohepatitis research: utility and clinical translation. *Liver Int.* **31**, 440–442 (2011).
14. Santhekadur, P. K., Kumar, D. P. & Sanyal, A. J. Preclinical models of non-alcoholic fatty liver disease. *J. Hepatol.* **68**, 230–237 (2018).
15. Im, Y. R. et al. A systematic review of animal models of NAFLD finds high-fat, high-fructose diets most closely resemble human NAFLD. *Hepatology* **74**, 1884–1901 (2021).
16. Gallage, S. et al. A researcher's guide to preclinical mouse NASH models. *Nat. Metab.* **4**, 1632–1649 (2022).
17. Boland, M. L. et al. Towards a standard diet-induced and biopsy-confirmed mouse model of non-alcoholic steatohepatitis: impact of dietary fat source. *World J. Gastroenterol.* **25**, 4904–4920 (2019).
18. Kampschulte, M. et al. Western diet in ApoE-LDLR double-deficient mouse model of atherosclerosis leads to hepatic steatosis, fibrosis, and tumorigenesis. *Lab. Invest.* **94**, 1273–1282 (2014).
19. Li, L. et al. A Western diet induced NAFLD in LDLR$^{-/-}$ mice is associated with reduced hepatic glutathione synthesis. *Free Radic. Biol. Med.* **96**, 13–21 (2016).
20. Saito, K. et al. Characterization of hepatic lipid profiles in a mouse model with nonalcoholic steatohepatitis and subsequent fibrosis. *Sci. Rep.* **5**, 12466 (2015).
21. Tsuchida, T. et al. A simple diet- and chemical-induced murine NASH model with rapid progression of steatohepatitis, fibrosis and liver cancer. *J. Hepatol.* **69**, 385–395 (2018).
22. Cubero, F. J. et al. TNFR1 determines progression of chronic liver injury in the IKKɣ/Nemo genetic model. *Cell Death Differ.* **20**, 1580–1592 (2013).
23. Anstee, Q. M. & Goldin, R. D. Mouse models in non-alcoholic fatty liver disease and steatohepatitis research. *Int. J. Exp. Pathol.* **87**, 1–16 (2006).
24. Han, M. A. T. et al. Rates of and factors associated with placebo response in trials of pharmacotherapies for nonalcoholic steatohepatitis: systematic review and meta-analysis. *Clin. Gastroenterol. Hepatol.* **17**, 616–629.e26 (2019).
25. Kleiner, D. E. et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology* **41**, 1313–1321 (2005).
26. Bedossa, P. et al. Histopathological algorithm and scoring system for evaluation of liver lesions in morbidly obese patients. *Hepatology* **56**, 1751–1759 (2012).
27. Azzu, V. et al. Suppression of insulin-induced gene 1 (INSIG1) function promotes hepatic lipid remodelling and restrains NASH progression. *Mol. Metab.* **48**, 101210 (2021).
28. Govaere, O. et al. Transcriptomic profiling across the nonalcoholic fatty liver disease spectrum reveals gene signatures for steatohepatitis and fibrosis. *Sci. Transl. Med.* **12**, eaba4448 (2020).
29. Hoang, S. A. et al. Gene expression predicts histological severity and reveals distinct molecular profiles of nonalcoholic fatty liver disease. *Sci. Rep.* **9**, 12541 (2019).
30. Vacca, M., Allison, M., Griffin, J. L. & Vidal-Puig, A. Fatty acid and glucose sensors in hepatic lipid metabolism: implications in NAFLD. *Semin. Liver Dis.* **35**, 250–261 (2015).
31. Houttu, V., Csader, S., Nieuwdorp, M., Holleboom, A. G. & Schwab, U. Dietary interventions in patients with non-alcoholic fatty liver disease: a systematic review and meta-analysis. *Front. Nutr.* **8**, 716783 (2021).

32. Promrat, K. et al. Randomized controlled trial testing the effects of weight loss on nonalcoholic steatohepatitis. *Hepatology* **51**, 121–129 (2010).

33. Wilding, J. P. H. et al. Once-weekly semaglutide in adults with overweight or obesity. *New Engl. J. Med.* **384**, 989–1002 (2021).

34. Mollerhoj, M. B. et al. Hepatoprotective effects of semaglutide, lanifibranor and dietary intervention in the GAN diet-induced obese and biopsy-confirmed mouse model of NASH. *Clin. Transl. Sci.* **15**, 1167–1186 (2022).

35. Seeley, R. J. & MacDougald, O. A. Mice as experimental models for human physiology: when several degrees in housing temperature matter. *Nat. Metab.* **3**, 443–445 (2021).

36. Azzu, V. & Valencak, T. G. Energy metabolism and ageing in the mouse: a mini-review. *Gerontology* **63**, 327–336 (2017).

37. Hunter, H. et al. Weight loss, insulin resistance, and study design confound results in a meta-analysis of animal models of fatty liver. *eLife* **9**, e56573 (2020).

38. Liang, W. et al. Establishment of a general NAFLD scoring system for rodent models and comparison to human liver pathology. *PLoS One* **9**, e115922 (2014).

39. Newsome, P. N. et al. A placebo-controlled trial of subcutaneous semaglutide in nonalcoholic steatohepatitis. *New Engl. J. Med.* **384**, 1113–1124 (2021).

40. Hu, S. et al. Dietary fat, but not protein or carbohydrate, regulates energy intake and causes adiposity in mice. *Cell Metab.* **28**, 415–431.e4 (2018).

41. Giles, D. A. et al. Thermoneutral housing exacerbates nonalcoholic fatty liver disease in mice and allows for sex-independent disease modeling. *Nat. Med.* **23**, 829–838 (2017).

42. Lee, E., Korf, H. & Vidal-Puig, A. An adipocentric perspective on the development and progression of non-alcoholic fatty liver disease. *J. Hepatol.* **78**, 1048–1062 (2023).

43. Castillo, M. et al. Disruption of thyroid hormone activation in type 2 deiodinase knockout mice causes obesity with glucose intolerance and liver steatosis only at thermoneutrality. *Diabetes* **60**, 1082–1089 (2011).

44. Nunes, J. R. C. et al. Thermoneutral housing does not accelerate metabolic dysfunction-associated fatty liver disease in male or female C57Bl/6J mice fed a Western diet. *Am. J. Physiol. Endocrinol. Metab.* **325**, E10–E20 (2023).

45. Oates, J. R. et al. Thermoneutral housing shapes hepatic inflammation and damage in mouse models of non-alcoholic fatty liver disease. *Front. Immunol.* **14**, 1095132 (2023).

46. Forcheron, F. et al. Nonalcoholic hepatic steatosis in Zucker diabetic rats: spontaneous evolution and effects of metformin and fenofibrate. *Obesity (Silver Spring)* **17**, 1381–1389 (2009).

47. Bilan, V. P. et al. Diabetic nephropathy and long-term treatment effects of rosiglitazone and enalapril in obese ZSF1 rats. *J. Endocrinol.* **210**, 293–308 (2011).

48. Minniti, M. E. et al. Insights from liver-humanized mice on cholesterol lipoprotein metabolism and LXR-agonist pharmacodynamics in humans. *Hepatology* **72**, 656–670 (2020).

49. Carvalho, B. S. & Irizarry, R. A. A framework for oligonucleotide microarray preprocessing. *Bioinformatics* **26**, 2363–2367 (2010).

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

51. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).

52. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

53. Zhang, Y., Jenkins, D. F., Manimaran, S. & Johnson, W. E. Alternative empirical Bayes models for adjusting for batch effects in genomic studies. *BMC Bioinf.* **19**, 262 (2018).

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

55. Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N. & Golani, I. Controlling the false discovery rate in behavior genetics research. *Behav. Brain Res.* **125**, 279–284 (2001).

56. Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M. & Tanabe, M. KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.* **49**, D545–D551 (2021).

57. Napolitano, F., Sirci, F., Carrella, D. & di Bernardo, D. Drug-set enrichment analysis: a novel tool to investigate drug mode of action. *Bioinformatics* **32**, 235–241 (2016).

58. Napolitano, F. et al. gene2drug: a computational tool for pathway-based rational drug repositioning. *Bioinformatics* **34**, 1498–1505 (2018).

59. Vacca, M. et al. Bone morphogenetic protein 8B promotes the progression of non-alcoholic steatohepatitis. *Nat. Metab.* **2**, 514–531 (2020).

60. Veyel, D. et al. Biomarker discovery for chronic liver diseases by multi-omics—a preclinical case study. *Sci. Rep.* **10**, 1314 (2020).

Acknowledgements

had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

Competing interests

Additional information

Michele Vacca [1,2,3,105] ✉, Ioannis Kamzolas [1,4,105], Lea Mørch Harder [5,105], Fiona Oakley [6], Christian Trautwein [7], Maximilian Hatting[7], Trenton Ross[8], Barbara Bernardo [8], Anouk Oldenburger[9], Sara Toftegaard Hjuler[5], Iwona Ksiazek [10], Daniel Lindén [11,12], Detlef Schuppan[13], Sergio Rodriguez-Cuenca[1], Maria Manuela Tonini[14], Tamara R. Castañeda[15], Aimo Kannt [16,17], Cecília M. P. Rodrigues [18], Simon Cockell[19], Olivier Govaere [20], Ann K. Daly[20], Michael Allison [21], Kristian Honnens de Lichtenberg[5], Yong Ook Kim [13], Anna Lindblom [11], Stephanie Oldham[22], Anne-Christine Andréasson[23], Franklin Schlerman [24], Jonathon Marioneaux[25], Arun Sanyal[26], Marta B. Afonso [18], Ramy Younes[20,27], Yuichiro Amano[28], Scott L. Friedman [29], Shuang Wang [29], Dipankar Bhattacharya[29], Eric Simon [30], Valérie Paradis[31], Alastair Burt[20,32], Ioanna Maria Grypari [33], Susan Davies[34], Ann Driessen[35,36], Hiroaki Yashiro[37], Susanne Pors[38], Maja Worm Andersen [38], Michael Feigh [38], Carla Yunis [39], Pierre Bedossa[20,40], Michelle Stewart[41], Heather L. Cater[41], Sara Wells [41], Jörn M. Schattenberg [42], Quentin M. Anstee[20,32], The LITMUS Investigators*, Dina Tiniakos [20,33,106] ✉, James W. Perfield [43,106] ✉, Evangelia Petsalaki [4,106] ✉, Peter Davidsen [5,104,106] ✉ & Antonio Vidal-Puig [1,44,106] ✉

[1]TVP Lab, WT/MRC Institute of Metabolic Science, University of Cambridge, Cambridge, UK. [2]Interdisciplinary Department of Medicine, University of Bari "Aldo Moro", Bari, Italy. [3]Laboratory of Liver Metabolism and MASLD, Roger Williams Institute of Hepatology, London, UK. [4]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridge, UK. [5]Research and Early Development, Novo Nordisk A/S, Måløv, Copenhagen, Denmark. [6]Newcastle Fibrosis Research Group, Biosciences Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK. [7]Department of Medicine III, University Hospital RWTH Aachen, Aachen, Germany. [8]Internal Medicine research Research Unit, Pfizer Worldwide Research and Development, Cambridge, MA, USA. [9]CardioMetabolic Diseases Research, Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach an der Riß, Germany. [10]Novartis Institutes for BioMedical Research, Novartis Pharma AG, Basel, Switzerland. [11]Bioscience Metabolism, Research and Early Development Cardiovascular, Renal and Metabolism (CVRM), AstraZeneca BioPharmaceuticals R&D, Gothenburg, Sweden. [12]Division of Endocrinology, Department of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden. [13]Institute of Translational Immunology and Research Center for Immunotherapy, Johannes Gutenberg University Medical Center, Mainz, Germany. [14]Luxembourg Institute of Health, Translational Medicine Operations Hub, Dudelange, Luxembourg. [15]R&D Diabetes & Portfolio Innovation and Excellence, Sanofi-Aventis Deutschland GmbH, Industriepark Hoechst, Frankfurt, Germany. [16]R&D Diabetes, Sanofi-Aventis Deutschland GmbH,

Industriepark Hoechst, Frankfurt, Germany. [17]Fraunhofer Institute for Translational Medicine and Pharmacology ITMP, Fraunhofer Innovation Center TheraNova and Goethe University, Frankfurt, Germany. [18]Research Institute for Medicines, Faculty of Pharmacy, Universidade de Lisboa, Lisbon, Portugal. [19]Bioinformatics Support Unit, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK. [20]Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK. [21]Liver Unit, Cambridge University Hospitals NHS Foundation Trust & Cambridge NIHR Biomedical Research Centre, Cambridge, UK. [22]Bioscience Metabolism, Research and Early Development Cardiovascular, Renal and Metabolism (CVRM), AstraZeneca BioPharmaceuticals R&D, Gaithersburg, MD, USA. [23]Bioscience Cardiovascular, Research and Early Development Cardiovascular, Renal and Metabolism (CVRM), AstraZeneca BioPharmaceuticals R&D, Gothenburg, Sweden. [24]Inflammation and Immunology Research Unit, Pfizer Worldwide Research and Development, Cambridge, MA, USA. [25]Fleur De Lis Holdings 10201 Dakins Dr. Richmond, Richmond, VA, USA. [26]Department of Internal Medicine, Virginia Commonwealth University, Richmond, VA, USA. [27]Boehringer Ingelheim International GmbH, Ingelheim am Rhein, Germany. [28]Research, Takeda Pharmaceutical Company Limited, Fujisawa, Japan. [29]Division of Liver Diseases, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [30]Global Computational Biology and Digital Sciences, Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach an der Riß, Germany. [31]Department of Imaging and Pathology, Université Paris Diderot and Hôpital Beaujon, Paris, France. [32]Newcastle NIHR Biomedical Research Centre, Newcastle upon Tyne Hospitals NHS Trust, Newcastle upon Tyne, UK. [33]Department of Pathology, Aretaeion Hospital, Medical School, National and Kapodistrian University of Athens, Athens, Greece. [34]Department of Cellular Pathology, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK. [35]Department of Pathology, Antwerp University Hospital, Edegem, Belgium. [36]Department of Molecular Imaging, Pathology, Radiotherapy, Oncology. Faculty of Medicine and Health Sciences, University of Antwerp, Wilrijk, Belgium. [37]Research, Takeda Pharmaceuticals Company Limited, Cambridge, MA, USA. [38]Gubra, Hoersholm, Denmark. [39]Pfizer, Inc.; Internal Medicine and Hospital, Pfizer Research and Development, Lake Mary, FL, USA. [40]LiverPat, Paris, France. [41]Mary Lyon Centre, MRC Harwell, Harwell Campus, Oxford, UK. [42]Department of Internal Medicine II, Saarland University Medical Centre, Homburg, Germany. [43]Lilly Research Laboratories, Eli Lilly and Company, Indianapolis, IN, USA. [44]Centro de Investigacion Principe Felipe, Valencia, Spain. [104]Present address: Ferring Pharmaceuticals A/S, International PharmaScience Center, Copenhagen, Denmark. [105]These authors contributed equally: Michele Vacca, Ioannis Kamzolas, Lea Mørch Harder. [106]These authors jointly supervised this work: Dina Tiniakos, James W Perfield, Evangelia Petsalaki, Peter Davidsen, Antonio Vidal-Puig. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: michele.vacca@uniba.it; Dina.Tiniakos@newcastle.ac.uk; Perfield_james_w@lilly.com; petsalaki@ebi.ac.uk; pkdavidsen@gmail.com; ajv22@medschl.cam.ac.uk

The LITMUS Investigators

Quentin M. Anstee[45], **Ann K. Daly**[45], **Simon Cockell**[45], **Dina Tiniakos**[45,107], **Pierre Bedossa**[45,107], **Alastair Burt**[45,107], **Fiona Oakley**[45], **Heather J. Cordell**[45], **Christopher P. Day**[45], **Kristy Wonders**[45], **Paolo Missier**[45], **Matthew McTeer**[45], **Luke Vale**[45], **Yemi Oluboyede**[45], **Matt Breckons**[45], **Jo Boyle**[45], **Patrick M. Bossuyt**[46], **Hadi Zafarmand**[46], **Yasaman Vali**[46], **Jenny Lee**[46], **Max Nieuwdorp**[46], **Adriaan G. Holleboom**[46], **Athanasios Angelakis**[46], **Joanne Verheij**[46,107], **Vlad Ratziu**[47], **Karine Clément**[47], **Rafael Patino-Navarrete**[47], **Raluca Pais**[47], **Valerie Paradis**[48,107], **Detlef Schuppan**[49], **Jörn M. Schattenberg**[49], **Rambabu Surabattula**[49], **Sudha Myneni**[49], **Yong Ook Kim**[49], **Beate K. Straub**[49,107], **Antonio Vidal-Puig**[50], **Michele Vacca**[50], **Sergio Rodrigues-Cuenca**[50], **Mike Allison**[50], **Ioannis Kamzolas**[50], **Evangelia Petsalaki**[50], **Mark Campbell**[50], **Chris J. Lelliott**[50], **Susan Davies**[50,107], **Matej Orešič**[51], **Tuulia Hyötyläinen**[51], **Aidan McGlinchey**[51], **Jose M. Mato**[52], **Óscar Millet**[52], **Jean-François Dufour**[53], **Annalisa Berzigotti**[53], **Mojgan Masoodi**[53], **Naomi F. Lange**[53], **Michael Pavlides**[54], **Stephen Harrison**[54], **Stefan Neubauer**[54], **Jeremy Cobbold**[54], **Ferenc Mozes**[54], **Salma Akhtar**[54], **Seliat Olodo-Atitebi**[54], **Rajarshi Banerjee**[55], **Elizabeth Shumbayawonda**[55], **Andrea Dennis**[55], **Anneli Andersson**[55], **Ioan Wigley**[55], **Manuel Romero-Gómez**[56], **Emilio Gómez-González**[56], **Javier Ampuero**[56], **Javier Castell**[56], **Rocío Gallego-Durán**[56], **Isabel Fernández-Lizaranzu**[56], **Rocío Montero-Vallejo**[56], **Morten Karsdal**[57], **Daniel Guldager Kring Rasmussen**[57], **Diana Julie Leeming**[57], **Antonia Sinisi**[57], **Kishwar Musa**[57], **Estelle Sandt**[58], **Maria Manuela Tonini**[58], **Elisabetta Bugianesi**[59], **Chiara Rosso**[59], **Angelo Armandi**[59], **Fabio Marra**[60], **Amalia Gastaldelli**[61], **Gianluca Svegliati**[62], **Jérôme Boursier**[63], **Sven Francque**[64], **Luisa Vonghia**[64], **An Verrijken**[64], **Eveline Dirinck**[64], **Ann Driessen**[64,107], **Mattias Ekstedt**[65], **Stergios Kechagias**[65], **Hannele Yki-Järvinen**[66], **Kimmo Porthan**[66], **Johanna Arola**[66,107], **Saskia van Mil**[67], **George Papatheodoridis**[68], **Helena Cortez-Pinto**[69], **Ana Paula Silva**[69], **Cecilia M. P. Rodrigues**[70], **Luca Valenti**[71], **Serena Pelusi**[71], **Salvatore Petta**[72], **Grazia Pennisi**[72], **Luca Miele**[73], **Antonio Liguori**[73], **Andreas Geier**[74], **Monika Rau**[74], **Christian Trautwein**[75], **Johanna Reißing**[75], **Guruprasad P. Aithal**[76], **Susan Francis**[76], **Naaventhan Palaniyappan**[76], **Christopher Bradley**[76], **Paul Hockings**[77], **Moritz Schneider**[77], **Philip N. Newsome**[78], **Stefan Hübscher**[78,107], **David Wenn**[79], **Jeremy Magnanensi**[80], **Aldo Trylesinski**[81], **Rebeca Mayo**[82], **Cristina Alonso**[82], **Kevin Duffin**[83], **James W. Perfield**[83], **Yu Chen**[83], **Mark L. Hartman**[83], **Carla Yunis**[84], **Melissa Miller**[84], **Yan Chen**[84], **Euan James McLeod**[84], **Trenton Ross**[84], **Barbara Bernardo**[84], **Corinna Schölch**[85], **Judith Ertle**[85], **Ramy Younes**[85], **Harvey Coxson**[85], **Eric Simon**[85], **Joseph Gogain**[86], **Rachel Ostroff**[86,108], **Leigh Alexander**[86], **Hannah Biegel**[86], **Mette Skalshøi Kjær**[87], **Lea Mørch Harder**[87], **Naba Al-Sari**[87], **Sanne Skovgård Veidal**[87], **Anouk Oldenburger**[87], **Jens Ellegaard**[88], **Maria-Magdalena Balp**[89], **Lori Jennings**[89], **Miljen Martic**[89], **Jürgen Löffler**[89], **Douglas Applegate**[89], **Richard Torstenson**[90], **Daniel Lindén**[90], **Céline Fournier-Poizat**[91], **Anne Llorca**[91], **Michael Kalutkiewicz**[92], **Kay Pepin**[92], **Richard Ehman**[92], **Gerald Horan**[93], **Gideon Ho**[94], **Dean Tai**[94], **Elaine Chng**[94], **Teng Xiao**[94], **Scott D. Patterson**[95], **Andrew Billin**[95], **Lynda Doward**[96], **James Twiss**[96], **Paresh Thakker**[97], **Zoltan Derdak**[97], **Hiroaki Yashiro**[97], **Henrik Landgren**[98], **Carolin Lackner**[99,107], **Annette Gouw**[100,107], **Prodromos Hytiroglou**[101,107], **Olivier Govaere**[102,108] & **Clifford Brass**[103,108]

[45]Newcastle University, Newcastle, UK. [46]AMC Amsterdam, Amsterdam, Netherlands. [47]Institute of Cardiometabolism And Nutrition, Paris, France. [48]Hôpital Beaujon, Assistance Publique Hopitaux de Paris, Paris, France. [49]University Medical Center Mainz, Mainz, Germany. [50]University of Cambridge, Cambridge, UK. [51]Örebro University, Örebro, Sweden. [52]Center for Cooperative Research in Biosciences, Derio, Spain. [53]University of Bern, Bern, Switzerland. [54]University of Oxford, Oxford, UK. [55]Perspectum, Oxford, UK. [56]Servicio Andaluz de Salud, Seville, Spain. [57]Nordic Bioscience, Herlev, Denmark. [58]Integrated Biobank of, Luxembourg, Luxembourg. [59]University of Torino, Torino, Italy. [60]Università degli Studi di Firenze, Florence, Italy. [61]Consiglio Nazionale delle Ricerche, Rome, Italy. [62]Università Politecnica delle Marche, Ancona, Italy. [63]University Hospital of Angers, Angers, France. [64]Antwerp University Hospital, Antwerp, Belgium. [65]Linköping University, Linköping, Sweden. [66]University of Helsinki, Helsinki, Finland. [67]UMC Utrecht, Utrecht, Netherlands. [68]Medical School of National & Kapodistrian University of Athens, Athens, Greece. [69]Faculdade de Medicina, Universidade de

Lisboa, Centro Hospitalar Vila Nova de Gaia, Lisbon, Portugal. [70]Faculty of Pharmacy, Universidade de Lisboa, Lisbon, Portugal. [71]Università degli Studi di Milano, Milan, Italy. [72]Università degli Studi di Palermo, Palermo, Italy. [73]Università Cattolica del Sacro Cuore, Rome, Italy. [74]University Hospital Würzburg, Würzburg, Germany. [75]RWTH Aachen University Hospital, Aachen, Germany. [76]University of Nottingham, Nottingham, UK. [77]Antaros Medical, Gothenburg, Sweden. [78]National Institute for Health Research, Biomedical Research Centre at University Hospitals Birmingham NHS Foundation Trust and the University of Birmingham, Birmingham, UK. [79]iXscient, London, UK. [80]Genfit, Lille, France. [81]Intercept Pharma, New York, NY, USA. [82]OWL, Coimbra, Portugal. [83]Eli Lilly and Company, Indianapolis, USA. [84]Pfizer, New York, NY, USA. [85]Boehringer-Ingelheim, Ingelheim am Rhein, Germany. [86]Somalogic, Boulder, CO, USA. [87]Novo Nordisk, Copenhagen, Denmark. [88]Ellegaard Göttingen Minipigs, Dalmose, Denmark. [89]Novartis Pharma AG, Basel, Switzerland. [90]AstraZeneca, Gothenburg, Sweden. [91]Echosens, Paris, France. [92]Resoundant, Rochester, USA. [93]Bristol-Myers Squibb, New York, NY, USA. [94]HistoIndex, Singapore, Singapore. [95]Gilead, Foster City, CA, USA. [96]RTI-HS, Research Triangle Park, NC, USA. [97]Takeda Pharmaceuticals Company Ltd, Osaka, Japan. [98]AbbVie, North Chicago, USA. [99]Medical University of Graz, Graz, Austria. [100]University of Groningen, Groningen, Netherlands. [101]Aristotle University of Thessaloniki, Thessaloniki, Greece. [102]KU Leuven, Leuven, Belgium. [103]Resolution Therapeutics, Paris, France. [107]Member of the LITMUS Histopathology Group (LHG). [108]Refers to people who significantly contributed to the project but who have left their partner institution and now have a different affiliation.
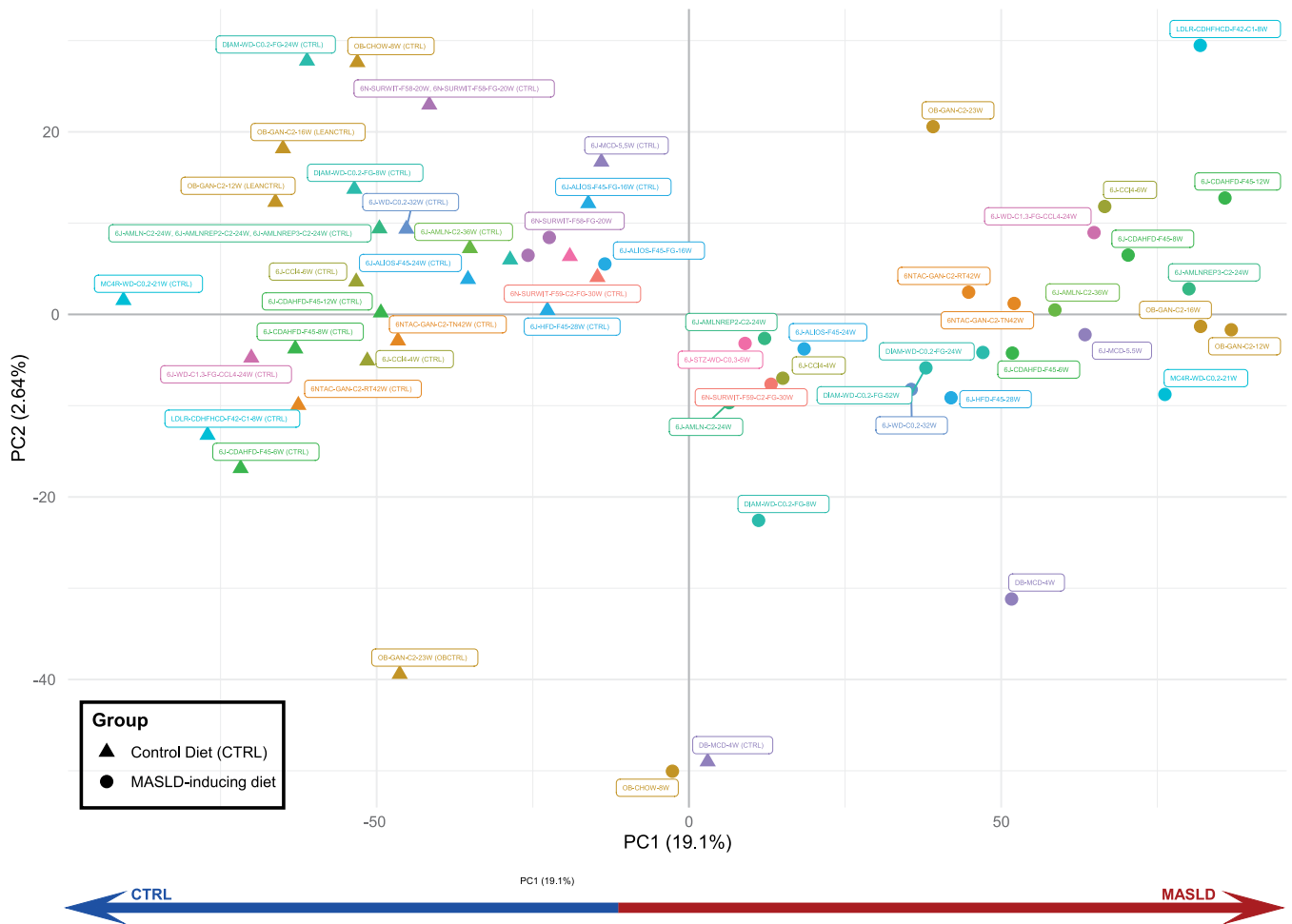
Extended Data Fig. 1 | PCA plot of RNASeq data comparing the different mouse MASLD models. The plot uses the normalised batch-corrected gene expression data. Each point represents a mouse model. The Controls (triangles) are clearly separated from the MASLD models (circles) according to the first principal component (PC1).
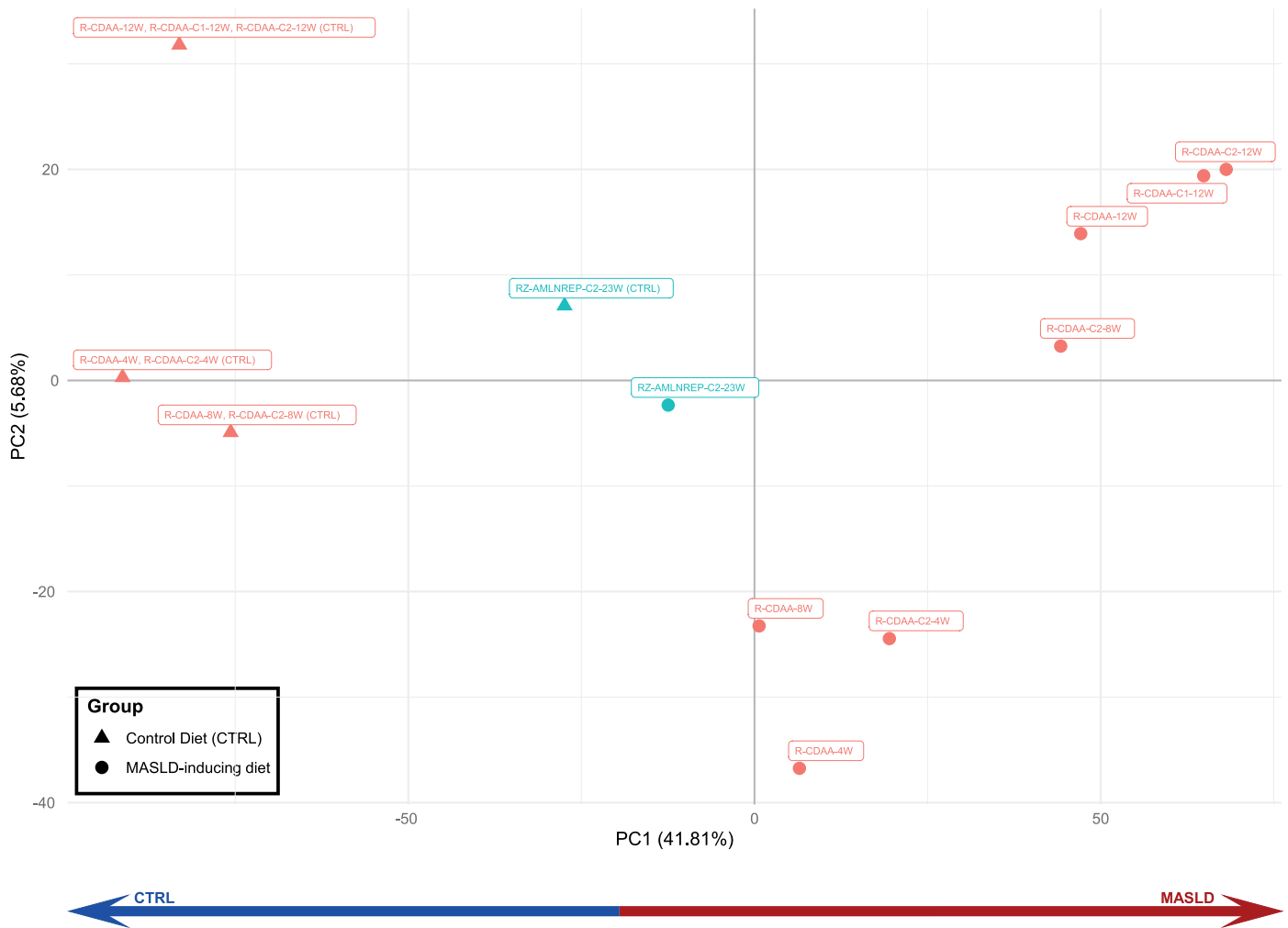
Extended Data Fig. 2 | PCA plot of RNASeq data comparing the different rat MASLD models. The plot uses the normalised batch-corrected gene expression data. Each point represents a rat model. The Controls (triangles) are separated from the MASLD models (circles) according to the first principal component (PC1).
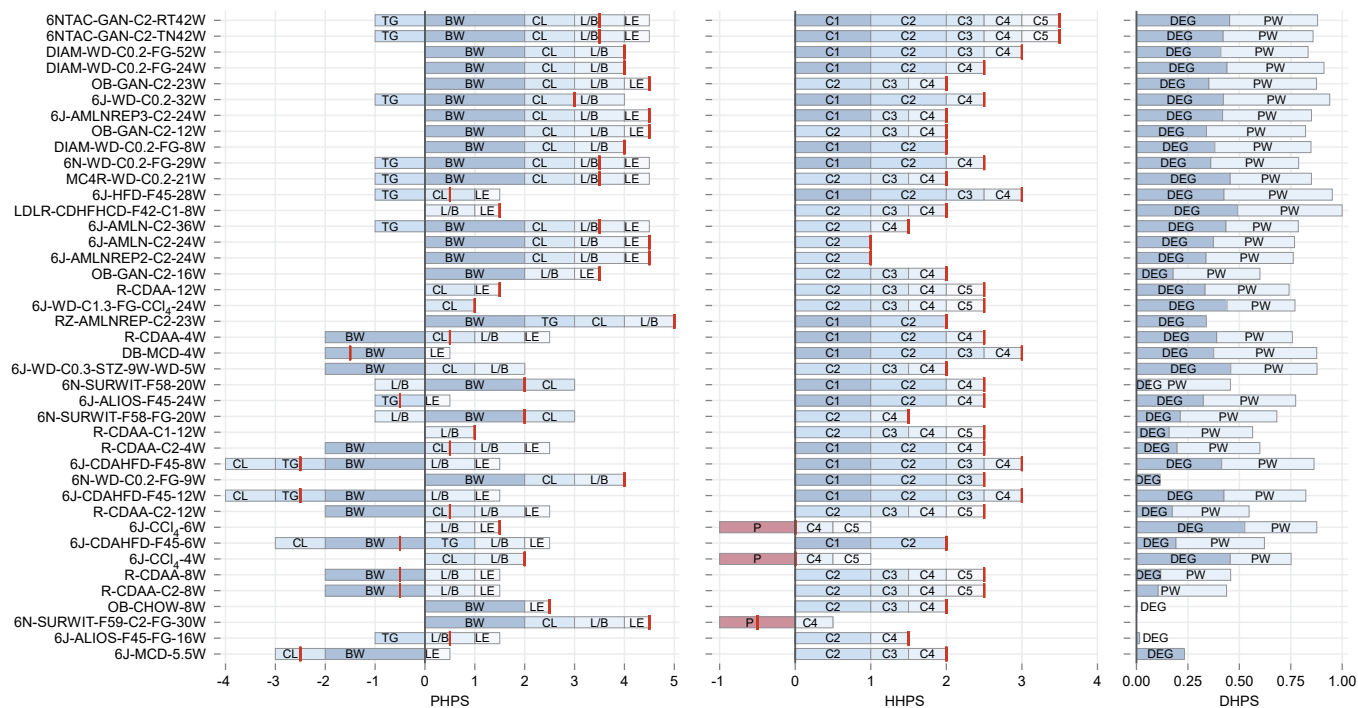
Extended Data Fig. 3 | Details of the MHPS components (PHPS, HHPS, DHPS) – Metabolic Relevance. Bar plots show the detailed composition of the MHPS components (PHPS, HHPS, and DHPS) utilised for the output focusing on metabolic relevance. Models are ordered globally according to the MHPS, where each barplot gives an overview of the total PHPS, HHPS or DHPS, the associated features, and their corresponding scores (see Supplementary Tables 5 and 6 and Supplementary Fig. 4 for an overview of the scoring strategies). For each model, the total scores of PHPS and HHPS are represented by red vertical lines to overcome the fact that some features are scored negatively. For PHPS and HHPS,

the length of the stacked bars represents the scores' absolute values (before normalising to the interval 0-1). DHPS directly shows the normalised score. Feature scores included are abbreviated as follows: **PHPS:** BW = Body weight, TG = Triglycerides, CL = Cholesterol, L/B = Liver Weight/Body weight (%), LE = Liver enzyme levels of ALT & AST; **HHPS**: C1 = Topography of histological lesions, C2 = Type of steatosis, C3 = Hepatocyte ballooning, C4 = Lobular inflammation, C5 = Mallory-Denk bodies; **DHPS**: DEG = Differentially Expressed Genes, DRP = Differentially Regulated Pathways.
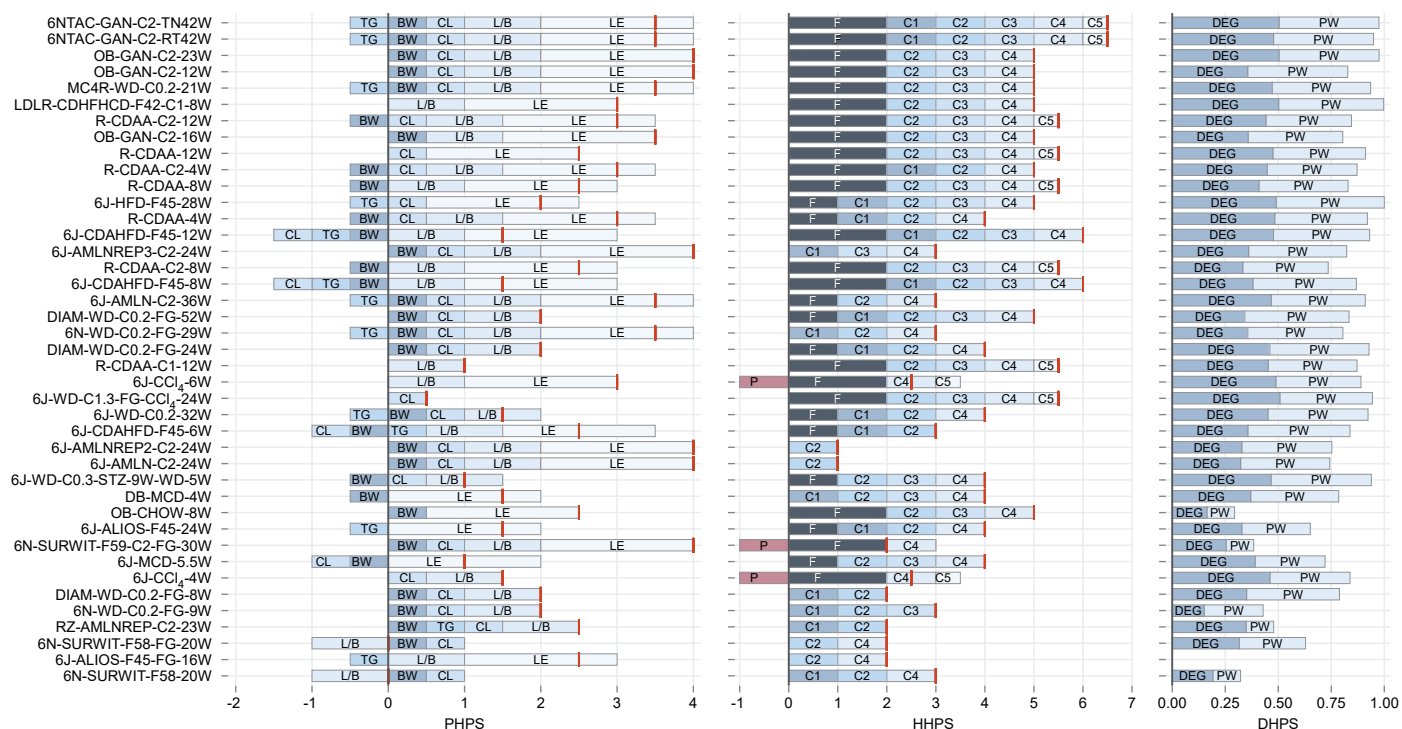
**Extended Data Fig. 4 | Details of the MHPS components (PHPS, HHPS, DHPS) – Ability to induce MASH-Fibrosis.** Bar plots show the detailed composition of the MHPS components (PHPS, HHPS, and DHPS) for the output focusing on the ability to induce MASH-Fibrosis. Models are ordered globally according to the MHPS, where each barplot gives an overview of the total PHPS, HHPS or DHPS, associated features, and corresponding scores (see Supplementary Tables 5 and 6 and Supplementary Fig. 4 for an overview of the scoring strategies). For each model, the total scores of PHPS and HHPS are represented by red vertical lines to overcome the fact that some features are scored negatively. For PHPS and HHPS,

the length of the stacked bars represents absolute scores' absolute values (before normalising to the interval 0-1). DHPS shows the normalised score directly. Feature scores included are abbreviated as follows: **PHPS**: BW = Body weight, TG = Triglycerides, CL = Cholesterol, L/B = Liver Weight/Body weight (%), LE = Liver enzyme levels of ALT & AST; **HHPS**: C1 = Topography of histological lesions, C2 = Type of steatosis, C3 = Hepatocyte ballooning, C4 = Lobular inflammation, C5 = Mallory-Denk bodies, F(C6) = Fibrosis; and for DHPS: DEG = Differentially Expressed Genes, DRP = Differentially Regulated Pathways.
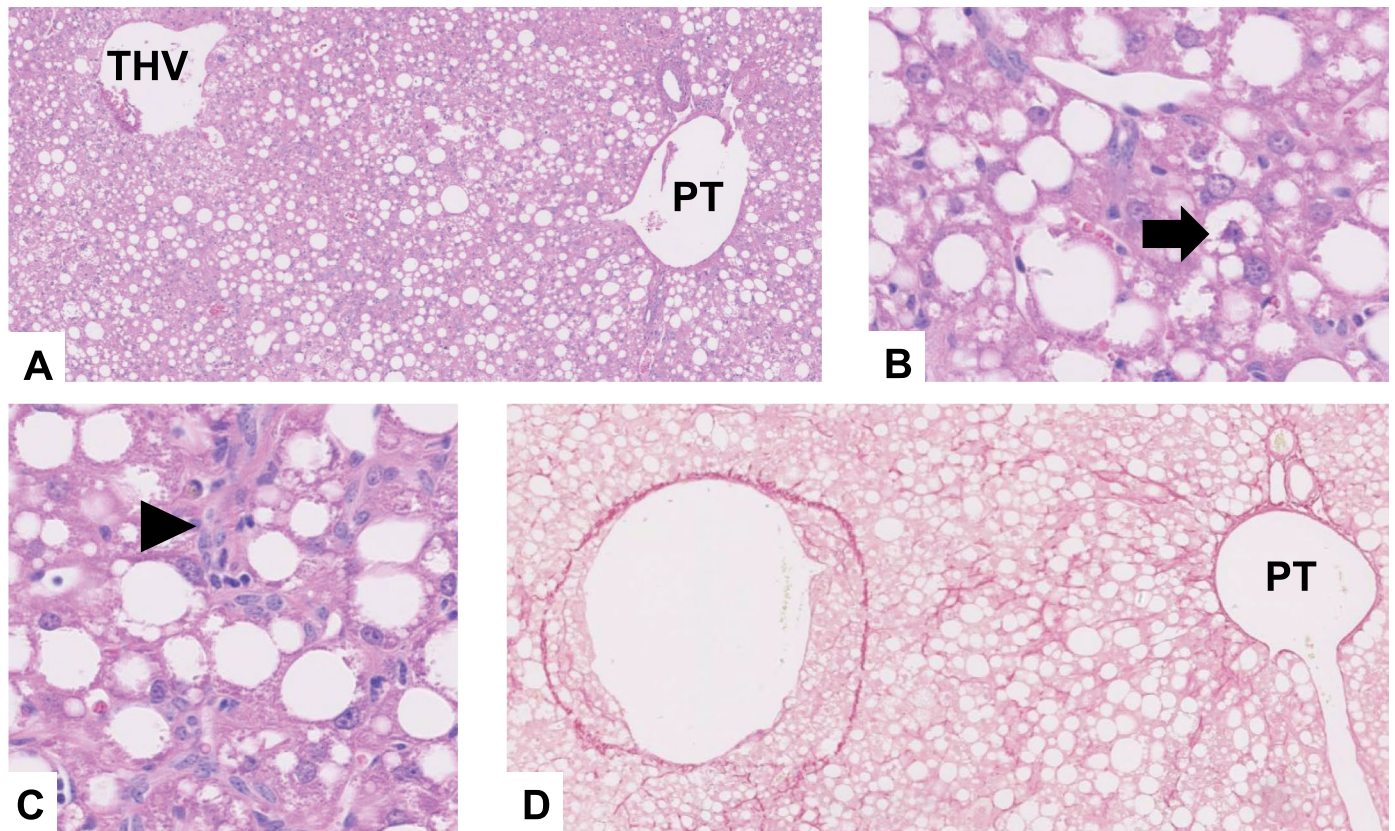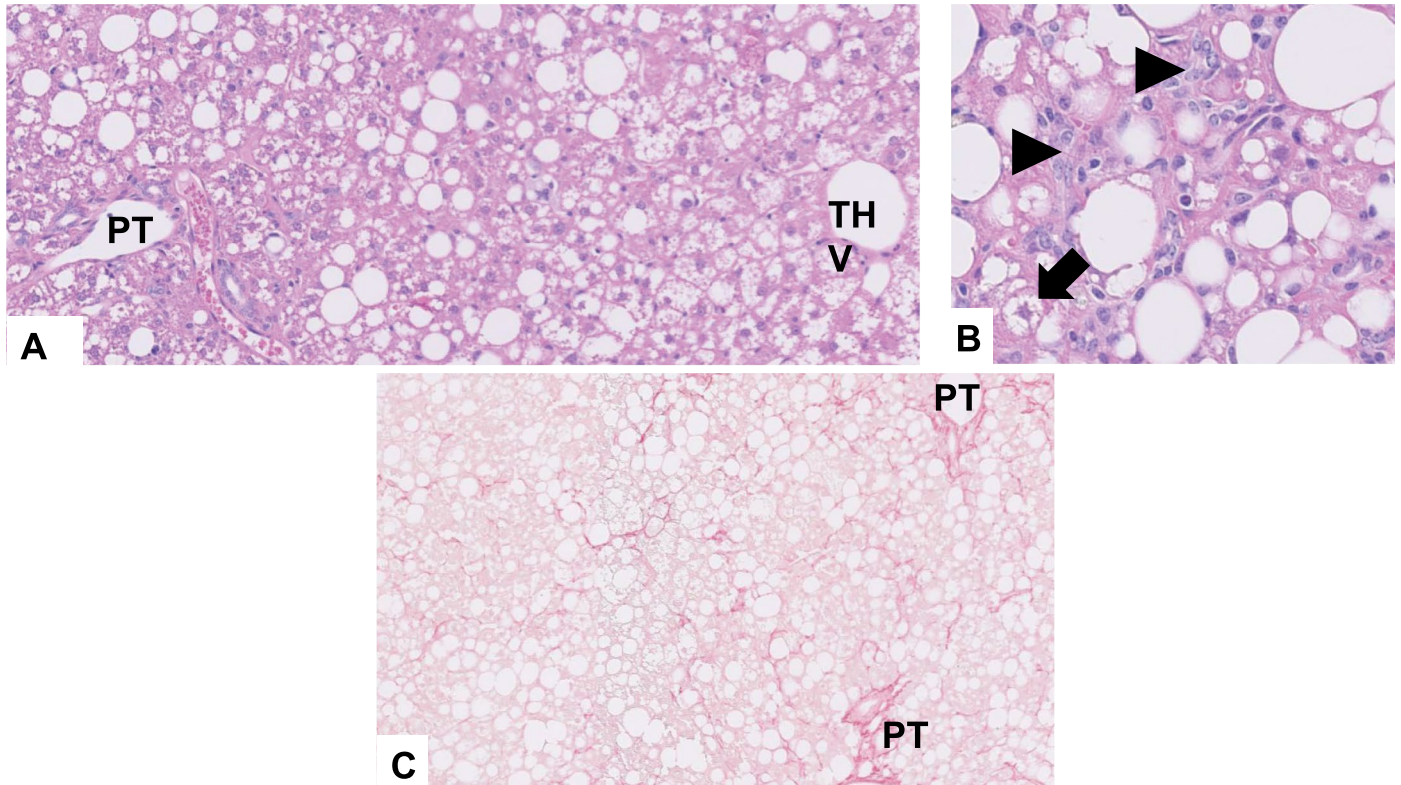
Extended Data Fig. 5 | Representative images of MASLD histopathology features in the 6NTAC-GAN-C2-TN42W model. a. Panlobular distribution of steatosis, grade 3, H&E stain, x50; **b**. Ballooned hepatocyte (arrow), H&E stain, x200; **c**. Necroinflammatory focus (arrowhead), H&E stain, x200; **d**. Periportal and extensive sinusoidal fibrosis, stage 2, Sirius red stain, x50 (THV terminal hepatic venule, PT portal tract).
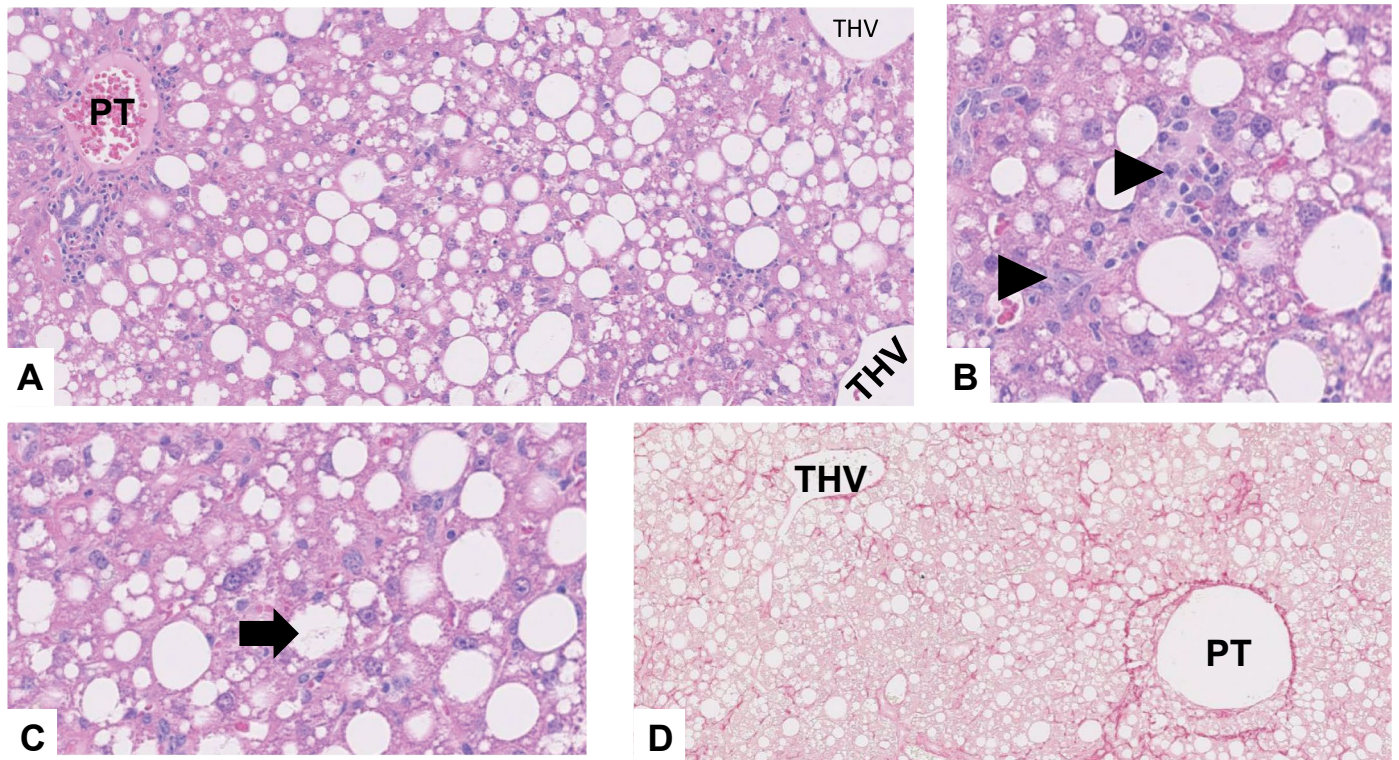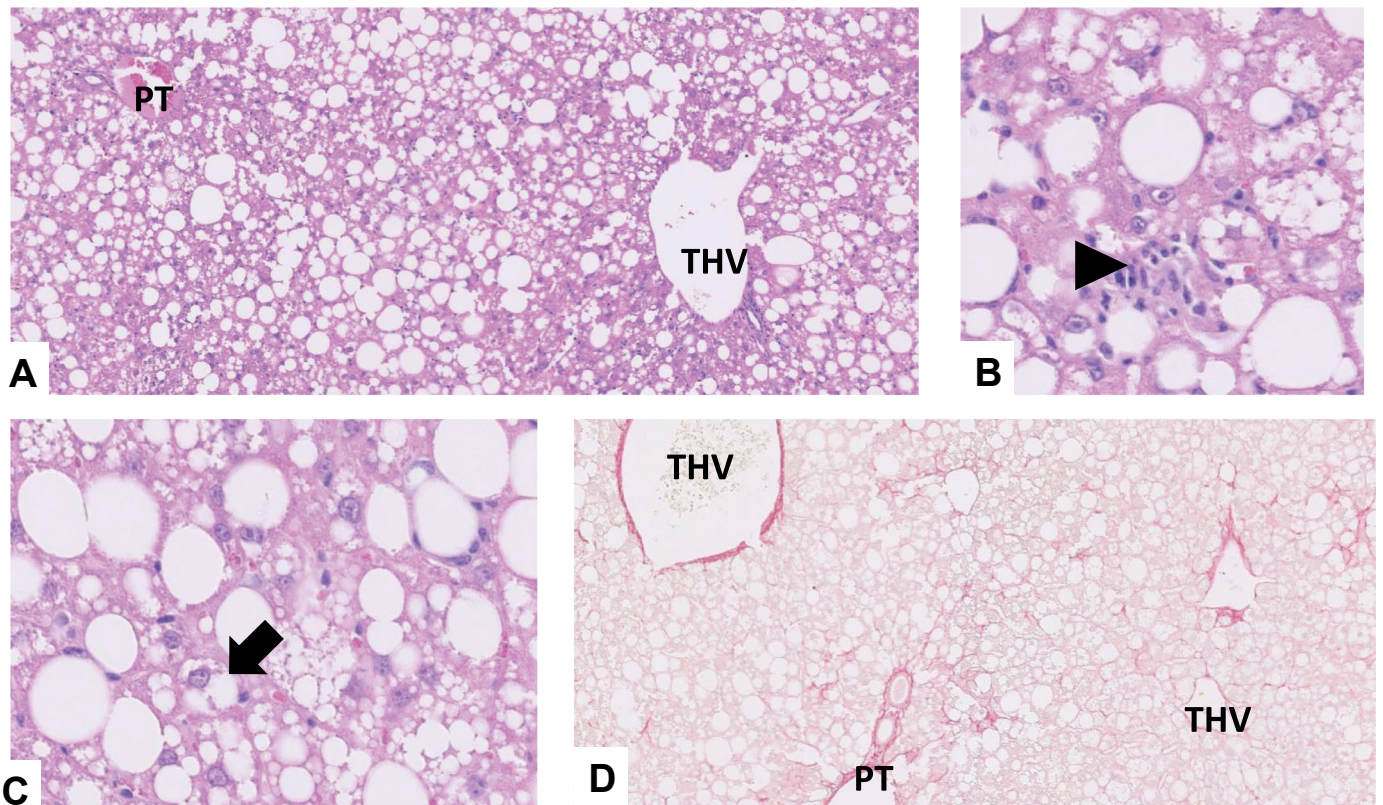
Extended Data Fig. 6 | Representative images of MASLD histopathology features in the 6NTAC-GAN-C2-RT42W model. a. Panlobular distribution of steatosis, grade 3, H&E stain, x50; **b**. Necroinflammatory foci (arrowheads), H&E stain, x200; **c**. Ballooned Hepatocyte (arrow), H&E stain, x200; **d**. Periportal and perisinusoidal fibrosis, stage 2, Sirius red stain, x50 (THV terminal hepatic venule, PT portal tract).
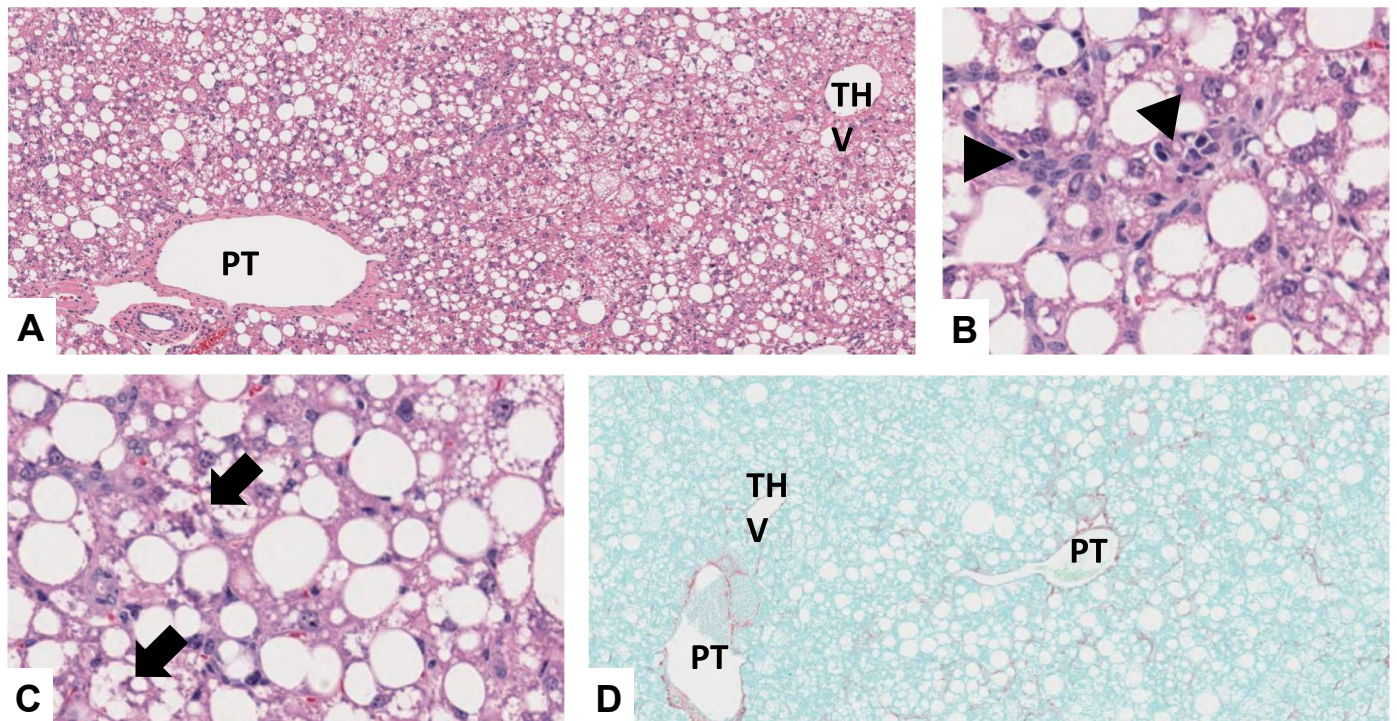
Extended Data Fig. 7 | Representative images of MASLD histopathology features in the AZ-OB-GAN-23W model. a. Panlobular distribution of steatosis, grade 3, H&E stain, x50; **b.** Non-classical ballooned hepatocyte (arrow) and necroinflammatory foci (arrowheads), H&E stain, x200; **c.** Periportal and perisinusoidal fibrosis, stage 2, Sirius red stain, x50 (THV terminal hepatic venule, PT portal tract).
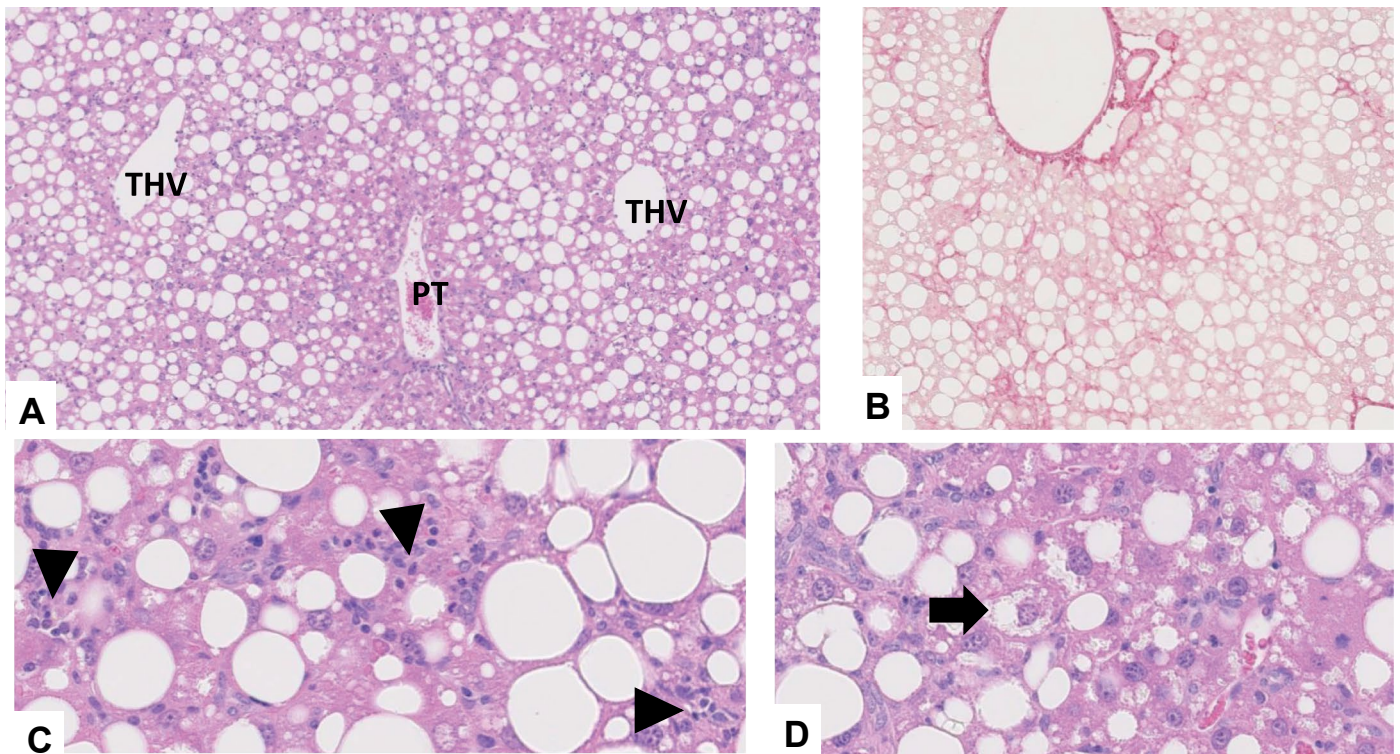
Extended Data Fig. 8 | Representative images of MASLD histopathology features in the OB-GAN-C2-12W model. a. Panlobular distribution of steatosis, grade 3, H&E stain, x50; **b**. Necroinflammatory focus (arrowhead), H&E stain, x200; **c**. Ballooned hepatocyte (arrow), H&E stain, x200; **d**. Periportal and perisinusoidal fibrosis, stage 2, Sirius red stain, x50 (THV terminal hepatic venule, PT portal tract).

Extended Data Fig. 9 | Representative images of MASLD histopathology features in the MC4R-WD-C0.2-21W model. a. Panlobular distribution of steatosis, grade 3, H&E stain, x50; **b.** Necroinflammatory foci (arrowheads), H&E stain, x200; **c.** Ballooned hepatocytes (arrows), H&E stain, x200; **d.** Periportal and perisinusoidal fibrosis, stage 2, Sirius red fast green stain, x50 (THV terminal hepatic venule, PT portal tract).

Extended Data Fig. 10 | Representative images of MASLD histopathology features in the LDLR-CDHFHCD-F42-C1-8W model. a. Panlobular distribution of steatosis, grade 3, H&E stain, x50; **b**. Periportal and perisinusoidal fibrosis, stage 2, Sirius red stain, x50; **c**) Necroinflammatory foci (arrowheads), H&E stain, x200; **d**. Ballooned hepatocyte (arrow), H&E stain, x200 THV terminal hepatic venule, PT portal tract).

nature portfolio

Corresponding author(s): Michele Vacca, Dina Tiniakos, James W Perfield, Evangelia Petsalaki, Peter Davidsen, Antonio Vidal-Puig

Last updated by author(s): Mar 22, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Microsoft Excel – Microsoft Office 2019
Data analysis	RStudio Inc, The R Foundation for Statistical Computing – v4.0.3; For NGS data analysis we used: FastQC - v0.11.9; https://github.com/s-andrews/FastQC, Hisat2 - v2.1.0, HTSeq - v0.11.1, biomaRt – v2.54.0, qnorm() - standard R function, Limma - v3.38.2, COMBAT - sva package -v3.38.0, DESeq2 – v1.26.0, Benjamini-Hochberg method – FDR calculation, cpm – edgeR -v3.32.1, FGSEA - https://github.com/ctlab/fgsea, KEGG database – v2019. For the phenotype and histology raw p-value calculation we used the Mann-Whitney U test (wilcox.test – stats v3.6.2). For the hypergeometric test we used the function phyper (base R package; REF: Johnson, N. L., Kotz, S., and Kemp, A. W. (1992) Univariate Discrete Distributions, Second Edition. New York: Wiley). For the human clinical feature characterisation we used ANOVA (base R package; for the continuous variables) and chi-squared test (for the chategorical variables). The function prcomp was used for the estimations of the principal components in the pca plots (package stats - v4.0.3). For the visualisation of the results we used the following tools: The heatmaps were produced with the package "Pheatmap" (v.1.0.12; https://rdrr.io/cran/pheatmap). The Sankey diagram was produced with the function "sankeyNetwork" from the package "networkD3" (v.0.4; https://rdocumentation.org/packages/networkD3/versions/0.4/topics/sankeyNetwork). The functions "ggscatter" (https://rdocumentation.org/packages/ggpubr/versions/0.5.0/topics/ggscatter) and "plot_grid" (https://rdocumentation.org/packages/cowplot/versions/1.1.1/topics/plot_grid) from the packages "ggpubr" (-v0.4.0) and "cowplot" (-v1.1.1), were used for the scatter plots, adding the regression lines, and calculating the R (Pearson correlation) scores and the corresponding p-values. The ROC-curves were produced using the "roc.curve" function from the "PRROC" package (https://rdocumentation.org/packages/PRROC/versions/1.3.1/topics/roc.curve), while sensitivity and specificity were calculated using the "caret" package (https://rdocumentation.org/packages/caret/versions/6.0-93). For visualisation, the ggplot2 (https://rdocumentation.org/packages/ggplot2/versions/3.4.0) and ggpubr packages were applied. The package ggplot was used for the visualisation of the pca plots. The partial least squares regression was performed using the "pls" function from the package

"mixOmics" (http://mixomics.org/, v2.8.0), while VIP scores and correlation structure between the study design and MHPS were extracted using the functions "vip" and "cim" functions respectively. The correlation structure result from cim was visualised using "Pheatmap" as described above.
Code is available upon request.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All the murine data in this manuscript are original and unpublished, except for the 6J-WD-C0.2-32W (GSE110404)49, R-CDAA (GSE134715)50, and GAN-C2 REV/SEMA (GSE196908)34 that have been previously published, and raw data reused. Murine Gene Expression datasets have been deposited in the Array Express database (NGS accession number: E-MTAB-12808; Microarrays accession number: E-MTAB-12817). All processed data used in or produced by this analysis have been deposited in Biostudies (Accession number: S-BSST1361, https://doi.org/10.6019/S-BSST1361), along with all murine metadata necessary for the interpretation, validation, and expansion of the findings presented in this study. For those animals with available expression data, all metadata have also been deposited to the Array Express database. Human Gene Expression datasets and some metadata are publicly available (E-MTAB-9815, GSE130970, GSE135251); additional metadata are available upon request from the authors that originally published these datasets.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	The human data (publicly available datasets) have been generated in males and females (sex). Gender information was not available to the investigators.
Population characteristics	The UCAM (University of Cambridge) / VCU (Virginia Commonwealth University) super-cohort consisted of 2 publicly available (E-MTAB-9815, GSE130970) datasets (TOTAL: 136 Patients) previously described by our teams. All the patients had a clinical diagnosis of NAFLD and histology scores according to the NASH CRN Scoring System. Patients were divided into CTRL (n = 4) and NAFLD (n = 132) sub-clustered against fibrosis (Mild i.e. F0: n = 52; Moderate i.e. F1-2: n = 50; Severe i.e. F3-4: n = 30). The EPoS (Newcastle University) Cohort is a large cohort of NAFLD patients with NGS available (GSE135251) recruited in different EU institutions and previously described by our teams. All the patients had a clinical diagnosis of NAFLD, and histology was centrally scored according to the NASH CRN Scoring System as previously described; 38 patients of the initial cohort were removed as they overlapped with the UCAM/VCU dataset. The remaining 168 NAFLD patients were sub-clustered against fibrosis (Mild i.e. F0: n = 47; Moderate i.e. F1-2: n = 64; Severe F3-4: n = 57). The covariate-relevant population characteristics included the gender for datasets E-MTAB-9815 and GSE130970 (publicly available), and the gender and site for dataset GSE135251 (requested and confidentially provided by the authors of the original study by email).
Recruitment	The Human data have been produced starting from already described studies with NGS data already available. The description of the recruitment was provided in the original publications (PMID: 33722690, 31467298, 33268509).
Ethics oversight	The relevant Ethics Committees (UCAM/VCU Cohort: East of England Research Ethics Committee, Virginia Commonwealth University; EPoS Cohort: multiple Ethical Committees in the participating countries) approved these studies as detailed in the original publications. The description of the relevant ethical approvals for this study has been provided in the original publications (PMID: 33722690, 31467298, 33268509).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Given the retrospective study design, statistical analyses were performed on the basis of data/sample availability. Power analysis has therefore not been performed.
Data exclusions	Animal models without centralised assessed histology, transcriptomics data, and phenotypic information were not included in the analyses. Transcriptomics data: For microarrays, technical outliers with only 1 or 2 samples processed for a given scan date were removed, retaining 13-15 samples for each timepoint/diet. RNA-seq inclusion criteria were the following: GC% content was approximately 50%, more than 10 million reads passed the quality filtering, and more than 80% of the reads per sample were mapped to the reference genome (all samples passed the QC).
Replication	Human data used to compare murine data were based on highly reproducible hits that resulted significant and with the same direction of modulation in the two datasets studied therefore ensuring high replicability among datasets. Each murine experiment was repeated once.
Randomization	Given the retrospective study design, randomisation is not applicable to this study.
Blinding	Given the retrospective study design, the blinding of the investigators during data acquisition is not applicable to this study. For the histology, both the histology team at IBBL and pathologists were blind when acquiring/assessing control/treatment diets. They were also unaware about the results of the other data layers not to influence their judgment.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	Antibodies
☒	Eukaryotic cell lines
☒	Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	Clinical data
☒	Dual use research of concern

Methods

n/a	Involved in the study
☒	ChIP-seq
☒	Flow cytometry
☒	MRI-based neuroimaging

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	Mouse: C57BL/6J, C57BL/6JRj, C57BL/N, C57B/6 S129J, C57BL/6J-LDLRKO, C57BL/6J-LEPTIN-KO, C57BL/6J-MC4RKO, db/db, ob/ob. Rat: RjHan:Wi, ZSF1-Leprfa Leprcp/Crl. Specific details for each individual model are included in Table S1 and Table S7
Wild animals	No wild animals were used in the study.
Reporting on sex	Experiments were mainly performed in Male rodents (21 F; 649 M). Sex information for each model is described in Tables S1 and S7.
Field-collected samples	No field collected samples were used in the study.
Ethics oversight	Relevant animal welfare authorities approved all the animal experiments that complied with national/international guidelines.

Note that full information on the approval of the study protocol must also be provided in the manuscript.