Variation in diagnostic NAFLD/NASH read-outs in paired liver samples from rodent models

Jensen, Victoria S.; Tveden-Nyborg, Pernille; Zacho-Rasmussen, Christina; Quaade, Michelle L.; Ipsen, David H.; Hvid, Henning; Fledelius, Christian; Wulff, Erik M.; Lykkesfeldt, Jens

Variation in diagnostic NAFLD/NASH read-outs in paired liver samples from rodent models

Victoria S. Jensen[a,b,*], Pernille Tveden-Nyborg[a], Christina Zacho-Rasmussen[a], Michelle L. Quaade[a], David H. Ipsen[a], Henning Hvid[b], Christian Fledelius[b], Erik M. Wulff[b], Jens Lykkesfeldt[a,*]

[a] *Section of Experimental Animal Models, Department of Veterinary and Animal Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, Ridebanevej 9, DK-1870 Frederiksberg, Denmark*
[b] *Global Research, Novo Nordisk A/S, Novo Nordisk Park 1, DK-2760 Maaloev, Denmark*

ARTICLE INFO

ABSTRACT

Introduction: In animal models of non-alcoholic fatty liver disease (NAFLD), assessment of disease severity and treatment effects of drugs rely on histopathological scoring of liver biopsies. However, little is known about the sampling variation in liver samples from animal models of NAFLD, even though several histopathological hallmarks of the disease are known to be affected by sampling variation in patients. The aim of this study was to assess the sampling variation in multiple paired liver biopsies from three commonly used diet-induced rodent models of NAFLD.

Methods: Eight male C57BL/6 mice, 8 male Sprague Dawley rats and 16 female Hartley guinea pigs were fed a NAFLD-inducing high-fat diet for 16 weeks (mice and rats), 20 or 24 weeks (guinea pigs). After the initial diet period, liver sections were sampled and subsequently assessed by histopathological scoring and biochemical analyses.

Results: Fibrosis was heterogeneously distributed throughout the liver in mice, manifesting as both intra- and interlobular statistically significant differences. Hepatic triglyceride content showed interlobular differences in mice, and both intra- and interlobular differences in guinea pigs (24-week time point) all of which were statistically significant. Also, hepatic cholesterol content was subject to significant intra-lobular sampling variation in mice, and hepatic glycogen content differed significantly between lobes in mice and guinea pigs.

Discussion: Dependent on animal model, both histopathological and biochemical end-points differed between sampling sites in the liver. Based on these findings, we recommend that sample site location is highly standardized and properly reported in order to minimize potential sampling variation and to optimize reproducibility and meaningful comparisons of preclinical studies of NAFLD.

1. Introduction

Non-alcoholic fatty liver disease (NAFLD) has become one of the most frequent causes of chronic liver disease in the Western world, affecting ≈25% of the population and currently ranking as the second most common aetiology among adults awaiting liver transplantation in the United States (Wong et al., 2015; Younossi et al., 2011; Younossi et al., 2016). The main diagnostic hallmark of NAFLD is the deposition of excess fat in the liver (hepatic steatosis); however, the term "NAFLD" encompasses several stages of liver disease with increasing pathological severity (Ipsen, Lykkesfeldt, & Tveden-Nyborg, 2018). Uncomplicated/simple hepatic steatosis (defined as accumulation of lipid in > 5% of hepatocytes) represents a relatively benign part of the disease spectrum. In some individuals, the disease progresses to non-alcoholic steatohepatitis (NASH), a condition characterized by coexisting hepatic inflammation, hepatocellular injury and increased oxidative stress, where

---

fibrosis may or may not be present (Loomba & Sanyal, 2013). In contrast to simple hepatic steatosis, NASH is associated with increased liver- and cardiovascular-related morbidity and mortality, and the accurate diagnosis and staging of NASH is therefore a vital prognostic tool (Anstee, Targher, & Day, 2013).

At present, histopathologic assessment of percutaneously obtained liver biopsies is the gold standard in the diagnosis of NASH and hepatic fibrosis in humans (Kleiner & Bedossa, 2015; Sanyal et al., 2015). However, the invasive nature of the procedure and the inherent risk imposed on the patients makes it unsuitable as a screening tool for population studies. This has prompted a search for alternative diagnostic modalities, such as the use of imaging techniques and clinically relevant surrogate biomarkers (Bedossa & Patel, 2016; Sanyal et al., 2015). However, no non-invasive diagnostic method has yet been found equivalent to the histopathologic evaluation of a liver biopsy (Sumida, Nakajima, & Itoh, 2014; Wieckowska, McCullough, & Feldstein, 2007).

Besides the risk associated with obtaining liver biopsies in general, the use of small-sample histopathology as primary diagnostic tool has also been questioned because of the potential risk of sampling variation (Ratziu et al., 2005). A standard biopsy approximates just 1/50000–1/65000 of the total mass of the human liver (Bravo, Sheth, & Chopra, 2001) and although guidelines for optimal biopsy size and length are available (Schiano et al., 2005; Vuppalanchi et al., 2007), it does not exclude potential measurement errors. Several studies in patients with NAFLD/NASH and other chronic liver diseases have reported heterogeneity in the distribution of histopathologic lesions when biopsies taken from different parts of the liver were compared (Abdi, Millan, & Mezey, 1979; Baunsgaard, Sanchez, & Lundborg, 1979; Bedossa, Dargere, & Paradis, 2003; Janiec, Jacobson, Freeth, Spaulding, & Blaszyk, 2005; Larson et al., 2007; Merriman et al., 2006; Ratziu et al., 2005; Regev et al., 2002).

It is not known if the distribution of NAFLD associated histopathologic lesions also differs between different anatomical parts of the liver in commonly used animal models of NAFLD. The topic has not received much attention in the literature and specific information on sampling methodology in experimental studies describing liver histopathology in animal models of NAFLD is often lacking or not sufficiently detailed (Brunt, 2008). In addition to taking essential measures to avoid bias and type 1 errors (false positives) by randomizing and blinding, sampling variation should also be considered as a contributor to increased variation and the risk of type 2 errors (false negatives), subsequently compromising the ability to detect a true difference should it be present. If these design-related issues are neglected the reproducibility and reliability of readouts from experimental studies, in this case in NAFLD/NASH, are reduced. Recent awareness on the diminished reproducibility of preclinical studies within several fields of research has emphasized the need for increased attention as to how to reduce variation and bias, and improve the validity of reported findings (Baker, 2016; Begley & Ellis, 2012; Goodman, Fanelli, & Ioannidis, 2016; Prinz, Schlange, & Asadullah, 2011). The aim of this study was to characterize sampling variation in commonly evaluated histopathologic and biochemical readouts from diet-induced rodent models of NAFLD. To explore potential species-differences and increase the general applicability of our findings, three routinely used rodent species (mice, rats and guinea pigs) were included in the study.

## 2. Materials and methods

2.1. Animals

Eight male Sprague-Dawley rats were purchased from Charles River Laboratories (Sulzfeld, Germany), eight male C57BL/6 mice from Taconic Biosciences (Lille Skensved, Denmark) and 16 female Hartley guinea pigs from Envigo (Venray, The Netherlands). Animals were acclimatized for one week prior to study initiation and were seven to eight weeks (guinea pigs, mice) or eight to nine weeks (rats) on arrival. Rats

and mice were housed two and ten per cage, respectively. Guinea pigs were group-housed in a floor pen. All animals had unrestricted access to non-chlorified, non-acidic tap-water and standard rodent chow (mice and rats: 1324 Altromin, Brogaarden, Denmark; guinea pigs: standard chow S9406-S020, Ssniff Spezialdiäten GmbH, Soest, Germany) prior to study initiation. Temperature in the animal rooms was maintained at 20–25 °C with a light/dark cycle of 12/12 h, air-change 8–15 times/h and a relative humidity of 30–70%. The study was approved by the Danish Animal Experiments Inspectorate under the Ministry of Food, Agriculture and Fisheries and in accordance with European Union directive 2010/63/EU.

2.2. Experimental design

After the acclimatization period, mice and rats were placed on a high-fat/high-fructose/high-cholesterol diet (NASH; (40% fat, 20% fructose, 2% cholesterol), D09100301, Research Diets, NJ, USA) for 16 weeks ($n = 8$/group) and guinea pigs on a high-fat/high-sucrose/high-cholesterol diet (HFS; (20% fat, 15% sucrose, 0.35% cholesterol), S9406-S025, Ssniff Spezialdiäten GmbH, Soest, Germany) for 20 or 24 weeks (n = 8/time point) based on the previously confirmed NAFLD/NASH induction in the models (Ipsen et al., 2016; Ipsen et al., 2018; Jensen et al., 2018; Trevaskis et al., 2012). Diets were stored at −20 °C throughout the study period and fresh aliquots thawed twice weekly to prevent putrefaction. After 16 weeks, mice and rats were euthanized by exsanguination under isoflurane anesthesia by incision of the heart. After 20 or 24 weeks, guinea pigs were anaesthetized and euthanized as described previously (Tveden-Nyborg et al., 2016). Immediately following euthanasia livers were excised and the left lateral lobe (L1), the right medial lobe (L2) and the caudate lobe (L3) were sampled for histologic and biochemical analyses. Three sampling sites were chosen in the distal (L1A), middle (L1B) and proximal (L1C) section of L1, and one sampling site from the middle section of L2 and L3 to facilitate evaluation of both intra- and interlobular sampling variation (Fig. 1). The chosen sampling sites correspond to those currently recommended by RITA (Registry of Industrial Toxicology Animal-data) for toxicopathological studies (Ruehl-Fehlert et al., 2003), with the addition of sample site L1A and L1C to facilitate intralobular comparison. From each sampling site, two adjacent 0.4 mm slices were collected and transferred to either 10% Neutral Buffered Formalin (Hounisen Laboratorieudstyr A/S, Skanderborg, Denmark) or snap-frozen in liquid nitrogen for biochemical analyses.
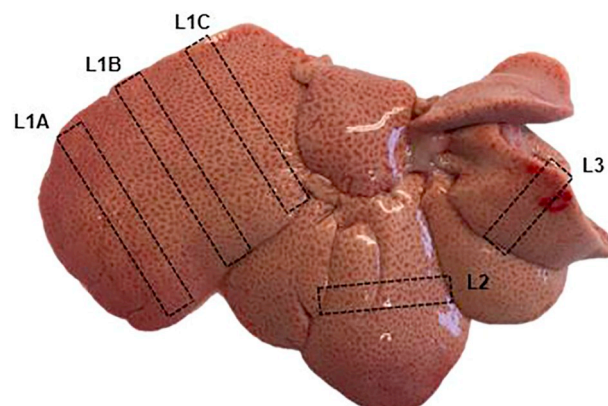
Fig. 1. Illustration of the five sampling sites used in the study sampling protocol. To facilitate assessment of the extent of both intra- and interlobular sampling variation, the left lateral lobe (L1A-C), the right medial lobe (L2) and the caudate lobe (L3) were sampled according to current RITA guidelines [24]. From each sampling site, two adjacent liver sections were sampled for histopathological and biochemical analyses.

Table 1
Histopathological readouts. Comparison of histopathological scores from different sampling sites in the liver of mice, rats and guinea pigs after 16–24 weeks on high-fat diets. Data are presented as medians with ranges (ranges calculated as maximum score – minimum score). Data were analyzed using Wilcoxon signed-rank tests. Scores designated [a] and [b] indicates statistically significant ($p < .05$) differences compared to sampling site L1A and L1B respectively.

	Sampling site					Cohens κ
	L1A	L1B	L1C	L2	L3	
Mice						
Steatosis	3 (1)	3 (0)	3 (1)	3 (2)	3 (0)	0.86
Lobular inflammation	1 (2)	1 (2)	1 (2)	0 (1)	1 (3)	0.84
Portal inflammation	0 (1)	0 (1)	0 (1)	0 (1)	0 (0)	0.72
Ballooning	0 (0)	0 (0)	0 (0)	0 (1)	0 (0)	−0.03
Fibrosis	**4 (4)**	**4 (3)**	**1 (2)[b]**	**1.5 (4)**	**1 (4)[ab]**	0.97
Rats						
Steatosis	3 (1)	3 (1)	3 (1)	2 (1)	3 (1)	0.87
Lobular inflammation	1 (0)	1 (1)	1 (1)	0 (1)	0.5 (2)	0.80
Portal inflammation	0.5 (1)	0 (1)	0.5 (1)	0 (1)	0 (1)	1
Ballooning	0 (1)	0 (1)	0 (0)	0 (0)	0 (1)	0.36
Fibrosis	1 (1)	0.5 (1)	0.5 (1)	0.5 (1)	0 (2)	0.91
Guinea pigs (20 weeks)						
Steatosis	1 (2)	1 (2)	1 (2)	1 (3)	1 (1)	0.90
Lobular inflammation	0 (1)	0.5 (1)	0 (1)	0 (1)	0 (1)	0.64
Portal inflammation	0 (0)	0 (1)	0 (0)	0 (0)	0 (1)	0.66
Ballooning	0 (1)	0.5 (1)	0 (0)	0 (1)	0 (1)	0.66
Fibrosis	1 (2)	1 (4)	1 (5)	1 (1)	1 (2)	0.95
Guinea pigs (24 weeks)						
Steatosis	2 (1)	2.5 (1)	3 (0)	3 (1)	3 (1)	0.90
Lobular inflammation	1 (2)	1 (2)	1 (2)	1 (1)	0.5 (3)	0.80
Portal inflammation	0 (0)	0 (1)	0 (1)	0 (1)	0 (1)	0.72
Ballooning	0 (1)	0 (1)	1 (1)	0.5 (1)	1 (1)	0.60
Fibrosis score	3.5 (4)	5 (4)	5 (3)	5 (3)	5 (0)	1

2.3. Histology

The proximal section from each sampling site was embedded in paraffin after fixation. Sections of 2–4 μm thickness were subsequently stained with Mayer's Haematoxylin (H&E) or Masson's Trichrome and used for evaluation of steatosis, lobular inflammation, portal inflammation, ballooning (H&E) and fibrosis (Masson's Trichrome). Histological evaluation was performed in a blinded fashion by two independent observers, according to the semi-quantitative scoring system suggested by (Kleiner et al., 2005), with few modifications as follows: Steatosis grade was evaluated on entire H&E-stained sections and scored from 0 to 3, where 0: < 5% lipid-containing hepatocytes, 1: 5–33%, 2: > 33%–66% and 3: > 66%. Lobular inflammation was evaluated by counting the number of inflammatory foci (the observation of at least 3 inflammatory cells in close proximity) per lobule in five randomly chosen lobules at $20\times$ magnification and was graded from 0 to 3, where 0: no foci, 1: < 2 foci, 2: 2–4 foci, 3: > 4 foci. Portal inflammation was scored as 0: not present, 1: present. The presence of ballooning hepatocytes was evaluated in five randomly chosen rectangular areas of 0.3 mm² at fields of $40\times$ magnification in areas with steatosis and scored as 0: not present, 1: few present (≤ 4 ballooning hepatocytes), 2: many present (> 4 ballooning hepatocytes). Lastly, fibrosis was evaluated on entire sections stained by Masson's Trichrome and scored from 0 to 4 where 0: not present; 1: perisinusoidal *or* periportal; 1A: mild, zone 3, perisinusoidal; 1B: moderate, zone 3, perisinusoidal; 1C: portal/periportal; 2: perisinusoidal *and* portal/periportal; 3: bridging fibrosis; 4: cirrhosis. Fibrosis scores were for data analysis purposes subsequently translated to a numeric scale from 0 to 6, where 0 = 0, 1 = 1A, 2 = 1B, 3 = 1C, 4 = 2, 5 = 3 and 6 = 4. Following the completion of the scorings, weighted or unweighted Cohen's Kappa (κ) index was calculated for each histological parameter to determine inter-observer variability (McHugh, 2012). κ-values < 0 indicate no agreement; κ = 0.01–0.2: None to slight; κ = 0.21–0.40: Fair agreement; κ = 0.41–0.60: Moderate agreement; κ = 0.61–0.80: Substantial agreement; and κ = 0.81–1.00: Almost perfect agreement.

## 2.4. Biochemistry

The distal section from each sampling site was snap-frozen in liquid nitrogen and stored at −80 °C prior to biochemical analyses. The levels of triglyceride, cholesterol and glycogen were analyzed on homogenized tissue from each sampling site as previously described (Jensen et al., 2018) using a Cobas 6000 c501 instrument (Roche Diagnostics GmbH, 68206 Mannheim, Germany) according to manufacturer's instructions.

2.5. Statistics

Statistical analyses were performed using GraphPad Prism version 8.02 (GraphPad Software Inc., La Jolla, CA, USA). Differences in histologic scores between and within hepatic lobes were analyzed using Wilcoxon signed-rank tests. Histological data are presented as medians with ranges (ranges calculated by subtracting the minimum histopathological score from the maximum histopathological score). Depending on scoring parameter, unweighted and weighted κ-values were calculated to estimate inter-observer reliability as outlined above. Due to significant inter-animal variation, biochemical data were normalized prior to analysis (normalization to the mean of all sample sites within each animal) in order to focus on the intra-animal sampling variation. After assuring that the assumption of no variance inhomogeneity was met, normalized data were analyzed by repeated-measures one-way ANOVA with Tukey's post hoc test for multiple comparisons. Post hoc tests were performed only when the level of F reached statistical significance ($p < .05$). Biochemical data are presented in graphs showing normalized data and in tables as means with standard deviations. Retrospective power calculations for detection of intra- and interlobular differences in steatosis grade and fibrosis stage were performed using the observed variation of the data set, and a significance level, α, of 0.05. Calculations of minimal detectable intra- and interlobular differences for biochemical parameters were performed using the observed variation of the data set; power, β, set at
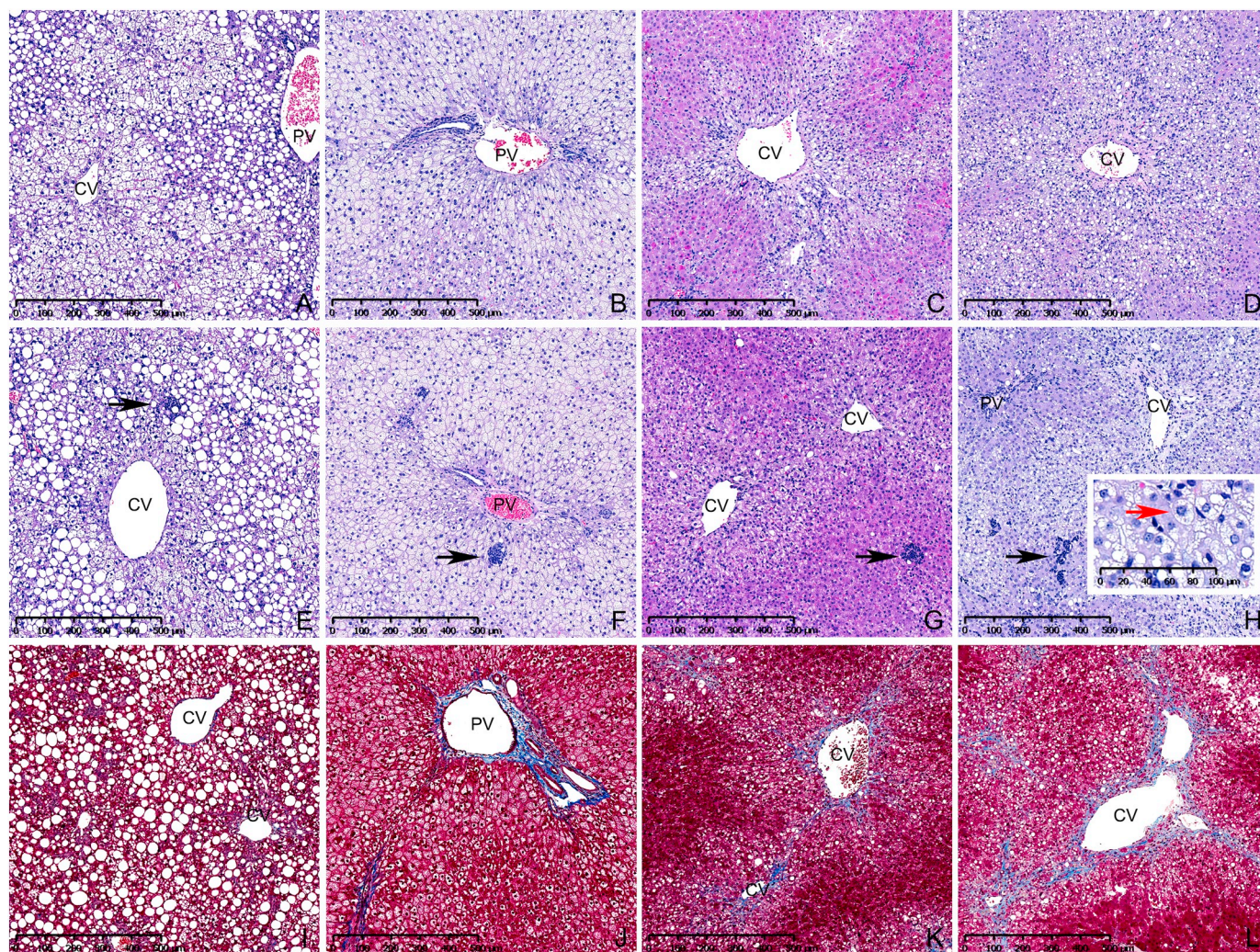
Fig. 2. Histological images of liver sections from mice, rats and guinea pigs. Row 1: Representative H&E stains of steatosis in mice (A); rats (B); guinea pigs 20 weeks (C) and guinea pigs 24 weeks (D). Row 2: Magnification of representative H&E stains showing inflammatory infiltrates (indicated by black arrows) in mice (E), rats (F); guinea pigs 20 weeks (G) and guinea pigs 24 weeks (H). Row 3: Representative Masson's Trichrome stains of fibrosis (blue staining) in mice (I), rats (J), guinea pigs 20 weeks (K) and guinea pigs 24 weeks (L). Insert in H: Example of ballooning hepatocyte in guinea pig liver (indicated by red arrow). CV: Central Vein; PV: Portal Vein.

0.80 and significance level, α, set at 0.05. For all reported data, *p*-values < .05 were considered statistically significant.

3. Results

3.1. Effects of diet on liver histology

Histological scores from all sampling sites in each species are provided in Table 1, with corresponding Cohen's κ-values confirming inter-observer reliability for histopathological readouts. Liver sections representative of histopathological changes and lesions from each species are shown in Fig. 2.

3.1.1. Mice

After 16 weeks on diet, mice developed hepatic steatosis, with a median steatosis grade of 3 at all 5 sampling sites. Across all evaluated sections, 2/40 (5%) were scored as grade 1, 3/40 (7.5%) as grade 2 and 35/40 (87.5%) as grade 3, in agreement with a majority of moderate to severe steatosis (Fig. 2A, Table 1). The steatosis was of mixed macro−/microvesicular type and involved both zone 1, 2 and 3 at the majority of the sampling sites. However, in sampling sites with moderate steatosis, the steatosis was found to originate in zone 3. Modest lobular

inflammation was present, indicated by a median lobular inflammation score of 1 at all sampling sites, except L2 (median score = 0). Portal inflammation and ballooning was only observed as occasional findings (grade 1 in 6/40 (15%) and ballooning in 2/40 (5%)). Importantly, no intra-individual sampling variation was observed for steatosis, lobular inflammation, portal inflammation or ballooning in mice. The diet induced variable degrees of fibrosis, indicated by a score ≥ 1 in 33/40 (82.5%) of the evaluated sampling sections (Table 1), spanning from observations of mild (25%) to moderate (17.5%) perisinusoidal fibrosis and combinations of perisinusoidal/periportal fibrosis (32.5%) to bridging lesions (7.5%). A predilection for the fibrotic lesions to the periphery of the parenchyma (subcapsular) was noted, but not quantified. For fibrosis score, intralobular and interlobular sampling variation was observed, indicated by a higher median score in L1B compared to L1C (4 vs. 1), and in both L1A and L1B compared to L3 (4 vs. 1, Table 1). Cohen's κ-values for inter-observer reliability ranged from substantial to almost perfect agreement for all scored parameters, except for ballooning (Table 1).

3.1.2. Rats

Similarly to mice, rats developed hepatic steatosis after 16 weeks, with a median steatosis score of 3 at all sampling sites, except L2

Table 2

Biochemical readouts. Absolute values of triglyceride, cholesterol and glycogen content from different sampling sites of the liver in mice, rats and guinea pigs after 16–24 weeks on high-fat diets. Data are presented as means with standard deviations. Values designated [a, b, c, d, e] indicate statistically significant differences ($p < .05$)) compared to L1A, L1B, L1C, L2 and L3, respectively. *Statistical analyses were performed on normalized data, not absolute values.

Biochemical parameter*	Sampling site				
	L1A	L1B	L1C	L2	L3
Mice					
Triglyceride (μmol/g liver)	239.7 ± 51.7	200.3 ± 53.7	192.3 ± 51.0	167.8 ± 53.5[e]	192.4 ± 55.4[d]
Cholesterol (μmol/g liver)	32.5 ± 9.5[c]	26.0 ± 10.5	24.6 ± 9.6[a]	26.7 ± 10.5	28.7 ± 11.3
Glycogen (μmol/g liver)	197.8 ± 45.6	210.1 ± 70.7	207.0 ± 65.6	230.1 ± 56.5[e]	175.9 ± 25.2[d]
Rats					
Triglyceride (μmol/g liver)	72.6 ± 9.4	71.5 ± 14.7	75.2 ± 12.2	71.0 ± 12.4	83.3 ± 16.0
Cholesterol (μmol/g liver)	66.2 ± 20.9	56.5 ± 10.4	53.0 ± 9.8	54.4 ± 5.9	60.4 ± 8.7
Glycogen (μmol/g liver)	200.1 ± 30.7	200.8 ± 24.3	186.6 ± 31.8	218.1 ± 23.0	178.9 ± 38.2
Guinea pigs (20 weeks)					
Triglyceride (μmol/g liver)	49.2 ± 14.8	49.2 ± 17.0	51.0 ± 10.2	50.9 ± 12.5	57.1 ± 12.9
Cholesterol (μmol/g liver)	32.3 ± 9.3	30.7 ± 11.2	30.3 ± 10.6	29.7 ± 7.1	31.6 ± 8.7
Glycogen (μmol/g liver)	68.4 ± 61.3[d]	61.5 ± 60.5	58.6 ± 55.4	60.5 ± 57.1[a]	53.5 ± 51.4
Guinea pigs (24 weeks)					
Triglyceride (μmol/g liver)	67.4 ± 11.6[be]	72.5 ± 13.2[ae]	69.9 ± 10.5	73.1 ± 13.4	79.3 ± 13.2[ab]
Cholesterol (μmol/g liver)	47.3 ± 8.6	48.5 ± 8.4	48.4 ± 10.3	46.4 ± 7.8	48.7 ± 8.6
Glycogen (μmol/g liver)	39.4 ± 28.2	37.1 ± 26.2	39.6 ± 30.4	34.9 ± 27.9	32.0 ± 25.3

(median score = 2). Across all evaluated sections 13/40 (32%) were scored as a grade 2 and 27/40 (67.5%) as a grade 3, indicating development of moderate to severe steatosis (Fig. 2B, Table 1). The steatosis was primarily of the microvesicular type and located in zone 1, with frequent involvement of zone 2 and occasionally zone 3. Mild lobular inflammation was present, with 27/40 (67.5%) of all sections scored as a grade 1, and 1/40 (2.5%) scored as a grade 2. The remaining sections were evaluated as a grade 0 (12/40 (30%), no inflammatory foci). Portal inflammation seemed more predominant in rats compared to mice, with 14/40 (35%) sections scored as grade 1. Ballooning was rarely observed, with occasional scorings of grade 1 in only 4/40 (10%) of sample sections evaluated. Fibrosis was not present in 19/40 (47.5%) sample sections, however, mild, perisinusoidal fibrosis located in zone 3 was observed in 20/40 (50%) sections, while only 1/40 (2.5%) sections displayed moderate, perisinusoidal zone 3 fibrosis. No significant intra-individual intralobular or interlobular sampling variation was observed for steatosis, lobular inflammation, portal inflammation, ballooning or fibrosis in rats. Cohen's κ-values for inter-observer reliability ranged from substantial to almost perfect agreement for all scored parameters, except for ballooning (Table 1).

3.1.3. Guinea pigs (20 weeks)

After 20 weeks on diet, guinea pigs developed mild hepatic steatosis, indicated by a median steatosis grade of 1 at all 5 sampling sites (Table 1). Across all evaluated sections, 5/40 (12.5%) were evaluated as grade 0; 23/40 (57.5%) as grade 1; 11/40 (27.5%) as grade 2 and only 1/40 (2.5%) liver sections as grade 3. The steatosis was mixed macro- and microvesicular and originated in zone 3. Lobular inflammation was modest, with the majority of sections (26/40, 65%) evaluated as grade 0 and 14/40 (35%) evaluated as grade 1. Portal inflammation was observed in only 2/40 (5%) scored sections, whereas ballooning was present in 8/40 (20%) of sections. The majority of animals displayed some degree of fibrosis, indicated by a median score of 1 at all 5 sampling sites. Across all evaluated sections, 4/40 (10%) was scored as non-fibrotic, 33/40 (82.5%) scored as mild, or moderate perisinusoidal zone 3 fibrosis and 3/40 (7.5%) sections evaluated as bridging fibrosis. No intra-individual sampling variation was seen for any of the histopathological traits. Cohen's κ-values for inter-observer reliability ranged from substantial to almost perfect agreement for all scored parameters.
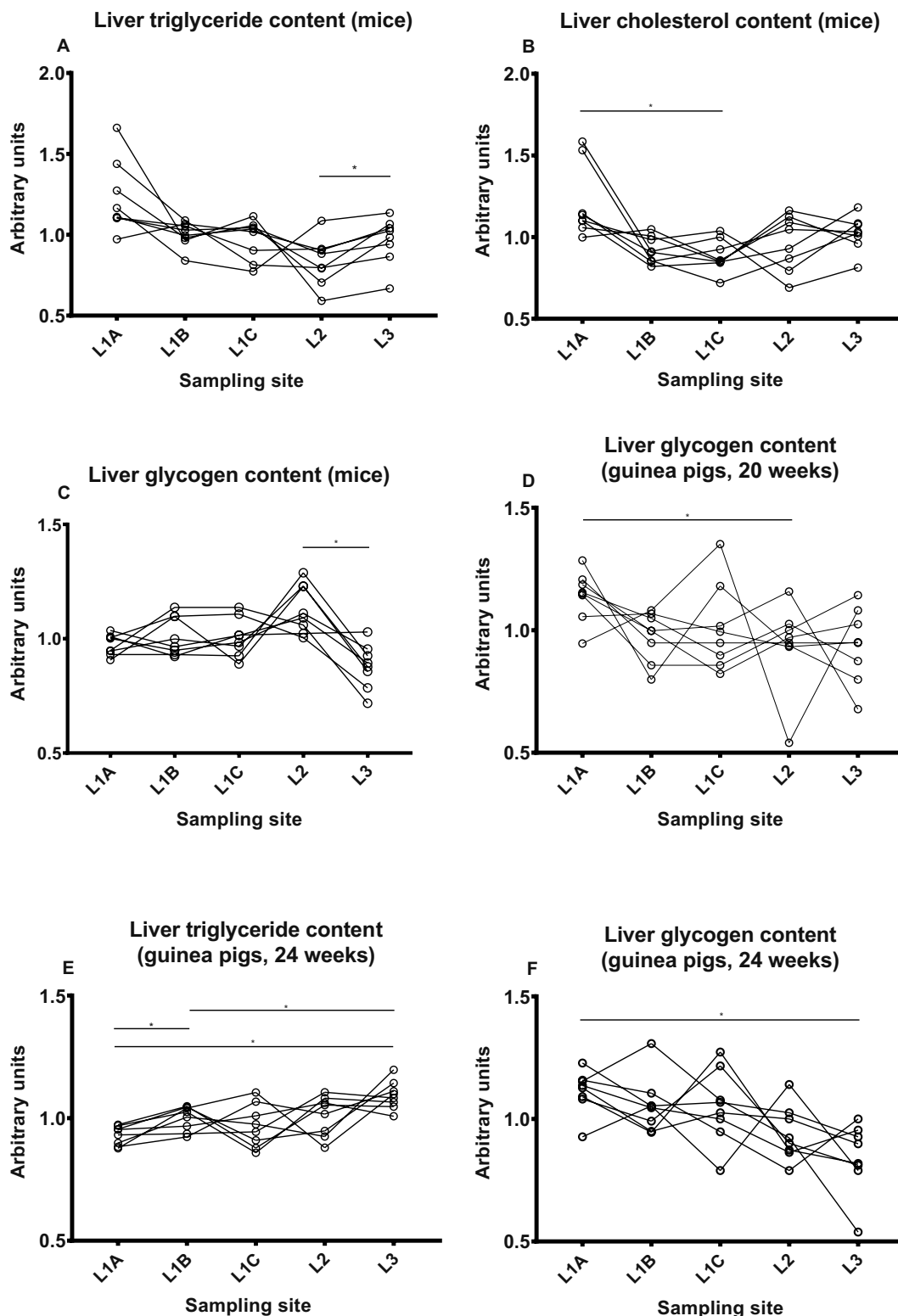
3.1.4. Guinea pigs (24 weeks)

After 24 weeks on diet, liver pathology in guinea pigs was markedly progressed compared to the 20-week time point (Table 1). Hepatic steatosis was present in all animals and characterized as moderate to severe, indicated by medians ranging from 2 to 3 depending on sampling site. The morphological characteristics and location of the steatosis was similar to the 20-week time point. Evaluated across all scored sections, 13/40 (32.5%) were scored as grade 2, while the majority (27/40; 67.5%) were scored as grade 3. Lobular inflammation was present at all 5 sampling sites, indicated by a median score of 1, except L3 where the median score was 0.5. Across all sections 15/40 (37.5%) was scored as grade 0 (no foci present); 20/40 (50%) as grade 1; 4/40 (10%) as grade 2 and just 1/40 (2.5%) as grade 3. Similarly to the 20-week time point, portal inflammation was a rare finding with only 6/40 (15%) sections receiving a score of 1. However, ballooning was more commonly identified compared to the 20-week time point, with 19/40 (47.5%) of all scored sections receiving a score of 1. The remaining sections (21/40; 52.5%) did not show ballooning. Fibrosis was extensively present at the 24-week time point, indicated by median scores of 3.5–5 depending on sampling site. The fibrosis was of varying severity, with 9/40 (22.5%) sections showing evidence of mild to moderate perisinusoidal fibrosis; however, the majority of scored sections (31/40; 77.5%) was evaluated as bridging fibrosis. As for the 20-week time point, no sampling variation was detected for any of the histopathological traits after 24 weeks on diet. Cohen's κ-values indicated substantial to almost perfect agreement for all traits, except for ballooning, where agreement was found to be fair.

3.2. Effects of diet on liver biochemistry

Absolute values of hepatic triglyceride, cholesterol and glycogen content at euthanasia are shown for each species in Table 2, and Fig. 3 displays the inter- and intralobular sampling variation for each biochemical readout.

3.2.1. Mice

Mean hepatic inter- and intralobular triglyceride levels in mice ranged between 167.8 and 239.7 μmol/g; cholesterol levels between 24.6 and 32.5 μmol/g; and glycogen levels between 175.9 and 230.1 μmol/g depending on sample site (Table 2). For both hepatic

Fig. 3. Sampling variation in biochemical parameters in mice, rats and guinea pigs after 16, 20 or 24 weeks on a high-fat diet. Triglyceride-, cholesterol- and glycogen content affected by sampling variation between and/or within individual lobes at study termination in mice (A-C), guinea pigs at the 20-week time point (D) and guinea pigs at the 24-week time point (E-F). Data is illustrated graphically as normalized data (normalization to the mean of all sample sites within each animal). Each circle on the graphs represents a normalized data point from an individual animal within the group. Normalized data were analyzed by repeated-measures one-way ANOVA with Tukey's post hoc test for multiple comparisons. * indicate statistically significant differences ($p < .05$).

triglyceride, cholesterol and glycogen content, statistically significant sampling variation was present. More specifically, triglyceride content was found to be significantly higher in L3 compared to L2 (Fig. 3A), cholesterol content was significantly higher in L1A compared to L1C (Fig. 3B) and lastly, glycogen content was significantly higher in L2 compared to L3 (Fig. 3C).

3.2.2. Rats

In rats, mean hepatic inter- and intralobular triglyceride levels ranged between 71.0 and 83.3 μmol/g, cholesterol levels between 53.0 and 66.2 μmol/g and glycogen levels between 178.9 and 218.1 μmol/g depending on sample site (Table 2). No statistically significant sampling variation was found for any of the biochemical readouts in rats.

3.2.3. Guinea pigs (20 weeks)

Mean hepatic triglyceride levels ranged between 49.2 and 57.1 μmol/g; cholesterol levels ranged between 29.7 and 32.3 μmol/g and glycogen levels ranged between 53.5 and 68.4 μmol/g depending on sampling site (Table 2). Statistically significant sampling variation was found only for glycogen content, where a higher level was observed in L1A compared to L2 (Fig. 3D).

3.2.4. Guinea pigs (24 weeks)

Biochemical readouts supported the histological findings and indicated further progression of hepatic steatosis compared to the 20-week time point, with mean hepatic triglyceride levels ranging between 67.4 and 79.3 μmol/g depending on sample site (Table 2). Similarly, cholesterol levels were higher and ranged between 46.4 and 48.7 μmol/g while glycogen levels had decreased and ranged between 32.0 and 39.6 μmol/g. For hepatic triglyceride content, statistically significant sampling variation (inter- and intralobular) was evident. More specifically, significantly higher triglyceride levels were found in L3 compared to both L1A and L1B and in L1B compared to L1A (Fig. 3E). Interlobular sampling variation in glycogen content was indicated by significantly higher levels in L1A compared to L3 (Fig. 3F). Cholesterol levels were not affected by sampling variation.

4. Discussion

Liver-lesions are commonly heterogeneously distributed and consequently, accurate staging and grading of NAFLD/NASH based on small liver-biopsies is a well-known challenge in patients (Sumida et al., 2014). Although the same may apply to animals, variation between sampling sites is often not considered in preclinical studies with experimental NAFLD models. In this study, we investigated sampling variation in three commonly applied rodent NAFLD models and found significant intra-individual differences in fibrosis score in mice and in several hepatic biomarkers in mice and guinea pigs. Rats did not display significant inter- or intralobular sampling variation. Our findings highlight that sampling variation may be a relevant concern when evaluating several commonly used endpoints in liver biopsies, at least in some rodent models of NAFLD.

With the increasing prevalence of NAFLD and NASH and no currently approved targeted pharmacological treatment, the need for preclinical testing of drug candidates is also increasing. Accurate and reproducible grading of liver fibrosis is of particular interest as it represents a primary diagnostic and prognostic hallmark in progressing NASH as well as an important treatment target (Chalasani et al., 2012), in agreement with the many potential anti-fibrotic drugs that are currently being explored in drug discovery programs (Musso, Cassader, & Gambino, 2016). Hence, the finding that liver fibrosis is affected by sampling variation in mice - the predominant species used in preclinical studies of NAFLD (Hansen et al., 2017) – is important. Moreover, preclinical NAFLD/NASH studies in mice are often challenged by the difficulty in obtaining an adequate fibrotic response in diet-induced models within a reasonable time frame. Furthermore, there is often considerable variability in the hepatic response (Duval et al., 2010; Farrell et al., 2014; Haczeyni et al., 2017), as also evident in the present study. Though this variability likely mirrors the severity and onset of NASH-related pathology reported from clinical settings, it is a concern that the variation in disease response may conceal treatment effects, for example in cases of unexpected mild hepatic histopathology ("low-responders") in either diet-induced or untreated groups (Hansen et al.,

2017). A proposed strategy to improve this is to collect pre-treatment liver biopsies (or small cone shaped wedges) from each animal, allowing stratification of animals according to disease stage prior to inclusion (Kristiansen et al., 2016). However, pre-treatment liver biopsies from animals possess the same risk factors as in humans and are also of limited size (i.e. 50–100 mg of liver tissue) (Kristiansen et al., 2016; Tolbol et al., 2018), constituting a much smaller fraction of total hepatic tissue than the samples included in our study. Furthermore, if pre-treatment biopsies are collected from another lobe than the biopsy collected at termination of the animal, it can potentially complicate detection of treatment-related effects, simply because fibrosis is heterogeneously distributed, as we demonstrate here.

A recent study examined the distribution of liver fibrosis by extensive image analysis on histological sections from the left lateral and right medial liver lobes in mice fed a NASH-diet similar to that of the present study, but for 30 weeks (Clapper et al., 2013). The authors found that within single liver sections in mice, fibrosis was in fact heterogeneously distributed, with 6.5-fold greater collagen content in the first 1 mm of the subcapsular parenchyma compared to areas > 3 mm from the section edges. Although we did not quantitatively compare the differences in peripheral and deeper tissue collagen deposition in single liver sections, we observed a similar trend in our study. The sampling protocol used by Clapper et al. only included the left lateral and right medial lobe of the liver and did not find differences in fibrosis between these two lobes. This is in line with our findings, where significant differences in fibrosis score were found only within the left lateral lobe (L1B vs L1C) and between the left lateral and the caudate lobe (L1A and L1B vs L3), whereas no significant difference was observed between the right medial lobe (L2) compared to the left lateral lobe (L1A and L1B). Besides a study that reported no differences in histopathological scoring criterions between the left lateral and the right medial lobe in a NAFLD/NASH guinea pig model (Tveden-Nyborg et al., 2016) there is to our knowledge, no other studies that have systematically examined intra- and interlobular sampling variation in rat or guinea pig models of NAFLD/NASH.

In humans with NAFLD/NASH and other chronic liver diseases, both fibrosis and other key histopathological features have been shown to be heterogeneously distributed (Abdi et al., 1979; Baunsgaard et al., 1979; Bedossa et al., 2003; Goldstein, Hastah, Galan, & Gordon, 2005; Janiec et al., 2005; Larson et al., 2007; Merriman et al., 2006; Ratziu et al., 2005; Regev et al., 2002). In a study examining single liver biopsies from 46 NASH-patients and 52 Hepatitis C Virus-affected patients, significant regional differences in fibrosis severity were present within the same biopsy, with a predilection of the lesions for the deeper parenchyma when compared to parenchyma closer to the capsule (Goldstein et al., 2005). Furthermore, studies evaluating paired liver biopsies confirm that NASH-related fibrosis is subject to sampling variation (Merriman et al., 2006; Ratziu et al., 2005). In one study, paired needle biopsies from the left and right lobe of the liver were examined in 41 obese patients with suspected NAFLD. Here, agreement in steatosis stage was good (κ = 0.88), while ballooning and lobular inflammation showed only fair agreement between liver lobes (κ = 0.20 and κ = 0.32 respectively), and sampling variation resulted in erroneous diagnosis of NASH and fibrosis in > 30% of cases (Merriman et al., 2006). Likewise, Ratziu et al. performed paired needle liver biopsies from different anatomical sites in the right liver lobe of 51 patients with suspected NAFLD. While agreement on steatosis score was relatively high between samples (78% of patients received the same steatosis score in both needle biopsies), there were inconsistencies of at least one category of fibrotic stage from paired samples in 41% of the patients. Furthermore, 35% of the patients diagnosed with bridging fibrosis in one sample, showed only mild or no fibrosis in the second (Ratziu et al., 2005). In contrast, another study performed on paired needle biopsies from the left and right lobe in 43 morbidly obese patients undergoing bariatric surgery showed excellent agreement between paired samples for both steatosis (κ = 0.91), ballooning

Table 3

Power calculations. Retrospective power calculations for detection of intra- and interlobular differences in steatosis grade and fibrosis stage were performed using a sample size of 8; the observed variation of the data set, and a significance level, α, of 0.05. Calculations of minimal detectable intra- and interlobular differences for biochemical parameters were performed using a sample size of 8; the observed variation of the data set; power, β, set at 0.80 and significance level, α, set at 0.05.

	Power to detect a difference of 1 in steatosis grade or a difference of 1 or 2 in fibrosis stage (%)			Minimal detectable difference in hepatic triglyceride content (%)
	Steatosis	Fibrosis (1 stage)	Fibrosis (2 stages)	
Mice	87	52	98	20
Rats	76	43	94	13
Guinea pigs (20 weeks)	52	54	98	21
Guinea pigs (24 weeks)	65	88	99	9

(κ = 0.73) and fibrosis (κ = 0.96) (Larson et al., 2007). One possible explanation for the discrepancy between these studies is the size and diameter of the obtained biopsy specimens, known to be an important factor in controlling sampling variation (Goldstein et al., 2005). The cause of heterogeneous distribution of liver fibrosis has been proposed to be caused by differences in sinusoidal blood levels of oxygens and antioxidants between and within different hepatic lobes (Ekataksin & Kaneda, 1999; Goldstein et al., 2005; Krogsgaard, Gluud, Henriksen, & Christoffersen, 1985; Takahashi, 1970). We suspect that these physiological factors are also likely to affect the distribution of fibrosis in animal models of the disease.

It is generally believed that steatosis, inflammation and ballooning precede the development of fibrosis (Cobbina & Akhlaghi, 2017) and these early lesions could therefore potentially also display a heterogeneous pattern of distribution. However, none of these histopathological endpoints displayed significant differences between sampling sites in our study. We cannot rule out that these features could be affected under different dietary regimens or by use of animals of different strains, ages, gender or disease progression state.

Another factor that might compromise the ability to detect differences in histopathological readouts between different sampling sites is the semi-quantitative nature of the grading and scoring system which could potentially mask minor albeit clinically relevant differences. Biochemical analysis revealed that hepatic lipid content was subject to sampling variation in both mice and guinea pigs, even though the differences were apparently too subtle to be detected histologically. Our ability to detect sampling variation in the biochemical readouts supports the importance of identifying quantitative surrogate markers for hepatic fibrosis, ballooning and inflammation (Bedossa & Patel, 2016). Once validated, such markers may provide a more accurate diagnosis, enable continued evaluation of disease progression/regression and thus identify smaller, yet potentially clinically relevant treatment effects in patients.

In order to estimate the reliability of our findings, we retrospectively calculated the statistical power for detecting intra- and interlobular differences of 1 in steatosis grade as well as differences of 1 or 2 stages in fibrosis score in all species (Table 3). Power calculations for lobular inflammation, portal inflammation and ballooning were not performed due to the limited occurrence of these histopathological traits in the study, and the modest inter-observer agreement for the majority of the ballooning scores. Despite the relatively modest sample sizes used in this study, these calculations returned a high power (power > 80%) for detecting differences of 1 and 2 stages in fibrosis score in guinea pigs (24 weeks), and for all species and time points for detecting a 2-stage difference (power > 90%). The statistical power for detecting differences in steatosis grade in mice and rats were also relatively high, but only modest for guinea pigs. We also calculated the minimal differences in hepatic triglyceride content this study design would enable us to detect. As seen in Table 3, the chosen sample sizes did allow for detection of relatively subtle differences (± 9–21% dependent on species) in hepatic triglyceride content, if such differences

existed. The results of the power analysis indicate that for most endpoints, the outcome variation was overall comparable between species. This increases the likelihood that the differences found between sample sites reflect sampling variation rather than species-specific variation in the biological endpoints.

5. Conclusion

The present study shows that liver fibrosis is heterogeneously distributed in a diet-induced mouse model of NASH, and that differences between sampling sites occur for biochemical readouts from both mice and guinea pigs. Liver biopsies from rats were apparently subject to less sampling variation. Even though statistically significant sampling variation was not observed for several of the assessed parameters, our findings still underline the importance of standardization of sampling site location when obtaining liver biopsies from experimental NASH-models, to minimize the risk of particularly type 2 errors and to facilitate meaningful comparisons of outcomes between independent experimental studies.

Author's contributions

The study was designed by VSJ, PTN, JL, HH, CF and EMW. The experiment was performed by VSJ, MLQ, CZ, DHI and PTN. Initial data analysis was carried out by VSJ, MLQ, CZ, DHI and JL followed by data interpretation by all authors. The draft manuscript was written by VSJ, revised by PTN, HH and JL and subsequently edited by all authors.

Declaration of Competing Interests

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

VSJ, HH, CF and EMW are current or former employees of Novo Nordisk and hold shares in the company. The other authors have no potential conflicts of interest to declare.

References

Abdi, W., Millan, J. C., & Mezey, E. (1979). Sampling variability on percutaneous liver biopsy. *Archives of Internal Medicine, 139*(6), 667–669.
Anstee, Q. M., Targher, G., & Day, C. P. (2013). Progression of NAFLD to diabetes

mellitus, cardiovascular disease or cirrhosis. *Nature Reviews. Gastroenterology & Hepatology, 10*(6), 330–344. https://doi.org/10.1038/nrgastro.2013.41.

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature, 533*(7604), 452–454. https://doi.org/10.1038/533452a.

Baunsgaard, P., Sanchez, G. C., & Lundborg, C. J. (1979). The variation of pathological changes in the liver evaluated by double biopsies. *Acta Pathologica et Microbiologica Scandinavica. Section A, 87*(1), 51–57.

Bedossa, P., Dargere, D., & Paradis, V. (2003). Sampling variability of liver fibrosis in chronic hepatitis C. *Hepatology, 38*(6), 1449–1457. https://doi.org/10.1016/j.hep.2003.09.022.

Bedossa, P., & Patel, K. (2016). Biopsy and noninvasive methods to assess progression of Nonalcoholic fatty liver disease. *Gastroenterology, 150*(8), 1811–1822. (e1814) https://doi.org/10.1053/j.gastro.2016.03.008.

Begley, C. G., & Ellis, L. M. (2012). Drug development: Raise standards for preclinical cancer research. *Nature, 483*(7391), 531–533. https://doi.org/10.1038/483531a.

Bravo, A. A., Sheth, S. G., & Chopra, S. (2001). Liver biopsy. *The New England Journal of Medicine, 344*(7), 495–500. https://doi.org/10.1056/NEJM200102153440706.

Brunt, E. M. (2008). Do you see what I see? The role of quality histopathology in scientific study. *Hepatology, 47*(3), https://doi.org/10.1002/hep.22185.

Chalasani, N., Younossi, Z., Lavine, J. E., Diehl, A. M., Brunt, E. M., Cusi, K., & Sanyal, A. J. (2012). The diagnosis and management of non-alcoholic fatty liver disease: Practice guideline by the American Association for the Study of Liver Diseases, American College of Gastroenterology, and the American Gastroenterological Association. *Hepatology, 55*(6), 2005–2023. https://doi.org/10.1002/hep.25762.

Clapper, J. R., Hendricks, M. D., Gu, G., Wittmer, C., Dolman, C. S., Herich, J., & Roth, J. D. (2013). Diet-induced mouse model of fatty liver disease and nonalcoholic steatohepatitis reflecting clinical disease progression and methods of assessment. *American Journal of Physiology. Gastrointestinal and Liver Physiology, 305*(7), G483–G495. https://doi.org/10.1152/ajpgi.00079.2013.

Cobbina, E., & Akhlaghi, F. (2017). Non-alcoholic fatty liver disease (NAFLD) - pathogenesis, classification, and effect on drug metabolizing enzymes and transporters. *Drug Metabolism Reviews, 49*(2), 197–211. https://doi.org/10.1080/03602532.2017.1293683.

Duval, C., Thissen, U., Keshtkar, S., Accart, B., Stienstra, R., Boekschoten, M. V., & Muller, M. (2010). Adipose tissue dysfunction signals progression of hepatic steatosis towards nonalcoholic steatohepatitis in C57BL/6 mice. *Diabetes, 59*(12), 3181–3191. https://doi.org/10.2337/db10-0224.

Ekataksin, W., & Kaneda, K. (1999). Liver microvascular architecture: An insight into the pathophysiology of portal hypertension. *Seminars in Liver Disease, 19*(4), 359–382. https://doi.org/10.1055/s-2007-1007126.

Farrell, G. C., Mridha, A. R., Yeh, M. M., Arsov, T., Van Rooyen, D. M., Brooling, J., & Larter, C. Z. (2014). Strain dependence of diet-induced NASH and liver fibrosis in obese mice is linked to diabetes and inflammatory phenotype. *Liver International, 34*(7), 1084–1093. https://doi.org/10.1111/liv.12335.

Goldstein, N. S., Hastah, F., Galan, M. V., & Gordon, S. C. (2005). Fibrosis heterogeneity in nonalcoholic steatohepatitis and hepatitis C virus needle core biopsy specimens. *American Journal of Clinical Pathology, 123*(3), 382–387. https://doi.org/10.1309/EY72-F1EN-9XCB-1KXX.

Goodman, S. N., Fanelli, D., & Ioannidis, J. P. (2016). What does research reproducibility mean? *Science Translational Medicine, 8*(341), 341ps312. https://doi.org/10.1126/scitranslmed.aaf5027.

Haczeyni, F., Poekes, L., Wang, H., Mridha, A. R., Barn, V., Geoffrey Haigh, W., & Farrell, G. C. (2017). Obeticholic acid improves adipose morphometry and inflammation and reduces steatosis in dietary but not metabolic obesity in mice. *Obesity (Silver Spring), 25*(1), 155–165. https://doi.org/10.1002/oby.21701.

Hansen, H. H., Feigh, M., Veidal, S. S., Rigbolt, K. T., Vrang, N., & Fosgerau, K. (2017). Mouse models of nonalcoholic steatohepatitis in preclinical drug development. *Drug Discovery Today, 22*(11), 1707–1718. https://doi.org/10.1016/j.drudis.2017.06.007.

Ipsen, D. H., Lykkesfeldt, J., & Tveden-Nyborg, P. (2018). Molecular mechanisms of hepatic lipid accumulation in non-alcoholic fatty liver disease. *Cellular and Molecular Life Sciences, 75*(18), 3313–3327. https://doi.org/10.1007/s00018-018-2860-6.

Ipsen, D. H., Rolin, B., Rakipovski, G., Skovsted, G. F., Madsen, A., Kolstrup, S., & Tveden-Nyborg, P. (2018). Liraglutide decreases hepatic inflammation and injury in advanced lean non-alcoholic Steatohepatitis. *Basic & Clinical Pharmacology & Toxicology, 123*(6), 704–713. https://doi.org/10.1111/bcpt.13082.

Ipsen, D. H., Tveden-Nyborg, P., Rolin, B., Rakipovski, G., Beck, M., Mortensen, L. W., & Lykkesfeldt, J. (2016). High-fat but not sucrose intake is essential for induction of dyslipidemia and non-alcoholic steatohepatitis in Guinea pigs. *Nutrition & Metabolism (London), 13*, 51. https://doi.org/10.1186/s12986-016-0110-1.

Janiec, D. J., Jacobson, E. R., Freeth, A., Spaulding, L., & Blaszyk, H. (2005). Histologic variation of grade and stage of non-alcoholic fatty liver disease in liver biopsies. *Obesity Surgery, 15*(4), 497–501. https://doi.org/10.1381/0960892053723268.

Jensen, V. S., Hvid, H., Damgaard, J., Nygaard, H., Ingvorsen, C., Wulff, E. M., & Fledelius, C. (2018). Dietary fat stimulates development of NAFLD more potently than dietary fructose in Sprague-Dawley rats. *Diabetology and Metabolic Syndrome, 10*, 4. https://doi.org/10.1186/s13098-018-0307-8.

Kleiner, D. E., & Bedossa, P. (2015). Liver histology and clinical trials for nonalcoholic steatohepatitis-perspectives from 2 pathologists. *Gastroenterology, 149*(6), 1305–1308. https://doi.org/10.1053/j.gastro.2015.09.015.

Kleiner, D. E., Brunt, E. M., Van Natta, M., Behling, C., Contos, M. J., Cummings, O. W., Nonalcoholic Steatohepatitis Clinical Research, N. (2005). Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology, 41(6), 1313–1321. doi:https://doi.org/10.1002/hep.20701.

Kristiansen, M. N., Veidal, S. S., Rigbolt, K. T., Tolbol, K. S., Roth, J. D., Jelsing, J., & Feigh, M. (2016). Obese diet-induced mouse models of nonalcoholic steatohepatitis-

tracking disease by liver biopsy. *World Journal of Hepatology, 8*(16), 673–684. https://doi.org/10.4254/wjh.v8.i16.673.

Krogsgaard, K., Gluud, C., Henriksen, J. H., & Christoffersen, P. (1985). Correlation between liver morphology and haemodynamics in alcoholic liver disease. *Liver, 5*(3), 173–177.

Larson, S. P., Bowers, S. P., Palekar, N. A., Ward, J. A., Pulcini, J. P., & Harrison, S. A. (2007). Histopathologic variability between the right and left lobes of the liver in morbidly obese patients undergoing roux-en-Y bypass. *Clinical Gastroenterology and Hepatology, 5*(11), 1329–1332. https://doi.org/10.1016/j.cgh.2007.06.005.

Loomba, R., & Sanyal, A. J. (2013). The global NAFLD epidemic. *Nature Reviews. Gastroenterology & Hepatology, 10*(11), 686–690. https://doi.org/10.1038/nrgastro.2013.171.

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochem Med (Zagreb), 22*(3), 276–282.

Merriman, R. B., Ferrell, L. D., Patti, M. G., Weston, S. R., Pabst, M. S., Aouizerat, B. E., & Bass, N. M. (2006). Correlation of paired liver biopsies in morbidly obese patients with suspected nonalcoholic fatty liver disease. *Hepatology, 44*(4), 874–880. https://doi.org/10.1002/hep.21346.

Musso, G., Cassader, M., & Gambino, R. (2016). Non-alcoholic steatohepatitis: Emerging molecular targets and therapeutic strategies. *Nature Reviews. Drug Discovery, 15*(4), 249–274. https://doi.org/10.1038/nrd.2015.3.

Prinz, F., Schlange, T., & Asadullah, K. (2011). Believe it or not: How much can we rely on published data on potential drug targets? *Nature Reviews Drug Discovery, 10*(9), 712–U781. https://doi.org/10.1038/nrd3439-c1.

Ratziu, V., Charlotte, F., Heurtier, A., Gombert, S., Giral, P., Bruckert, E., & Group, L. S. (2005). Sampling variability of liver biopsy in nonalcoholic fatty liver disease. *Gastroenterology, 128*(7), 1898–1906.

Regev, A., Berho, M., Jeffers, L. J., Milikowski, C., Molina, E. G., Pyrsopoulos, N. T., & Schiff, E. R. (2002). Sampling error and intraobserver variation in liver biopsy in patients with chronic HCV infection. *The American Journal of Gastroenterology, 97*(10), 2614–2618. https://doi.org/10.1111/j.1572-0241.2002.06038.x.

Ruehl-Fehlert, C., Kittel, B., Morawietz, G., Deslex, P., Keenan, C., Mahrt, C. R., & Group, N (2003). Revised guides for organ sampling and trimming in rats and mice–part 1. *Experimental and Toxicologic Pathology, 55*(2–3), 91–106.

Sanyal, A. J., Friedman, S. L., McCullough, A. J., Dimick-Santos, L., & American Association for the Study of Liver, D., United States, F., & Drug, A (2015). Challenges and opportunities in drug and biomarker development for nonalcoholic steatohepatitis: Findings and recommendations from an American Association for the Study of Liver Diseases-U.S. Food and Drug Administration joint workshop. *Hepatology, 61*(4), 1392–1405. https://doi.org/10.1002/hep.27678.

Schiano, T. D., Azeem, S., Bodian, C. A., Bodenheimer, H. C., Jr., Merati, S., Thung, S. N., & Hytiroglou, P. (2005). Importance of specimen size in accurate needle liver biopsy evaluation of patients with chronic hepatitis C. *Clinical Gastroenterology and Hepatology, 3*(9), 930–935.

Sumida, Y., Nakajima, A., & Itoh, Y. (2014). Limitations of liver biopsy and non-invasive diagnostic tests for the diagnosis of nonalcoholic fatty liver disease/nonalcoholic steatohepatitis. *World Journal of Gastroenterology, 20*(2), 475–485. https://doi.org/10.3748/wjg.v20.i2.475.

Takahashi, T. (1970). Lobular structure of the human liver from the viewpoint of hepatic vascular architecture. *The Tohoku Journal of Experimental Medicine, 101*(2), 119–140.

Tolbol, K. S., Kristiansen, M. N. B., Hansen, H. H., Veidal, S. S., Rigbolt, K. T. G., Gillum, M. P., & Feigh, M. (2018). Metabolic and hepatic effects of liraglutide, obeticholic acid and elafibranor in diet-induced obese mouse models of biopsy-confirmed nonalcoholic steatohepatitis. *World Journal of Gastroenterology, 24*(2), 179–194. https://doi.org/10.3748/wjg.v24.i2.179.

Trevaskis, J. L., Griffin, P. S., Wittmer, C., Neuschwander-Tetri, B. A., Brunt, E. M., Dolman, C. S., & Roth, J. D. (2012). Glucagon-like peptide-1 receptor agonism improves metabolic, biochemical, and histopathological indices of nonalcoholic steatohepatitis in mice. *American Journal of Physiology. Gastrointestinal and Liver Physiology, 302*(8), G762–G772. https://doi.org/10.1152/ajpgi.00476.2011.

Tveden-Nyborg, P., Birck, M. M., Ipsen, D. H., Thiessen, T., Feldmann, L. B., Lindblad, M. M., & Lykkesfeldt, J. (2016). Diet-induced dyslipidemia leads to nonalcoholic fatty liver disease and oxidative stress in Guinea pigs. *Translational Research, 168*, 146–160. https://doi.org/10.1016/j.trsl.2015.10.001.

Vuppalanchi, R., Cummings, O. W., Saxena, R., Ulbright, T. M., Martis, N., Jones, D. R., & Chalasani, N. (2007). Relationship among histologic, radiologic, and biochemical assessments of hepatic steatosis: A study of human liver samples. *Journal of Clinical Gastroenterology, 41*(2), 206–210. https://doi.org/10.1097/01.mcg.0000225515.28536.3a.

Wieckowska, A., McCullough, A. J., & Feldstein, A. E. (2007). Noninvasive diagnosis and monitoring of nonalcoholic steatohepatitis: Present and future. *Hepatology, 46*(2), 582–589. https://doi.org/10.1002/hep.21768.

Wong, R. J., Aguilar, M., Cheung, R., Perumpail, R. B., Harrison, S. A., Younossi, Z. M., & Ahmed, A. (2015). Nonalcoholic steatohepatitis is the second leading etiology of liver disease among adults awaiting liver transplantation in the United States. *Gastroenterology, 148*(3), 547–555. https://doi.org/10.1053/j.gastro.2014.11.039.

Younossi, Z. M., Koenig, A. B., Abdelatif, D., Fazel, Y., Henry, L., & Wymer, M. (2016). Global epidemiology of nonalcoholic fatty liver disease-meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology, 64*(1), 73–84. https://doi.org/10.1002/hep.28431.

Younossi, Z. M., Stepanova, M., Afendy, M., Fang, Y., Younossi, Y., Mir, H., & Srishord, M. (2011). Changes in the prevalence of the most common causes of chronic liver diseases in the United States from 1988 to 2008. *Clinical Gastroenterology Hepatology, 9*(6), 524–530. e521; quiz e560 https://doi.org/10.1016/j.cgh.2011.03.020.