

# Quantitative Image Analysis of Hepatic Zonation in Cytochrome P450 and Steatosis Using Whole Slide Scans

## Masterarbeit

zum Erwerb des akademischen Grades Master of Science  
im Fach  
Biophysics

**Humboldt-Universität zu Berlin**  
**Lebenswissenschaftliche Fakultät**  
Institut für Biologie

vorgelegt von

Küttner, Jonas  
geb. am 03.05.1998, Rochlitz  
Matrikelnummer: 589388  
E-mail: jonas.kuettner98@gmail.com

Erstprüfer: Dr. Matthias König  
Zweitprüfer: Prof. Dr. Hans-Peter Herzel

Berlin, 29. August 2024

# Eigenständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig verfasst habe und sämtliche Quellen, einschließlich Internetquellen, die unverändert oder abgewandelt wiedergegeben werden, insbesondere Quellen für Texte, Grafiken, Tabellen und Bilder, als solche kenntlich gemacht habe. Ich versichere, dass ich die vorliegende Abschlussarbeit noch nicht für andere Prüfungen eingereicht habe. Mir ist bekannt, dass bei Verstößen gegen diese Grundsätze ein Verfahren wegen Täuschungsversuchs bzw. Täuschung gemäß der fachspezifischen Prüfungsordnung und/oder der Fächerübergreifenden Satzung zur Regelung von Zulassung, Studium und Prüfung der Humboldt-Universität zu Berlin (ZSP-HU) eingeleitet wird.

*Berlin, 29. August 2024*

---

Jonas Küttner

# Disclaimer

The image analysis workflow for lobule detection (Section 3.1), the statistical analysis of the lobule geometries (Section 3.3.1), and the zonated expression patterns (Section 3.4.1) that I developed in the context of this thesis, including the figures, were published as part of a research paper [1]. Figures that have been published as part of this publication are marked with a note in the figure caption.

## Summary

The mammalian liver is organized into three-dimensional structures called lobules that are integral to its function. According to the widely accepted hexagonal model, the cross-section of each lobule is characterized by a central vein and portal triads at the corners. These portal triads consist of a hepatic portal vein, a hepatic arteriole, and a bile duct. Blood is delivered through the hepatic artery and portal vein and flows inward from the outer periportal region toward the central vein through sinusoids lined with hepatocytes. These hepatocytes express a wide variety of metabolic enzymes, including cytochrome P450 (CYP) isoforms that are critical for xenobiotic metabolism. Notably, many hepatic enzymes, including CYPs, exhibit differential expression along the periportal-to-perivenous axis, a spatial variation known as liver zonation.

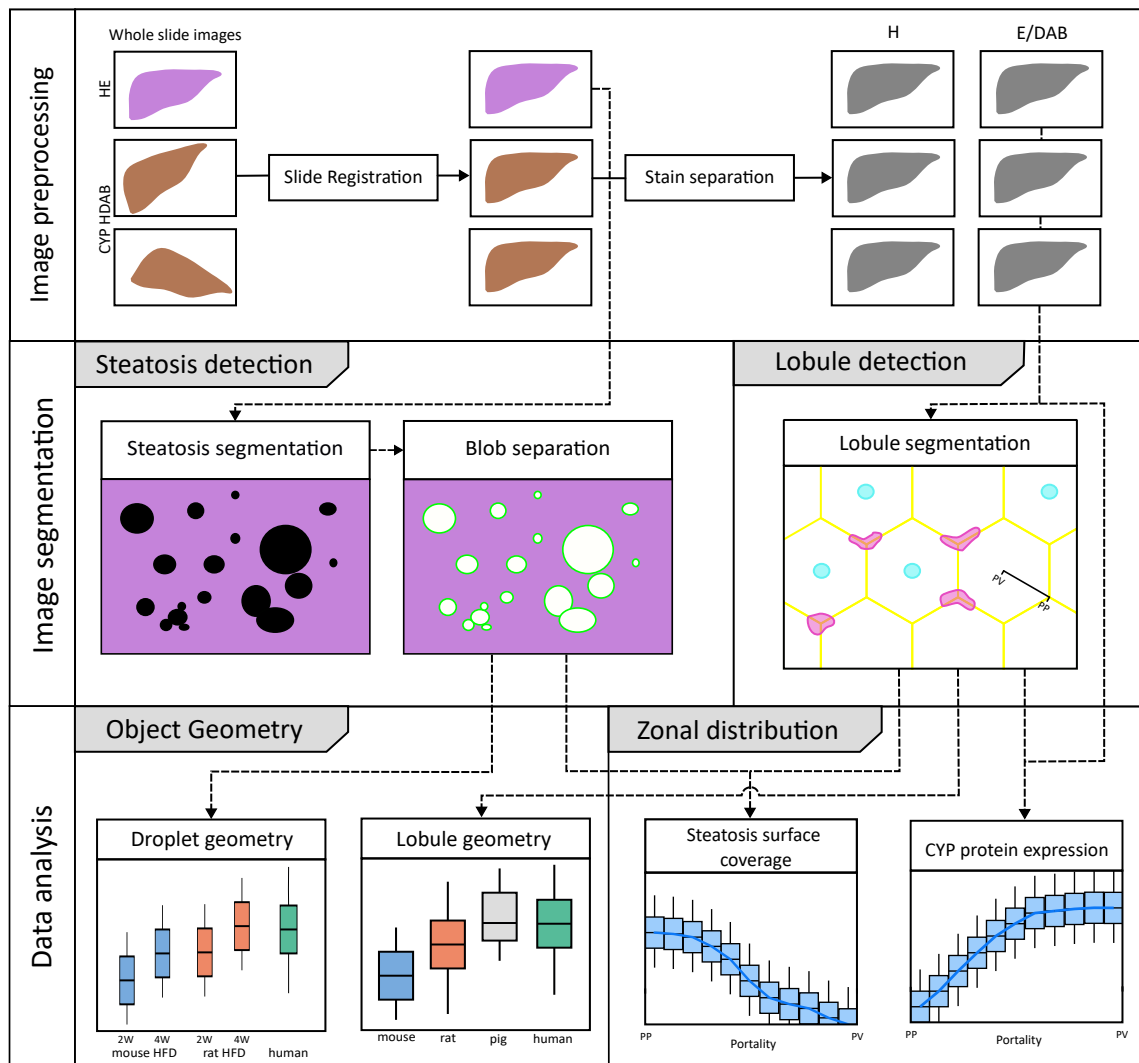
Non-alcoholic fatty liver disease (NAFLD) is one of the most common liver diseases characterized by the accumulation of fat in the liver (steatosis) in the absence of significant alcohol consumption. Histopathologically, NAFLD is characterized by the presence of microvesicular fat droplets within hepatocytes.

Although liver zonation has been extensively studied, the protein gradients of CYPs and their alterations in response to steatosis have not been systematically quantified or compared across species. This study aimed to develop an image analysis workflow to accurately derive these lobular expression gradients and quantify macrosteatosis in whole slide images (WSI) of liver histology.

A comprehensive analysis was performed on whole slide images (WSIs) of mouse, rat, pig, and human liver specimens. These images included hematoxylin and eosin (H&E) stained slides and immunohistochemically stained slides for glutamine synthetase (GS) and cytochrome P450 isoforms 1A2, 2E1, 2D6, and 3A4. To achieve the study objectives, two image analysis workflows were developed: one using classical methods and the other using convolutional neural networks (CNN). These workflows were designed to perform two key tasks: (1) segmentation of liver lobules and (2) detection of macrosteatosis droplets.

Key findings were that there was significant variability in lobular geometry within subjects, with strong consistency across species, with inter-species differences observed primarily in lobular size. A slight increase in lobular size was observed in steatotic samples, with a moderate to strong correlation between lobular size and steatosis content. In addition, interspecies differences in zonation patterns were observed for CYP1A2 in pigs, CYP2E1 in rats, and CYP3A4 in humans. Importantly, zonation patterns in steatotic subjects were not qualitatively different from those in non-steatotic subjects across species.

In summary, I have successfully developed and implemented an image analysis workflow for whole slide images (WSIs) that allows detailed examination of liver lobule geometry and macrosteatosis droplets. Furthermore, the resulting datasets were integrated to facilitate a comprehensive analysis of the zonal distribution of CYP protein expression and macrosteatosis within the lobules.



**Graphical abstract of the image analysis workflow presented in this thesis:** The workflow begins with a set of whole slide images (WSIs), including a hematoxylin and eosin (H&E) stained WSI and immunohistochemically stained WSIs for glutamine synthetase (GS) and cytochrome P450 isoforms 1A2, 2E1, 2D6, and 3A4. During preprocessing, these images were aligned, and the hematoxylin and eosin/DAB staining components were separated to produce grayscale images representing each staining component. The DAB components from the GS and CYP images were then processed through a lobule detection pipeline to identify lobule and vessel boundaries. Meanwhile, the registered H&E slide was processed through a steatosis detection pipeline to identify macrosteatosis droplets. The resulting datasets from both pipelines were then analyzed to assess (1) the geometric properties of lobule boundaries and macrosteatosis droplets, and (2) the zonal distribution of protein expression and macrosteatosis.

## Zusammenfassung

Die Leber von Säugetieren ist in dreidimensionale Strukturen, sogenannte Lobuli, organisiert, die für ihre Funktion von entscheidender Bedeutung sind. Gemäß dem weit verbreiteten hexagonalen Modell ist der Querschnitt jedes Lobulus durch eine zentrale Vene und einem Verbund aus einer Pfortader, einer Leberarteriole und einem Gallengang an den Ecken gekennzeichnet. Das Blut aus der periportal Region durch Sinusoide, die mit Hepatozyten ausgekleidet sind, in Richtung der zentralen Vene. Diese Hepatozyten exprimieren eine Vielzahl von Stoffwechsellzymen, darunter Cytochrom P450 (CYP)-Isoformen, die für den Abbau von Xenobiotika entscheidend sind. Bemerkenswerterweise zeigen viele Leberenzyme, einschließlich der CYPs, eine inhomogene Expression entlang der periportal-perivenösen Achse, eine räumliche Variation, die als Leberzonierung bekannt ist.

Nicht-alkoholische Fettlebererkrankung (NAFLD) ist eine der häufigsten Lebererkrankungen, die durch die Ansammlung von Fett in der Leber (Steatose) in Abwesenheit eines signifikanten Alkoholkonsums gekennzeichnet ist. Histopathologisch ist NAFLD durch das Vorhandensein von mikrovesikulären Fetttropfen innerhalb der Hepatozyten charakterisiert.

Trotz intensiver Forschung zur Leberzonierung wurden die Protein-Gradienten der CYPs und ihre Veränderungen bei Steatose bisher nicht systematisch quantifiziert oder zwischen Spezies verglichen. Diese Studie entwickelte einen Bildanalyse-Workflow, um diese lobulären Expressionsgradienten präzise zu erfassen und Makrosteatose in Ganzschnittbildern (WSI) der Leberhistologie zu quantifizieren.

Eine umfassende Analyse wurde an Whole-Slide-Images (WSI) von Lebergewebe-proben von Maus, Ratte, Schwein und Mensch durchgeführt. Diese Bilder umfassten Hämatoxylin- und Eosin-(H&E)-gefärbte Schnitte sowie immunhistochemisch gefärbte Schnitte für Glutaminsynthetase (GS) und Cytochrom P450-Isoformen 1A2, 2E1, 2D6 und 3A4. Um die Studienziele zu erreichen, wurden zwei Bildanalyse-Workflows entwickelt: einer unter Verwendung klassischer Methoden und der andere unter Verwendung von Convolutional Neural Networks (CNN). Diese Workflows wurden entwickelt, um zwei Hauptaufgaben zu erfüllen: (1) Segmentierung von Leberlobuli und (2) Erkennung von Makrosteatose-Tropfen.

Die Ergebnisse zeigen signifikante Variabilitäten in der Lobulusgeometrie zwischen den Probanden, konsistente Muster zwischen den Spezies und eine Zunahme der Lobulusgröße bei steatotischen Proben. Interspezifische Unterschiede in den Zonierungsmustern wurden für CYP1A2 bei Schweinen, CYP2E1 bei Ratten und CYP3A4 beim Menschen festgestellt. Interessanterweise unterschieden sich die Zonierungsmuster bei steatotischen Proben nicht qualitativ von denen bei nicht-steatotischen.

Zusammenfassend habe ich erfolgreich einen Bildanalyse-Workflow für Whole-Slide-Images (WSI) entwickelt und implementiert, der eine detaillierte Untersuchung der Geometrie der Leberlobuli und von Makrosteatose-Tropfen ermöglicht. Darüber hinaus wurden die resultierenden Datensätze integriert, um eine umfassende Analyse der zonalen Verteilung der CYP-Proteinausdrücke und der Makrosteatose innerhalb der Lobuli zu ermöglichen.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	The Liver . . . . .	1
1.2	Hepatic Zonation . . . . .	1
1.3	Hepatic Steatosis . . . . .	4
1.4	Questions, Scope, and Hypotheses . . . . .	8
<b>2</b>	<b>Materials and Methods</b>	<b>9</b>
2.1	Human and Animal Samples . . . . .	9
2.1.1	Rodent Samples . . . . .	9
2.1.2	Pig Samples . . . . .	9
2.1.3	Human Samples . . . . .	10
2.2	Ethics Statement . . . . .	10
2.3	Histology . . . . .	10
2.4	Whole slide image (WSI) preprocessing . . . . .	11
2.4.1	WSI Registration . . . . .	11
2.4.2	Stain Separation . . . . .	12
2.5	Detection of Macrosteatosis . . . . .	13
2.5.1	U-Net Model Framework . . . . .	13
2.5.2	Training Dataset . . . . .	13
2.5.3	Preparation of the Training Data . . . . .	13
2.5.4	Training of the U-Net Model . . . . .	14
2.5.5	Semantic Segmentation on WSI . . . . .	15
2.5.6	Postprocessing and Clump Separation . . . . .	15
2.5.7	Model Validation . . . . .	16
2.6	Statistical Analysis . . . . .	18
2.7	Code Dependencies . . . . .	18
2.8	Code Availability . . . . .	19
<b>3</b>	<b>Results</b>	<b>20</b>
3.1	Image Analysis Workflow for Lobule Segmentation . . . . .	21
3.1.1	Lobule Detection Workflow . . . . .	21
3.1.2	Protein Portality Map . . . . .	26
3.2	Quantification of macrosteatosis droplets in whole-slide images (WSI) . . . . .	28
3.2.1	Steatosis Segmentation Performance . . . . .	28

3.2.2	Quantification of Macrosteatosis . . . . .	29
3.3	Lobular Geometry . . . . .	32
3.3.1	Species-Specific Lobular Geometry . . . . .	32
3.3.2	Impact of Macrosteatosis on Lobular Geometry . . . . .	37
3.4	Hepatic Zonation . . . . .	40
3.4.1	Quantification of Zonated Protein Expression . . . . .	40
3.4.2	Effects of Macrosteatosis on Zonated Protein Expression . . . . .	42
3.4.3	Zonated Distribution of Macrosteatosis . . . . .	46
<b>4</b>	<b>Discussion</b>	<b>49</b>
4.1	Image Analysis Workflow for Lobule Segmentation . . . . .	49
4.2	Quantification of Macrosteatosis . . . . .	50
4.3	Lobular Geometry . . . . .	51
4.4	Zonated Expression of Cytochrome P450 Isoforms . . . . .	52
<b>5</b>	<b>Outlook</b>	<b>53</b>
	<b>References</b>	<b>56</b>
<b>S</b>	<b>Supplementary material</b>	<b>63</b>
S.1	Tables . . . . .	63



# 1 Introduction

## 1.1 The Liver

The liver is a vital organ that serves as a central hub for a wide range of essential physiological processes. It plays a critical role in macronutrient metabolism, blood volume regulation, immune system support, endocrine control of growth signaling pathways, lipid and cholesterol homeostasis, and detoxification of xenobiotic compounds, including many pharmaceutical drugs. Among its key functions, the liver's ability to process, distribute, and metabolize macronutrients is fundamental because it provides the energy necessary to sustain these physiological activities [11].

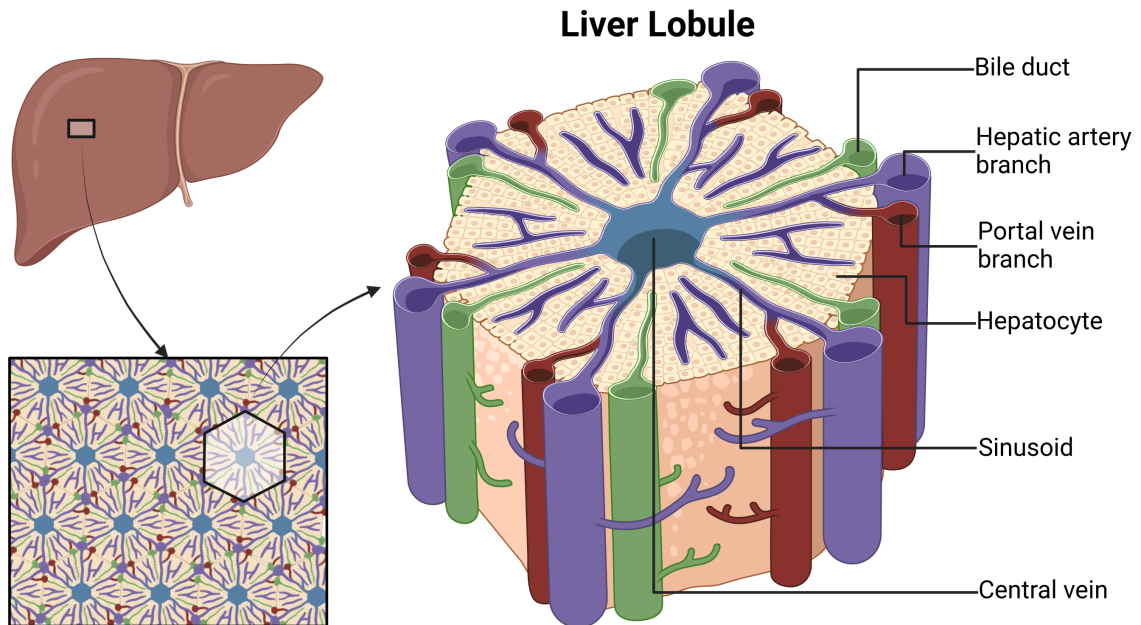
One of the primary functions of the liver is to regulate glucose levels. It stores glucose as glycogen during periods of feeding and produces glucose through gluconeogenesis in response to fasting [43, 33]. In addition, the liver is central to lipid metabolism, not only oxidizing lipids for energy but also packaging excess lipids for secretion and storage in other tissues, such as adipose tissue [3, 40].

The liver also plays a critical role in protein and amino acid metabolism. It is responsible for synthesizing a significant portion of the proteins found in the blood, both in terms of total mass and protein diversity. In addition, the liver processes amino acids for energy production and manages the disposal of nitrogenous waste from protein degradation through the urea cycle [60].

## 1.2 Hepatic Zonation

The liver is a highly structured organ composed of three-dimensional anatomical units called liver lobules. These lobules are traditionally described by a hexagonal model (Figure 1.1). In this model, each lobule features a central vein at its core, with branches of the hepatic artery, portal vein, and bile duct situated at the corners, forming what is known as the portal triad. Oxygen-rich blood enters the liver lobule through the portal tracts and flows inward toward the central vein, while bile produced by hepatocytes flows outward toward the bile ducts in the portal triads [18]. Hepatocytes, the primary functional cells of the liver, are arranged in layers along the portal-venous axis, absorbing oxygen and nutrients as blood passes through. This directional blood flow creates a gradient of oxygen, nutrients, and signaling molecules, resulting in a zoned microenvironment that drives the heterogeneous expression of various enzymes and metabolic pathways within the lobule. As a result, metabolic functions are not uniformly distributed: energetically demanding

processes such as protein synthesis, urea production, and gluconeogenesis predominantly occur in the oxygen-rich periportal zones, while processes like xenobiotic metabolism, bile acid biosynthesis, and glycolysis, which are less dependent on high oxygen levels, are concentrated in the pericentral zones [49]. This spatial partitioning of metabolic functions within the liver is known as liver zonation, and it is essential for the organ’s ability to efficiently manage its diverse roles [41, 10].



**Figure 1.1: The hexagonal model of a liver lobule.** Created with Biorender.com.

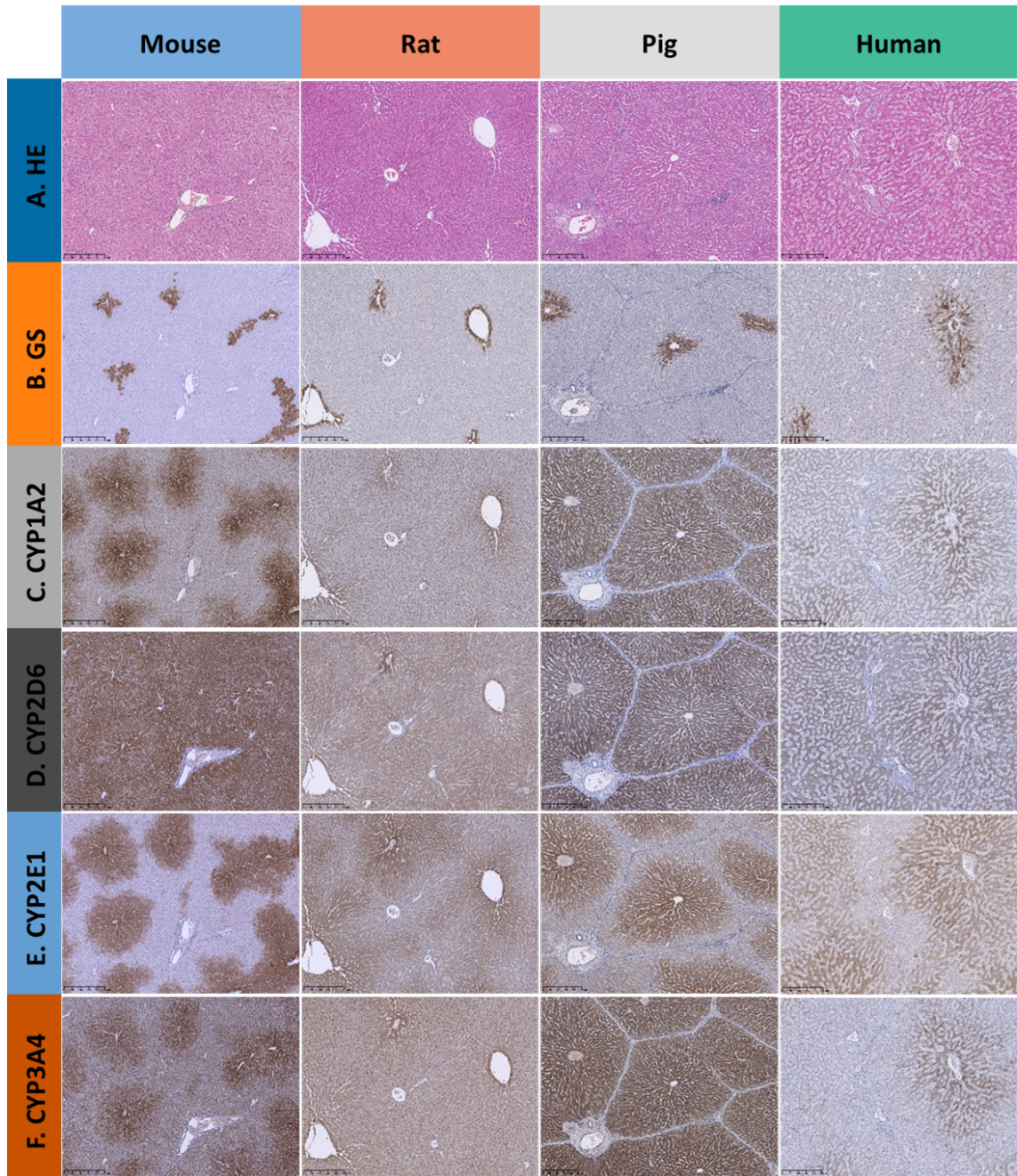
Liver zonation provides insight into the localized drug-induced damage observed in liver pathology. For example, acetaminophen is the leading cause of drug-induced acute liver failure in the US, UK, South America, Australia, and Sweden [20]. This damage is primarily due to the accumulation of reactive metabolites produced by the metabolizing enzymes CYP2E1 and CYP1A2, which are predominantly expressed in the pericentral zones resulting in pericentral necrosis [36].

The cytochrome P450 (CYP) superfamily is expressed primarily in the liver and is responsible for catalyzing over 90% of all reported enzymatic reactions [58]. The CYP isoforms 3A4 (20%), 1A2 (10%), 2D6 (10%), and 2E1 (5%) collectively contribute to about 45% of these reactions. In humans, approximately 80% of oxidative metabolism and 50% of overall drug elimination can be attributed to CYP enzymes [76].

It has long been established that most members of the CYP superfamily are preferentially expressed in the periportal zone of liver lobules in rats [31, 9, 13]. Notably, Bühler et al. [13] analyzed the expression of CYP isoforms 2A1, 2B, 2E1, and 3A in immunochemically stained adjacent liver sections and found nearly identical expression patterns for these proteins. However, this research was conducted exclusively in rats [54], and little is known about the quantitative expression patterns within liver lobules or their potential differences between species.

Example sections of immunostained WSI for GS and the CYPs 1A2, 2D6, 2E1, 3A4 are

shown in Figure 1.2 for normal mouse, rat, pig, and human samples.



**Figure 1.2: Overview of HE and zoned expression of GS and CYP in liver tissue in normal mice, rats, pigs, and humans.** The different stainings are depicted in rows, with columns corresponding to the different species. Adapted from Albadry et al. [1].

A thorough analysis of CYP expression patterns requires a method to automatically detect liver lobules in histological whole slide images (WSI) in order to establish the spatial portal-venous reference system across which protein expression can be analyzed. Identifying liver lobules in WSI is a challenging task and existing methods require manually annotated image data. Peleman et al. [56] employed a semi-automated approach by manually annotating central veins on tissue slides and then using Voronoi tessellation to segment the lobules. Similarly, Lau et al. [46] used a fully automated approach in which slides were stained

for glutamine synthetase (GS), a protein expressed exclusively in the hepatocyte layers adjacent to the central veins. Markers were created from the GS signal to calculate the Voronoi diagram. However, the borders obtained from tessellation ignore the location of the periportal vessels. This problem was circumvented by manual adjustments. Another approach involved manual annotation of the periportal vessels to assign each pixel a "portality" value ranging between 0 and 1, calculated based on the distance of the pixel to the nearest central and portal locations. Lobule boundaries were then derived using a watershed algorithm with central veins locations serving as markers [66].

However, these methods rely heavily on manual annotation to produce accurate results. For a comprehensive and scalable analysis, this reliance on manual input is not feasible due to the significant time and expertise required to effectively annotate large datasets.

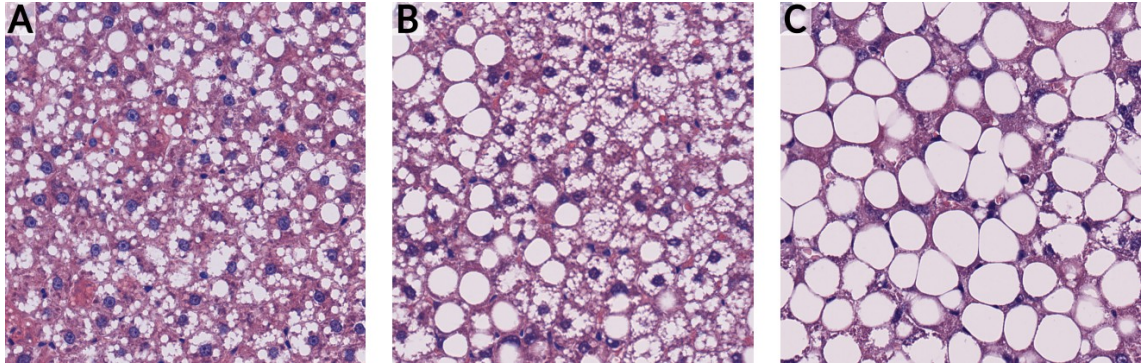
### 1.3 Hepatic Steatosis

Hepatic steatosis, also known as fatty liver disease, is characterized by the accumulation of fat, predominantly triglycerides, within hepatocytes [6]. It represents the mildest form on the spectrum of liver disease, which can range from simple fatty liver to more severe conditions such as steatohepatitis and can eventually progress to fibrosis, cirrhosis, and hepatocellular carcinoma [57]. There are two major forms of steatosis: alcoholic and non-alcoholic fatty liver disease (NAFLD). NAFLD has a global prevalence of 25% and is primarily observed in patients with central obesity, Type 2 diabetes mellitus, dyslipidemia, and metabolic syndrome [57]. The histological damage in NAFLD closely resembles that seen in alcoholic liver disease; however, by definition, NAFLD is not induced by alcohol [6]. Experts have reached a consensus that the term NAFLD no longer adequately reflects current scientific understanding, and they have proposed the use of metabolic (dysfunction)-associated fatty liver disease (MAFLD) as a more accurate and encompassing term [23, 28].

Several non-invasive imaging techniques are available for the diagnosis of liver steatosis, including ultrasound [24], vibration-controlled transient elastography (VCTE) [70], CT, MRI, and magnetic resonance spectroscopy (MRS) [69]. Despite these advances, liver biopsy remains the gold standard for the diagnosis of hepatic steatosis and is particularly important for assessing its severity [57]. Accurate quantification of hepatic steatosis is also critical in assessing the suitability of donor livers for transplantation, as donors with macrovesicular steatosis in more than 25% of hepatocytes have a higher risk of early graft dysfunction in liver transplant recipients [4].

However, the absence of standardized grading criteria for steatosis has led to highly variable scores and inconsistent assessments of liver transplant suitability among pathologists [53]. To address this issue, Neil et al. [53] proposed a set of standards to guide the assessment of hepatic steatosis in liver histology images. They categorized fat in histological slides into three distinct types: **Large droplet fat (LDF)**: Characterized by a single large fat droplet within the cell that causes significant cellular expansion, making the affected hepatocyte larger than adjacent non-steatotic or minimally steatotic hepatocytes. If the nucleus is visible in the section, it is typically displaced to the cell's periphery (Figure

1.3C). **Smaller droplet fat (SDF)**: Comprising all fat droplets that do not qualify as LDF or 'true' microvesicular steatosis (Figure 1.3B). **"True" microvesicular steatosis**: Tiny droplets within hepatocytes, which expand and fill the cells, resulting in a foamy appearance (Figure 1.3A, B).



**Figure 1.3: Histologic images of the steatosis types** (A) True microvesicular steatosis and Smaller Droplet Fat (SDF). (B) Large Droplet Fat (LDF) in the left and microvascular steatosis in the right portion of the image. (C) Mainly LDF. All examples are  $225\ \mu\text{m} \times 225\ \mu\text{m}$  sections from a 40x WSI of rat liver samples.

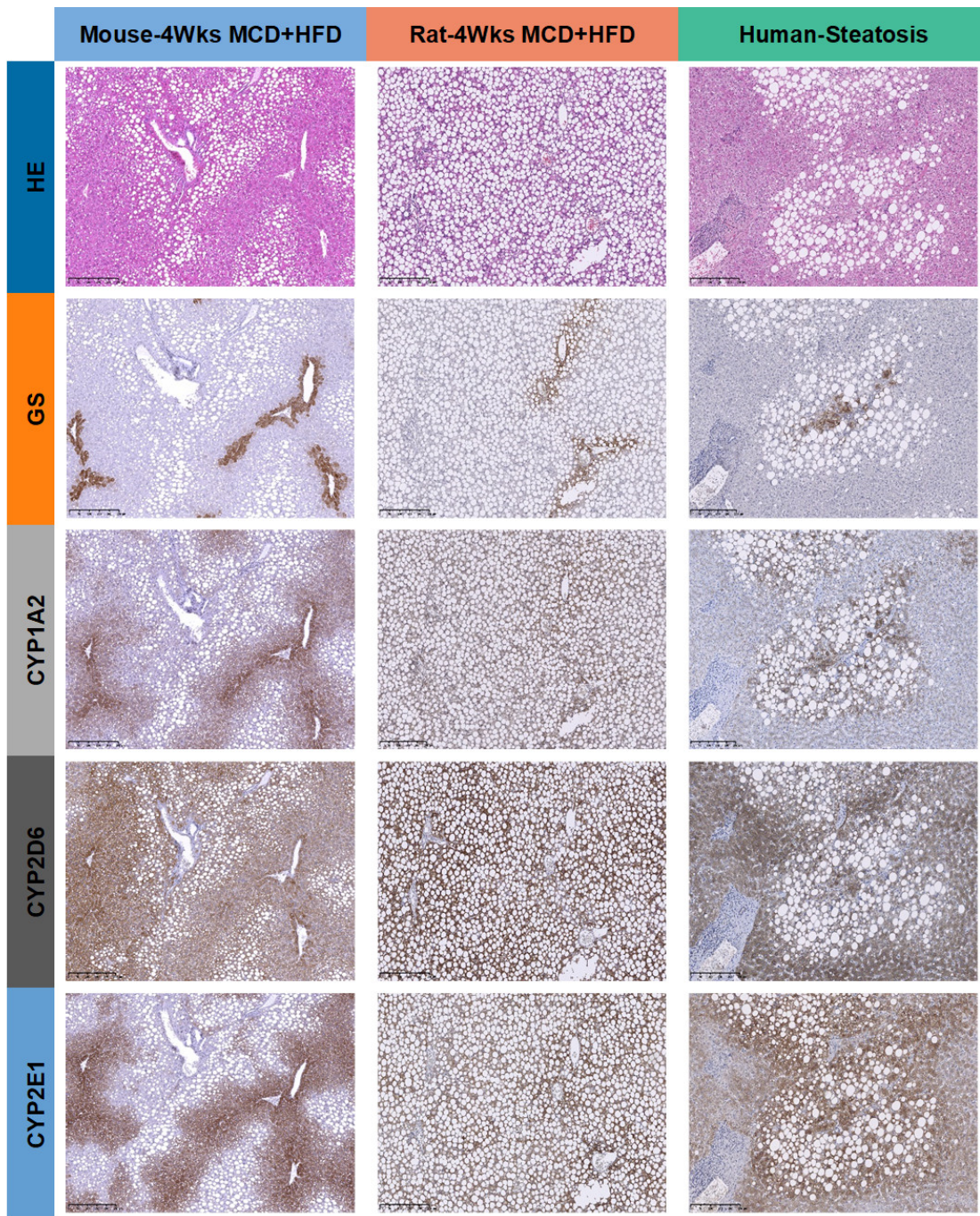
Automated detection of steatosis in whole slide images (WSI) is a significant challenge in histological image analysis. Various analytical techniques have been employed to address this problem, ranging from classical rule-based methods to fully deep learning-based approaches. Classical image analysis techniques often exploit the characteristic appearance of steatosis blobs, which present as white, circular objects on Hematoxylin and Eosin (H&E) stained WSIs [12]. In these methods, the initial step involves extracting potential objects by thresholding the input image. The extracted objects are then classified based on their geometric properties, such as size and shape [37, 62, 52].

Roy et al. [62] used an unsupervised approach to distinguish between steatosis droplets and clumped droplets. Potential objects were filtered based on size and solidity, a measure of how much an object deviates from its convex hull. For clumped objects, points of high curvature along the boundary were identified, and adjacent points were connected to split the object into two. This separation was validated by fitting ellipses to both the separated objects and the original object. In contrast, Homeyer et al. [37] used a supervised approach, where shape features of the candidate objects were used to train a random forest classifier to distinguish between steatosis and non-steatosis objects. Munsterman et al. [52] used a logistic regression model for classification. After filtering objects by size, the regression model classified blobs as steatosis or non-steatosis based on geometric parameters.

In recent years, neural networks have become increasingly popular for the quantification of steatosis. Salvi et al. [64] used a hybrid approach that combined a UNET architecture with a pre-trained ResNet34 backbone, along with several classical image processing techniques. They first applied Gabor filters to the input images, followed by thresholding to identify potential steatosis objects. The semantic segmentation results from the UNET were then used to eliminate false positives generated during the initial thresholding. In addition, an active deformable model was employed to separate fused droplets. Roy et al. [63]

implemented a methodology that integrated three different neural network types: a UNET structure to identify steatosis regions, a Holistically Nested Neural Network (HNN) to analyze the characteristics of steatosis boundaries, and a Fully Convolutional Network (FCN) to generate the final prediction map.

Steatosis has been shown to significantly influence transcription and protein expression in the liver, including drug-detoxifying enzymes such as CYPs [16]. Research indicates that CYP3A activity is elevated in humans, rats, and mice in steatosis [16]. Similarly, CYP2D6 levels and activity tend to increase in human liver tissue affected by nonalcoholic steatohepatitis (NASH). In contrast, CYP1A mRNA and protein expression are reduced in rodent models of nonalcoholic fatty liver disease (NAFLD) and corresponding human tissues. Conversely, CYP2E1 levels and activity are consistently elevated in both humans and rodents. These findings were further corroborated by Albadry et al. [2], who reported decreased enzymatic activity for CYP1A and CYP3A, alongside increased activity for CYP2E1, in steatotic mice. Example sections of the extent of steatosis and protein expression for the investigated species and proteins are presented in Figure 1.4.



**Figure 1.4: Overview of H&E, extent of zoned expression and steatosis in liver tissue in mice (control, 2W HFD, 4W HFD), rats (control, 2W HFD, 4W HFD), and humans.** The different stainings are depicted in rows and the species are in the columns. Species are color-coded as follows: blue for mice, orange for rats, and green for humans. Scale bars are 250  $\mu$ m. Adapted with permission of Mohamed Albadry (unpublished results).

## 1.4 Questions, Scope, and Hypotheses

It is well established that CYPs are expressed in the liver lobule in a zoned pattern, with most CYPs depicting high expression in the pericentral zone and lower expression in the periportal zone. While the geometry of liver lobules has been the subject of some studies, systematic quantification and comparison of lobular geometry and CYP zonation patterns across species have not yet been conducted. Also, the effect of steatosis on the zonation patterns and geometries has not been compared between species so far.

This thesis aims to develop and implement an image analysis workflow for whole slide images (WSIs) that allows detailed examination of liver lobule geometry and macrosteatosis droplets to address the following questions:

- Are there differences in lobular geometry between species?
- Are there inter-species differences in zoned expression patterns of the cytochrome P450 isoforms?
- Does steatosis affect lobular geometry and zoned expression patterns?



## 2 Materials and Methods

### 2.1 Human and Animal Samples

#### 2.1.1 Rodent Samples

Formalin-fixed and paraffin-embedded liver samples were collected from normal and steatotic mice, rats, pigs, and humans, with six samples per group. For mice, liver samples were taken from the left lateral lobe (LLL), median lobe (ML), right lobe (RL), and caudate lobe (CL). For rats, only the median lobe (ML) was sampled. Pig liver samples were collected from the LLL. Human liver samples were obtained during clinically indicated hepatic surgeries performed in 2019 at the University Hospital of Jena, Germany.

For the rodent studies, male C57BL6/J mice (ex-breeders) from Janvier, France, weighing 28–30 g and aged 8–9 months (n=6/group), and male Lewis rats from Charles River, France, weighing 300–400 g and aged 3 months (n=6/group) were used. The rodents were housed in groups of three with *ad libitum* access to food and water, under controlled environmental conditions: a 12-hour light/dark cycle, a temperature of  $21 \pm 2^\circ\text{C}$ , and 45%–65% relative humidity. Animals were humanely euthanized using an overdose of isoflurane followed by exsanguination.

To induce steatosis in rodents, a high-fat diet with low methionine and choline content (HF-diet) (E15652-94 EF R/M, high-fat MCD mod, Ssniff Spezialdiäten GmbH, Sulzfeld, Germany) was administered for two (2W HFD) or four weeks (4W HFD), leading to varying degrees of steatosis. Control groups were maintained on a standard diet from Altromin Spezialfutter GmbH, Germany. Body weight and food intake were monitored daily throughout the study.

#### 2.1.2 Pig Samples

Liver tissue from pigs was collected from three male and three female clinically healthy Prestice Black-Pied pigs, weighing 25–33 kg and aged 3 months. These pigs were housed under similar controlled conditions with a 12-hour light/dark cycle, a temperature of  $21 \pm 2^\circ\text{C}$ , and 60% relative humidity. Anesthesia was induced with ketamine (10 mg/kg), azaperone (5 mg/kg), and atropine (1 mg), and maintained with intravenous propofol (5–10 mg/kg). Pigs were euthanized with an intravenous injection of T61 solution under general anesthesia.

### 2.1.3 Human Samples

Human liver samples were obtained from 12 individuals, including normal (n=6, aged 45–59 years, 2 males, 4 females) and steatotic tissue (n=6, aged 45–69 years, 5 males, 1 female, steatosis grade 30-70%) samples. Given that most patients undergoing liver surgery present with additional conditions such as steatosis, fibrosis, or hepatic inflammation, careful selection of samples was necessary. Samples were independently assessed by four scientists, including a board-certified pathologist, to ensure accurate categorization. The absence of abnormalities, including steatosis and fibrosis, was confirmed by a board-certified hepato-pathologist. Steatosis was considered absent if less than 5% of hepatocytes contained lipid droplets, following established guidelines [42, 22]. Fibrosis was deemed absent when liver tissue samples scored 0 for necrosis, inflammation, and fibrosis according to the Ishak score (modified Knodell score) [39] based on Hematoxylin and Eosin (HE) and Elastin Van Gieson (EvG) stained sections.

## 2.2 Ethics Statement

All animal experiments and housing were conducted in compliance with current with the German Law on the Protection of Animals/European Communities Council Directive (Directive 2010/63/EU), animal welfare guidelines, and the ARRIVE Guidelines for Reporting Animal Research. The mouse study was approved by the Thüringer Landesamt für Verbraucherschutz, Thuringia (Approval Number: UKJ-19-020) (see also Albadry et al. [2, 1]), and the rat study was approved under Reg.-Nr. 02-043/10 and Reg.-Nr.02-018/14.

The pig study received approval from the Commission for Work with Experimental Animals under the Czech Republic’s Ministry of Agriculture, project ID: MSMT-15629/2020-4.

The human study was approved by the institutional review board of the University Hospital of Jena, Germany (ethical vote: UKJ\_2018-1246-Material). This approval confirms that the study adhered to the ethical standards outlined in the 1964 Declaration of Helsinki and its subsequent amendments or comparable ethical standards.

## 2.3 Histology

Histological experiments were conducted as described in Albadry et al. [2, 1]: Paraffin-embedded liver tissue samples were processed for H&E staining by first cutting 3  $\mu$ m sections to evaluate lobular architecture. The staining process began with deparaffinization and rehydration through descending grades of alcohol. The sections were then immersed in hematoxylin and eosin solutions, followed by dehydration using ascending grades of alcohol and subsequent clearing with xylene. Additionally, paraffin-embedded human liver tissue samples were subjected to EvG staining to confirm the absence of fibrous connective tissue. After cutting 3  $\mu$ m sections, the tissue sections were deparaffinized and hydrated in distilled water. The sections were immersed in Verhoeff’s solution for 1 hour until the tissue appeared completely black. Following this, the sections were rinsed in tap water

and differentiated with 2% ferric chloride for 1–2 minutes. The slides were then washed with tap water, treated with 5% sodium thiosulfate for 1 minute, and counterstained with Van Gieson’s solution for 3–5 minutes. The sections were subsequently dehydrated using ascending grades of alcohol and cleared with xylene.

Immunohistochemistry was employed to assess and quantify the spatial distribution of glutamine synthetase (GS) and four different cytochrome P450 (CYP) enzymes, as previously outlined Albadry et al. [2]. The staining was conducted on consecutive 3  $\mu\text{m}$  thick sections of formalin-fixed, paraffin-embedded liver tissue. Five different antibodies were used to detect GS, CYP1A2, CYP2D6, CYP2E1, and CYP3A4. After deparaffinization and rehydration using descending grades of ethanol, heat-induced epitope retrieval was performed using a trisodium citrate buffer at pH 6.1, with steaming at 100 °C for 30 minutes. The samples were then allowed to cool to room temperature for 20 minutes. Endogenous peroxidase activity was inhibited using a 3% hydrogen peroxide solution. To block non-specific binding, a commercially available protein block (ab64226, Abcam, Germany) was applied. The sections were incubated overnight at 4 °C with the appropriate CYP antibody. For the rabbit polyclonal primary antibodies (CYP2D6, 2E1, and 3A4), the rabbit-specific HRP/DAB IHC detection system (ab236469, Abcam) was used for 40 minutes at room temperature. For mouse monoclonal primary antibodies (GS and CYP1A2), additional primary antibody biotinylation was conducted using the Dako Animal Research Kit Peroxidase (K3954, Dako, Denmark). Endogenous avidin and biotin activity was blocked using the avidin/biotin blocking kit (ab64212, Abcam), followed by application of the avidin-HRP complex. To visualize the reaction, DAB chromogen (GV825, Dako, Denmark) was applied for approximately 3 minutes. Counterstaining was carried out with Dako hematoxylin (CS700, Dako, Denmark) for 6–8 minutes. For each staining procedure, a single slide was used as a negative reagent control, where the primary antibody was omitted. A consistent protocol was used for staining the liver tissue samples across the four species. The stained slides were mounted and digitized using a whole slide scanner (L11600, Hamamatsu, Japan) equipped with NDP.view2 Plus Image viewing software (version U12388-02) at  $\times 40$  magnification.

## 2.4 Whole slide image (WSI) preprocessing

### 2.4.1 WSI Registration

To aggregate protein expression data across slides, aligning adjacent regions of interest (ROIs) was necessary. I utilized the VALIS Python library [25] for this purpose. VALIS identifies image features and performs both rigid and non-rigid registrations to effectively align slides. Initially, rigid registration aligns slides based on overall features, while non-rigid registration refines the alignment using micro features at higher resolution levels.

VALIS requires each image set to contain a single ROI for reliable alignment. Therefore, I extracted ROIs from the target images. When multiple ROIs were present in the source image, I used QPath [8] to annotate these ROIs. In the lowest resolution of the source image, I applied a static threshold to segment the tissue and employed OpenCV contour detection

to outline the tissue contours. Each ROI annotation from QPath was then matched against these contours. When a match was found, I extracted the corresponding subsection of the source image defined by the contour’s bounding box and saved it as a separate `ome.tiff` file. This process produced one or more ROI subdirectories per subject, each containing the ROI for every stain.

For each subject and ROI, slides were processed using VALIS, which organized the aligned slides into a consistent directory structure. This ensured that each subject and ROI had a set of aligned slides for downstream analysis. A graphical example is shown in Figure 3.2A.

## 2.4.2 Stain Separation

For segmentation based on protein expression and quantitative analysis of zoned expression, crucial information is extracted from the DAB signal in whole slide image (WSI) scans. A Python implementation [65] of the stain separation technique proposed by Macenko et al. [47] to separate the Hematoxylin and DAB staining components was used. The method begins by transforming RGB tuples into optical density (OD) space. Instead of using singular value decomposition, this implementation applies the eigendecomposition of the covariance matrix. The two eigenvectors corresponding to the largest eigenvalues define a plane onto which OD-transformed pixels are projected.

To identify the extreme vectors, angles are computed for each projected vector, and the minimum and maximum angles are determined. To enhance robustness against outliers, the 0.01 and 0.99 percentiles are used to represent the minimum and maximum angles, respectively. Stain vectors are then derived by calculating the dot product between these eigenvectors and the vectors corresponding to the extreme angles.

Stain concentrations are calculated by solving the equation  $A \cdot c = P$ , where  $A$  is the stain matrix composed of stain vectors,  $P$  represents the OD-transformed pixels, and  $c$  denotes the concentration vector. Given the large size of megapixel images, which cannot be loaded into memory all at once, a tile-based approach is employed for this algorithm. As suggested by Anghel et al. [5], the algorithm operates exclusively on foreground pixels. The implemented procedure consists of three key steps: (1) Identifying foreground pixels; (2) Calculating stain vectors; (3) Decomposing each image tile and writing the results to a target array.

In the first step, foreground pixels are identified by randomly sampling pixels from each tile and using Otsu’s method [55], implemented in OpenCV [17], to determine a threshold. In the second step, stain vectors are computed from the selected foreground pixels based on the threshold. Finally, the tiles are decomposed using these stain vectors, and the resulting grayscale images for each stain are stored in a target ZARR array with JPEG compression. To manage computational demands efficiently, only 0.1 percent of the pixels are used for calculating the threshold and the stain vectors.

One significant drawback of this approach is the need to load each tile into memory three times. To reduce the time spent on decompression, the uncompressed tile is written to a temporary file on disk after the first read during threshold computation. This allows

subsequent accesses to bypass decompression, thereby streamlining the computation of stain vectors and tile decomposition.

An example result of the stain separation algorithm is shown in Figure 3.2B.

## 2.5 Detection of Macrosteatosis

A hybrid approach combining classical image analysis methods with neural network technology was developed to detect macrosteatosis droplets. The first part of the detection pipeline utilized a U-Net model to segment the input image into steatosis and non-steatosis regions. The regions identified as steatosis were then further processed using a blob disentanglement algorithm.

### 2.5.1 U-Net Model Framework

A Python library published by Iakubovskii [38], based on the TensorFlow framework [50], which provides a U-Net model [61] with pre-trained backbones, was used. Specifically, a U-Net model with a VGG16 encoder, trained on the ImageNet dataset [19], was employed in this work.

### 2.5.2 Training Dataset

QPath was utilized to select points in the WSI from which 1024x1024 pixel tiles were extracted from resolution-level zero (40x magnification). Macrosteatosis annotations were created for these images as described in the section 2.5.3. In total, 136 tiles were selected. Detailed information about the selection is documented in Table 2.1.

**Table 2.1:** Selection of Training Image Tiles from the WSI Dataset

	2W HDF	4HDF	Steatosis	Control
Mouse	26	24		4
Rat	24	24		4
Human			24	6

The training data was split into training, test, and validation data in an 80/10/10 ratio.

### 2.5.3 Preparation of the Training Data

To train a segmentation model, it was necessary to prepare training images along with corresponding masks that designate the class of each pixel. Manual annotation of these images is typically a labor-intensive and time-consuming task. To reduce the annotator’s workload, I implemented a strategy that generated preliminary proposals for potential steatotic blobs, allowing annotators to refine these proposals instead of annotating each image from scratch. Additionally, I simplified the annotation process by using keypoint annotations rather than polygon annotations, requiring annotators to mark just a single point within each steatotic blob.

Since steatotic droplets appear as distinct white drops on slides, potential objects were identified through image thresholding. This process generated a set of background objects, including single steatotic droplets, clumped droplets, vessels, microsteatotic areas, and other artifacts. These objects were then extracted using contour detection and filtered based on their size. To differentiate steatotic objects from other structures, a clustering approach was used based on the assumption that different classes exhibit inherent geometric differences. Geometric features were computed for each contour, and principal component analysis (PCA) was used to identify distinguishing features. After dimensionality reduction, the objects were clustered into two classes, with labels assigned through visual assessment. Steatotic objects were then annotated with polygons and keypoints and uploaded as proposals for further annotation.

Once the annotators had refined these proposals, the optimized annotations were downloaded. The training images were then processed using watershed segmentation, utilizing the keypoints as markers and the background mask as the true background. This approach resulted in an array where each steatotic object was uniquely labeled with a distinct identifier. For training the U-Net model, only a binary mask was required, so the labeled array was converted into a binary mask by setting all steatotic pixels to true and the background pixels to false.

#### 2.5.4 Training of the U-Net Model

The U-Net model with a pre-trained VGG16 encoder was first trained on the macrosteatosis training dataset published by [63] for 25 epochs with a batch size of 4. This dataset was designed for detecting macrosteatosis that fitted the definition of large droplet fat (LDF) by Neil et al. [53]. Therefore, the obtained model weights were still insufficient to detect small droplet fat (SDF). The pretrained model was then trained on 136 training images, with a batch size of 4, over 50 epochs. The images used for training were 256x256 pixel tiles (10x magnification). Consequently, the training image tiles and masks were downscaled by a factor of 4. To enhance robustness, on-the-fly augmentation techniques such as random flipping, scaling, and rotation were applied during training.

I used the binary focal dice loss for binary segmentation as recommended by the author of the segmentation-model framework [38], which is defined as the sum of dice loss  $L_{DL}$  and binary focal loss  $L_{BF}$

$$L(\mathbf{G}, \mathbf{P}) = L_{BF} + L_{DL} \quad (2.1)$$

The matrices  $\mathbf{G} \in \{0, 1\}^{m \times n}$  and  $\mathbf{P} \in (0, 1)^{m \times n}$  represent the ground truth mask and the prediction array, respectively. The binary focal loss is defined as

$$L_{BF}(\mathbf{G}, \mathbf{P}) = \frac{1}{m \cdot n} \sum_{k=1}^m \sum_{l=1}^n -g_{k,l} \cdot \alpha (1 - p_{k,l})^\gamma \log(p_{k,l}) - (1 - g_{k,l}) \alpha \cdot p_{k,l}^\gamma \log(1 - p_{k,l}) \quad (2.2)$$

with the default parameter values  $\alpha = 0.25$  and  $\gamma = 2.0$ . The dice loss is defined as

$$L_{DL}(TP, FP, FN) = \frac{(1 + \beta^2) \cdot TP}{(1 + \beta^2) \cdot TP + \beta^2 \cdot FN + FP} \quad (2.3)$$

with the default parameter value  $\beta = 1.0$ . The arguments  $TP$ ,  $FP$ , and  $FN$  denote true positive, false positive, and true negative pixels, respectively. The true positive value  $TP$  is defined as

$$TP = \sum_{k=1}^m \sum_{l=1}^n g_{k,l} \cdot p_{k,l} \quad \text{and} \quad (2.4)$$

and the false positive and false negative values  $FP$  and  $FN$  are defined as

$$FP = \left( \sum_{k=1}^m \sum_{l=1}^n p_{k,l} \right) - TP \quad \text{and} \quad FN = \left( \sum_{k=1}^m \sum_{l=1}^n g_{k,l} \right) - TP. \quad (2.5)$$

### 2.5.5 Semantic Segmentation on WSI

The U-Net model, being a fully convolutional network, is capable of handling variable input image sizes. While training requires significant computational resources, the prediction phase is less resource-intensive, allowing for the use of larger image sizes. However, due to the substantial size of whole slide images (WSI), processing them in a single pass is not feasible, making it necessary to divide the WSIs into smaller tiles. To ensure comprehensive coverage and accurate predictions, I extracted overlapping tiles from the WSI and fed them into the network. Predictions for the non-overlapping regions of each tile were then written to a target ZARR array.

The use of overlapping tiles is critical for mitigating prediction performance issues at the edges of the images, where boundary effects can lead to incomplete or inaccurate predictions. The final output is a binary array in which all pixels corresponding to steatosis are marked as true, while all other pixels are marked as false. Since the detected steatotic objects can be either individual droplets or clusters of fused droplets, further separation is necessary for subsequent analysis and accurate quantification.

### 2.5.6 Postprocessing and Clump Separation

To perform instance segmentation, the binary macrosteatosis mask was processed to identify and separate contours for each droplet instance. Initially, clumped droplets were filtered based on size and solidity thresholds. Contours with a surface area corresponding to 700 pixels ( $36 \mu\text{m}^2$ ) were discarded. Objects with a solidity greater than 0.95 were classified as singular droplets and added to the final result, while objects with a solidity smaller than 0.6 were discarded. The remaining contours were considered clumped droplets.

Solidity (SS) is defined as the ratio of the area of the object  $A_O$  to the area of its convex hull  $A_H$ , expressed as  $S = A_O/A_H$ . To further analyze clumped droplets, I utilized convexity defects, similar to the method outlined by [62]. Convexity defects represent indentations or cavities where the contour deviates from the convex hull. OpenCV functions were used to identify these defects, measuring the distance from the convex hull's farthest points to each defect.

To process the contour, I first evaluated the defects. Contours with fewer than two defects were discarded. The remaining defects were sorted by depth, and the deepest defect was identified. I then calculated the distance to the nearest second defect, selected it, and

split the contour by connecting these inflection points. The splitting was validated by fitting ellipses to the original and the two post-split objects. When the Intersection over Union of the ellipses and the respective objects improved for the split objects, the split was accepted otherwise the original object was kept. This process was repeated for the split objects until only individual droplet instances remained.

This method was adapted for use on image tiles, where contours could be inadvertently split at the tile edges, requiring post-processing to merge them. To address this, I loaded the ZARR array from semantic segmentation in tiles. For each tile, neighboring tiles (right, lower-right, and lower) were also loaded to identify and merge contours that touched tile edges or corners to ensure continuity.

Finally, attributes such as centroid coordinates, area, and circumference were computed for the resulting contours and stored in a structured dataframe for further analysis.

### 2.5.7 Model Validation

The segmentation task in this work is framed as a binary instance segmentation problem. The U-Net model semantically segments the image into two classes: background and macrosteatosis. Instance segmentation is then achieved through a blob separation process that disentangles clumped objects and assigns a unique instance label to each connected component of the macrosteatosis class.

Choosing appropriate validation metrics is critical, as they must address the specific challenges of detection, classification, and segmentation. It is essential to select metrics that are appropriate for the specific problem being solved. In this work, metrics for both semantic and instance segmentation are reported, as suggested by [48]. To identify suitable metrics for the macrosteatosis segmentation task, the online tool presented in their study was utilized.

As a metric for evaluating the U-Net semantic segmentation, the Dice Similarity Coefficient (DSC), also known as the F1-Score, is reported. The DSC is the harmonic mean of precision and recall. The precision is defined as

$$\text{Precision} = \frac{|TP|}{|TP| + |FP|} \in [0, 1] \quad (2.6)$$

where  $|TP|$  represents the number of true positive pixels, and  $|FP|$  represents the number of false positive pixels. Precision measures the fraction of pixels correctly predicted as macrosteatosis out of the total pixels classified as macrosteatosis. The recall is given by

$$\text{Recall} = \frac{|TP|}{|TP| + |FN|} \in [0, 1] \quad (2.7)$$

where  $|FN|$  denotes the number of false negative pixels. Recall measures the fraction of correctly predicted macrosteatosis pixels out of all the ground truth macrosteatosis pixels. It is also referred to as the true positive rate. The Dice Similarity Coefficient (DSC) is



calculated as

$$\text{DSC} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2|TP|}{2|TP| + |FP| + |FN|} \in [0, 1] \quad (2.8)$$

The DSC provides a single metric that balances both precision and recall, making it a comprehensive measure of segmentation performance.

For instance segmentation, the Panoptic Quality (PQ) is reported, which combines the quality of segmentation, recognition, and detection. The PQ metric integrates the average Intersection over Union (IoU) scores for all true positive instances and the Dice Similarity Coefficient (DSC) for detection quality [48]. The panoptic quality is defined as

$$PQ = \frac{\sum_{(i,j) \in TP} \text{IoU}(i,j)}{|TP| + 0.5(|FP| + |FN|)} \in [0, 1] \quad (2.9)$$

where  $\text{IoU}(i,j)$  denotes the IoU score of every matched pair of a prediction instance  $P_i$  and a ground truth instance  $G_j$ .  $|TP|$ ,  $|FP|$ , and  $|FN|$  represent the number of true positive, false positive, and false negative instances, respectively.

Additionally, [48] suggested reporting a boundary-based metric to emphasize the correct prediction of object boundaries. For this purpose, the Normalized Surface Distance (NSD) was calculated. The NSD is similar to the Dice Similarity Coefficient (DSC) but focuses solely on the object boundary regions.

The NSD measures how well the boundaries of the predicted and ground truth objects overlap. Specifically, it calculates the intersection of the boundary pixels of one object with the extended boundary of the other object, using a tolerance  $\tau$ . The metric is defined for a matched pair of a prediction  $P_i$  and ground truth instance  $G_j$  as

$$\text{NSD}(P_i, G_j)^{(\tau)} = \frac{|S_{P_i} \cap \mathcal{B}_{G_j}^\tau| + |S_{G_j} \cap \mathcal{B}_{P_i}^\tau|}{|S_{P_i}| + |S_{G_j}|} \in [0, 1] \quad (2.10)$$

where  $S$  denotes the set of boundary pixels and  $\mathcal{B}^\tau$  the set of pixels within the boundary extended by the tolerance  $\tau$ . The final metric for a test image is calculated by averaging over all TP matches:

$$\text{NSD}(P, G)^{(\tau)} = \frac{1}{|TP|} \sum_{(i,j) \in TP} \text{NSD}(P_i, G_j)^{(\tau)} \quad (2.11)$$

The tolerance parameter  $\tau$  was set to 3 px on 10x magnified images corresponding to an extension of the boundary of 1.36  $\mu\text{m}$ .

I calculated matrices  $\mathbf{C}_{\text{IoU}}$ , and  $\mathbf{C}_{\text{NSD}}$  where the element  $\mathbf{c}_{i,j}$  is the IoU, or NSD of the  $i$ th prediction element  $P_i$ , and the  $j$ th ground truth  $G_j$ , respectively.

The Hungarian matching [45] was employed to match each ground truth instance with one prediction instance in order to maximize the total IoU. The "linear\_sum\_assignment" function from the SciPy library [75] was used to solve this assignment problem. This algorithm used  $\mathbf{C}_{\text{IoU}}$  as the cost matrix to optimally match the ground truth and the prediction instances by maximizing the sum of IoU scores across all matches. The output

was a set of index pairs that represented the optimal matches.

Based on these index pairs, the number of number of false positives and false negatives were calculated. Let  $P$  and  $G$  be the set of predicted and ground truth objects and  $n$  and  $m$  be the number of predictions, and ground truth objects, respectively:

$$P = \{P_1, P_2, \dots, P_n\} \quad \text{and} \quad G = \{G_1, G_2, \dots, G_m\}. \quad (2.12)$$

Further, the set of index pairs  $(i, j)$  matched by linear sum assignment is given as

$$TP \subseteq \{1, 2, \dots, n\} \times \{1, 2, \dots, m\} \quad (2.13)$$

so that the prediction  $P_i$  is matched with the ground truth object  $G_j$ . The number of false positives  $|FP|$  is given as the number of unmatched prediction indices:

$$|FP| = n - |I_P| \quad \text{where} \quad I_P = \{i | (i, j) \in TP \text{ for some } j\} \quad (2.14)$$

Similarly, the number of false negatives  $|FN|$  is given as the number of unmatched ground truth indices

$$|FN| = m - |I_G| \quad \text{where} \quad I_G = \{j | (i, j) \in TP \text{ for some } i\}. \quad (2.15)$$

It is obvious that in the case that  $n = 0$  and  $m = 0$  (i.e. the prediction and the ground truth set are empty), the DSC, PQ, and NSD are not defined as  $|TP| = |FP| = |FN| = 0$ . In that case, the respective metric is set to 1, as this is the correct result.

## 2.6 Statistical Analysis

Data distributions were assessed by eye. When the data was assessed to be log-normally distributed, it was transformed to log space, statistics calculated in log space and back-transformed and the results were plotted on a log scale.

For group comparison tests, we used the Kruskal-Wallis tests [44]. When significant differences between groups were found, we employed a Dunns post-hoc to group-wise test the central tendency of the samples. For group-wise testing, Bonferroni correction for the p-value in multiple comparisons was used [21]. A corrected p-value  $< 0.05$  was considered significant.

## 2.7 Code Dependencies

QPath [8] was used as a viewer for whole slide images and to annotate ranges of interest. For image analysis and manipulation, the computer vision library OpenCV [17] for Python and NumPy [34] was utilized. For reading and writing pyramidal WSIs, the tiffle library [29], the openslide-python (the Python bindings for OpenSlide) [30], and zarr-python [51] were used. For training the U-Net model, segmentation-models [38] and TensorFlow [50] were used. For K-Means clustering in the lobule segmentation workflow and image

data preparation, as well as for matching prediction and ground truth instances, the SciPy implementations of KMeans and linear sum assignment [75] were used. For data augmentation during training, the albumentations library [14] was utilized. Kruskal-Wallis and Dunn’s post-hoc testing was conducted using the SciPy and scikit-posthoc packages [75, 73]. For handling and processing result and interim datasets, pandas [71] was used. For data visualization, the matplotlib library [72] was used.

## 2.8 Code Availability

The image analysis code is distributed across four projects. The main project containing the code for the image analysis pipeline and the code generating the datasets and conducting statistical testing is available from <https://github.com/matthiaskoenig/zonation-image-analysis>. The code for the training of the U-Net model and the macro-steatosis prediction on the WSI is available from <https://github.com/matthiaskoenig/steatosis-U-Net>. Image manipulation and analysis routines developed for the preprocessing pipeline, the lobule segmentation pipeline, or the macrosteatosis segmentation workflow, are available in <https://github.com/matthiaskoenig/image-utils>. The code for preprocessing and creating the dataset for annotation, and creation of the training dataset from the annotations, is available in <https://github.com/matthiaskoenig/trainingdata-curation>. The code used to handle the upload and download annotation data from and to the label studio instance is available in <https://github.com/matthiaskoenig/labelstudio>.

### 3 Results

In this work, an image analysis workflow was developed to determine lobular geometries, analyze lobular expression gradients, and quantify macrosteatosis in whole slide images (WSIs) of liver histology (Figure 3.1). The main results include the development of an image analysis workflow for lobule segmentation (Section 3.1), the quantification of macrosteatosis droplets (Section 3.2), the determination of lobular geometries (Section 3.3), and the assessment of hepatic zonation of protein expression and the effects of steatosis (Section 3.4).

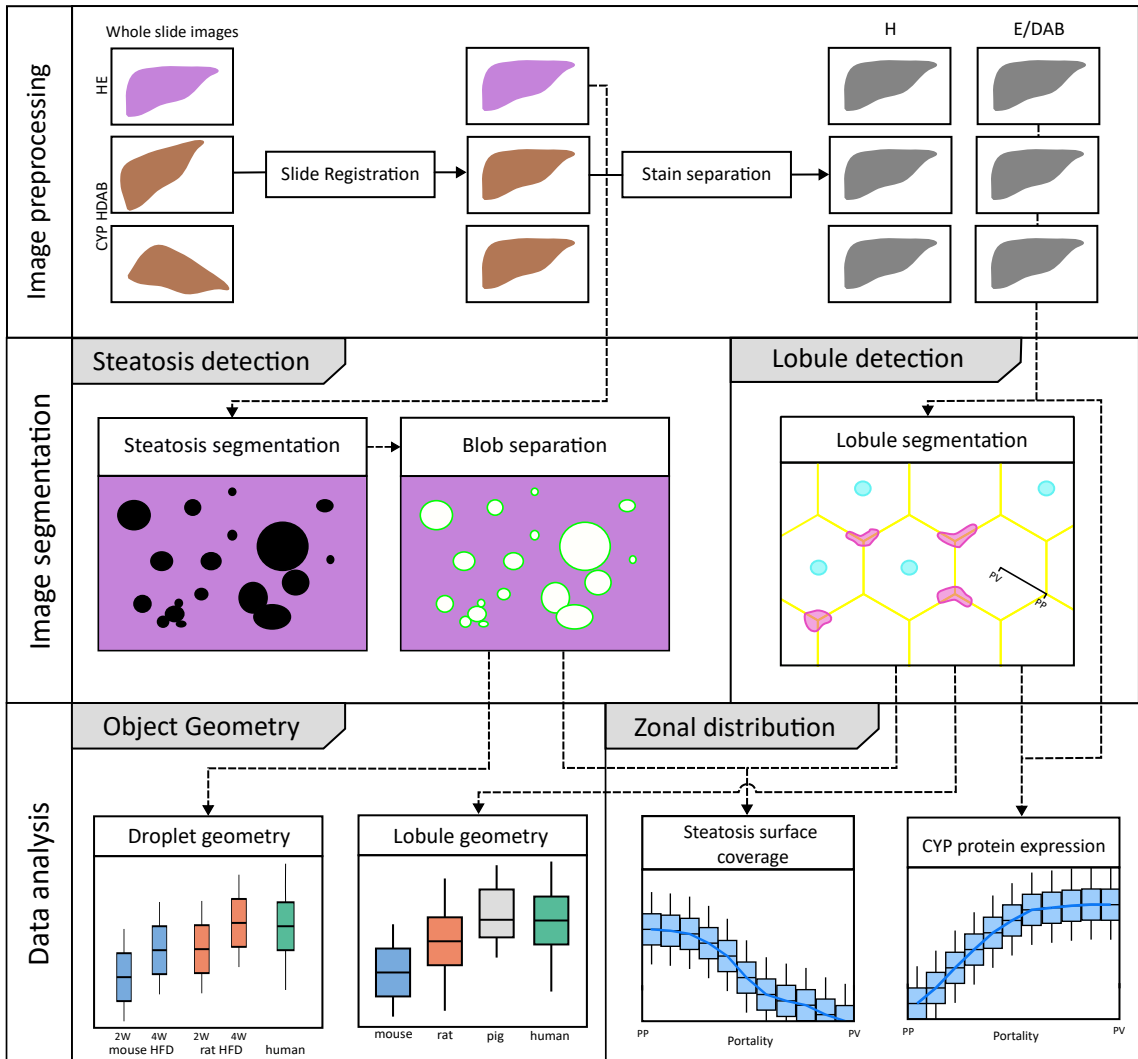


Figure 3.1: Overview of the developed image analysis workflow.

**Figure 3.1 (cont.):** A set of whole-slide images (WSI) was provided, including a Hematoxylin and Eosin (H&E) stained WSI and immunohistologically stained WSIs for the proteins Glutamine Synthetase (GS) and Cytochrome P450 isoforms 1A2, 2E1, 2D6, and 3A4. During preprocessing, these RGB images were first aligned, and then the stain components of Hematoxylin and Eosin/DAB were separated, resulting in two grayscale images representing each stain component. The DAB components from the GS and CYP images were processed through the lobule detection pipeline to identify lobule and vessel boundaries. The registered H&E slide was passed into the steatosis detection pipeline to detect macrosteatosis droplet instances. Subsequently, the resulting datasets from both pipelines were analyzed for (1) geometric properties of the lobule boundaries and macrosteatosis droplets, and (2) the zoned distribution of protein expression and macrosteatosis.

## 3.1 Image Analysis Workflow for Lobule Segmentation

This work involved developing an image analysis workflow to detect lobules and quantify the zoned expression of cytochrome proteins in liver whole slide images (WSI). The following section details the image processing steps for detecting lobule boundaries.

### 3.1.1 Lobule Detection Workflow

Previous work on the segmentation of liver lobules has used neural networks and Voronoi diagrams, which rely heavily on manual annotation of periportal and pericentral vessels in the given whole slide images (WSIs). Annotation of these slides is a time-consuming task that requires specialized pathologists to obtain reliable results. Lobules are most distinct in pig data, where they are delineated by interlobular septa. In contrast, interlobular septa are absent in humans, rats, and mice. Furthermore, the observed structures deviate significantly from the idealized symmetric hexagons and show considerable heterogeneity. This variability can be attributed to both the plane of sectioning and to biological differences. The variability and lack of physiological boundaries make the recognition of portal triads and central veins particularly challenging, especially for those without a histopathological background.

To address these challenges, I developed a method for liver lobule segmentation on WSIs that exploits the zoned expression patterns of cytochrome P450 (CYP) proteins within the lobules and does not rely on manual annotation. The established model of liver zonation describes a three-dimensional substructure called a lobule. The cross-section of a lobule appears to be polygonal, often described as a hexagon. The periportal vessels and bile ducts are located at the corners of the hexagon, while the central vessels are located in the center. Cytochrome P450 expression is heterogeneously distributed along this radial axis. Despite differences in the expression of specific isoforms, the majority of CYPs are expressed more abundantly in the pericentral region and less abundantly in the periportal region. Although the expression pattern of specific CYP enzymes on stained slides can suggest lobular substructures, the protein signal often stays at background levels in large regions between two lobule centers, making it difficult to confidently delineate boundaries. Therefore, the developed method integrates the signals of multiple CYP species to increase robustness in

the boundary locations. The method requires a set of the DAB stain components of the HDAB stained slides for different CYP/GS. The DAB component provides information on the protein expression level and is obtained using the stain separation method described in the Materials and Methods section (Section 2.4.2). The DAB stain images are the input for the lobule detection workflow. From the aforementioned model of the lobule, the following rules for locating lobule boundaries were derived:

- The boundaries should be placed in the periportal region (i.e. region of lowest CYP expression)
- Boundaries should be closed
- Portal vessels are supposed to be placed on the boundaries
- Central vessels should be located within the boundaries

Based on these rules, the method converts the provided slides into grayscale images, where dark areas correspond to pericentral zones (high CYP expression), and light areas correspond to periportal zones (low expression). These grayscale images are then skeletonized and the lobule boundaries are derived. The specific details of this method are presented in the following subsection. The processing steps of the lobule detection algorithm are shown in Figure 3.2C.

The lobule segmentation algorithm consists of nine steps:

1. Load the DAB stain components for the CYPs/GS and stack the images into a 5-channel array, each channel representing a CYP or GS
2. Apply filters for smoothing and adaptive histogram normalization
3. Cluster 5-channel image into superpixels using the SLIC algorithm
4. Classify superpixels as either foreground (tissue) or background (background/vessels)
5. Classify foreground superpixels into periportal, midzone, and pericentral and classify vessels into periportal, pericentral
6. Create grayscale portality map
7. Run thinning algorithm
8. Extract line segments from the thinned image
9. Polygonize line segments

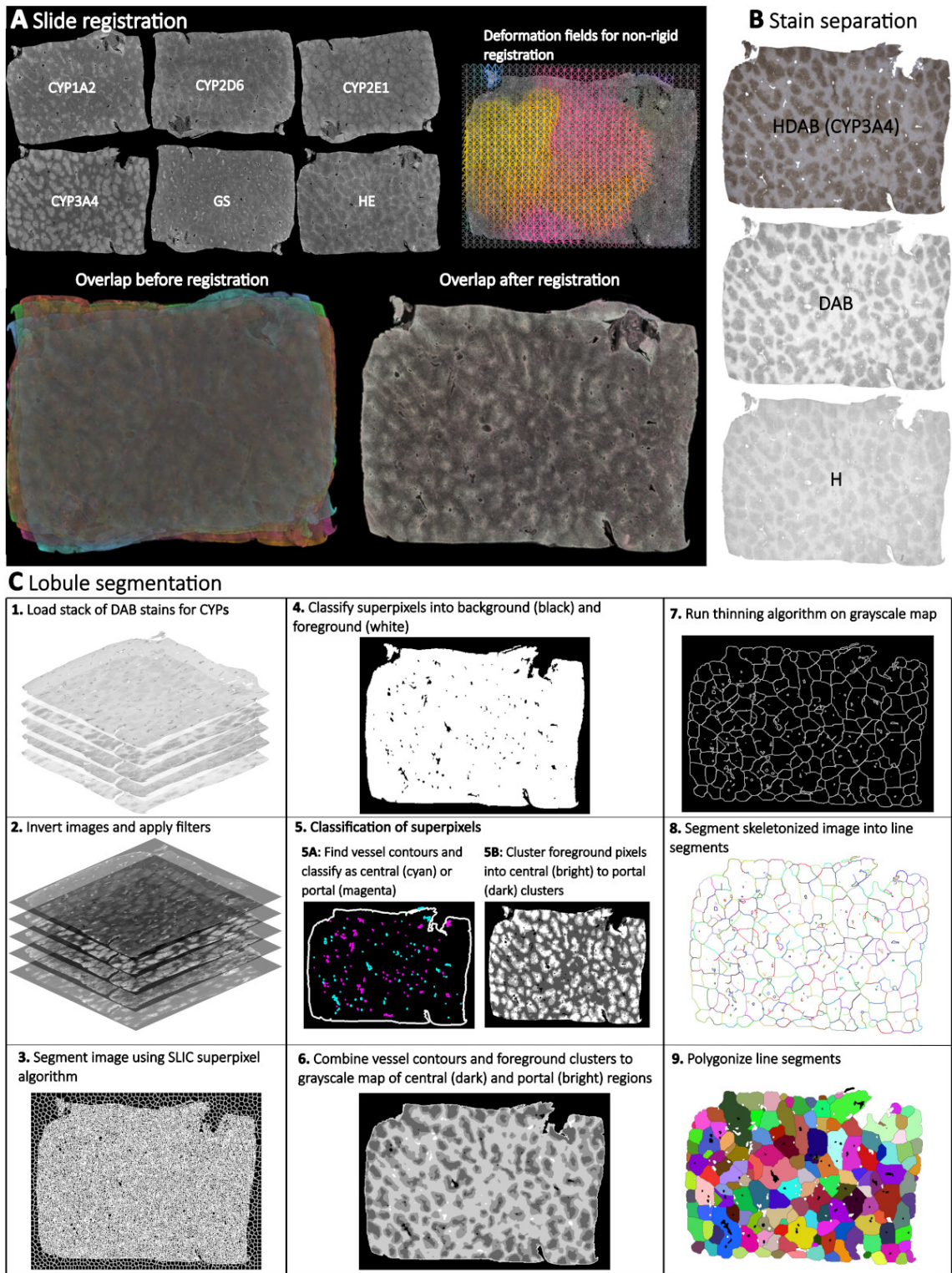


Figure 3.2: Overview of the analysis pipeline for quantifying lobular geometry and zonation patterns

**Figure 3.2 (cont.):** The pipeline consists of the following steps: **(A)** Registration of HE, GS, and CYP whole slide images (WSI) using VALIS [25]. **(B)** Stain separation of the WSI. WSIs are separated into two grayscale images representing the Hematoxylin and the Eosin or DAB component for H&E or HDAB stained slides, respectively. **(C)** Lobular segmentation of WSIs consists of nine steps: (1) Loading a stack of DAB stains for CYPs and GS. (2) Black and white image inversion and filter application. (3) Image segmentation using SLIC (Simple Linear Iterative Clustering) superpixel algorithm to generate uniform size and regular contour superpixels. (4) Classify superpixels into background (black color) and foreground (white color). (5B) Cluster foreground pixels, into central (bright) to portal (dark) clusters. (5A) Find vessel contours and classify vessels as central (cyan) and portal (magenta). (6) Combine vessel contours and foreground clusters to create a grayscale map of central (dark) and portal (light) regions. (7) Apply a thinning algorithm to the grayscale map to create a skeleton. (8) Segment the skeletonized image into line segments. (9) Polygonize the line segments to create closed polygons. The figure was adapted from Albadry et al. [1].

In the following these steps are described in detail: **(1)** To find robust boundaries, the expression signals of all CYPs were integrated into a stacked image. The registered DAB stain images for GS, CYP1A2, CYP2D6, GYP2E1, and CYP3A4 were loaded at resolution level 5 (1.25x magnification) for each region of interest (ROI). At this level, the expression patterns and vessels remain visible. Images were stacked to create a 5-channel image. Images, where the number of foreground pixels exceeded 80% of the median number of foreground pixels across all images, were discarded to exclude images with large tears or missing tissue.

**(2)** The images were inverted so that regions of high absorbance (high expression) correspond to high pixel values. Pixels were set to zero in all images in the stack if the pixel was zero in at least one of the images. A median filter was then applied and the image was convoluted to resolution level six (0.625x). Subsequently, adaptive histogram normalization was applied to reduce global differences in illumination and staining. Finally, the image was convoluted to resolution level seven (0.3125x) after again applying a median filter. Each channel was then normalized to the maximum intensity of the channel. This approach worked well for non-steatotic samples. However, in highly steatotic samples, macrosteatotic blobs appear as background, causing the median filter to produce lower values that do not accurately represent protein abundance. Therefore, the median filter was replaced by a dilation followed by a Gaussian filter. Dilation assigns the maximum value within the selected kernel to all the pixels in that kernel. This preserves the integrity of the protein abundance signal.

**(3)** The resulting 5-channel images were segmented using OpenCV's SLIC (simple linear iterative clustering) superpixelization implementation, which assimilates similar pixels into larger superpixels. Using a superpixelization approach allows the aggregation of similar pixels in terms of protein expression and spatial location.

**(4)** These superpixels were then classified as either background or foreground pixels. Specifically, superpixels with more than 10% of the pixels having a value of 0 (background) were classified as background.

**(5)** Foreground pixels were clustered into three regions: periportal, pericentral, or midzone.



For this, each superpixel was reduced to a 5-channel vector, where each element represented the mean intensity of the channel in the superpixel. K-means clustering was then used to group these vectors into three clusters. The cluster centers were sorted in ascending order by Euclidian distance. Given the high expression of the CYPs and GS in the pericentral, and low expression in the periportal zone, the cluster labels corresponded to periportal, midzone, and pericentral. These labels were mapped back to the superpixel representation labeling each foreground pixel as either pericentral, periportal, or midzone. A mask was created from the background pixels and contour detection was used to obtain the vessels and tissue boundary on the slide. Using the information from the clustered foreground pixels, the vessels were classified as pericentral or periportal. To achieve this, each vessel contour was transformed into a binary mask. In addition, a second mask was created by dilating the vessel contour mask, and the difference of the two masks was obtained using the XOR operator. The difference was used to select pixels adjacent to the vessel from the foreground image and the number of periportal, midzone, and pericentral labeled pixels in this selection was counted. As anticipated, the number of pericentral labels was higher for pericentral vessels and vice versa for periportal vessels. The count vectors for all vessels were clustered into two clusters using the K-means algorithm. Contours in the cluster with a higher number of pericentral pixels were classified as pericentral while the other vessels were labeled as periportal.

(6) The intermediate result was an array labeling the pixels as background, periportal, midzone, or pericentral, along with a list of periportal and pericentral vessel contours. A grayscale representation was created from the clustered foreground and vessel contours, with pericentral vessels colored black (zero) and periportal vessels colored white (255). The foreground zones were uniformly distributed across the grayscale spectrum in between, with dark and light corresponding to pericentral and periportal zones, respectively.

(7) Finally, an OpenCV implementation of a thinning algorithm was used to skeletonize the grayscale representation. The remaining lines are in the expected center of the periportal zones and vessels (bright zones), marking the potential boundaries.

(8) To facilitate polygonization, a pixel-walking algorithm was developed that recursively aggregates neighboring true pixels into line segments. Line segments are ordered lists of coordinates representing adjacent pixels. The algorithm operates on a binary image where 8-connected white pixels mark the lines. First, the images were converted into a list of coordinates of the white pixels. Then three lists were initialized: one to store completed segments, one for segments awaiting processing, and one for nodes, which are pixels with multiple adjacent pixels. The algorithm proceeds by iterating through the list of pixels. The neighbors of the first pixel are examined to initialize the first line segments, which are appended to the segments-to-do list. As long as there are segments to process, the next segment is retrieved and analyzed. Neighboring pixels are evaluated for the last pixel in the current segment. If a single neighboring pixel is found, it is appended to extend the segment, continuing the process. If no neighbors exist, the segment is finished and added to the list of finished segments. If multiple neighbors exist, new line segments are created from the current pixel to each neighboring pixel, and these segments are added to

the segments-to-do list. Finally, the current segment is completed and added to the list of finished segments.

(9) The result is a list of line segments representing the boundaries. These segments were then converted to polygons using the Shapely [27] library. All line segments that were not part of a closed path were discarded. The polygons delineating the lobule boundaries were integrated with the vessel data from the vessel classification steps, and the combined data was exported as GeoJSON.

### 3.1.2 Protein Portality Map

One objective of this thesis was to quantify the differences in zoned protein expression within liver lobules across different species. After identifying the boundaries of the lobules, the next step was to integrate this spatial information with the protein signals to derive expression gradients. Therefore, the relative position of each pixel in the lobule was calculated. The portality  $p$  was defined as

$$p(x, y) = 1 - \frac{d_c(x, y)}{d_c(x, y) + d_p(x, y)} \in [0, 1] \quad (3.1)$$

where  $d_c$  and  $d_p$  denote the distance of a pixel to the nearest central and portal pixel, respectively.

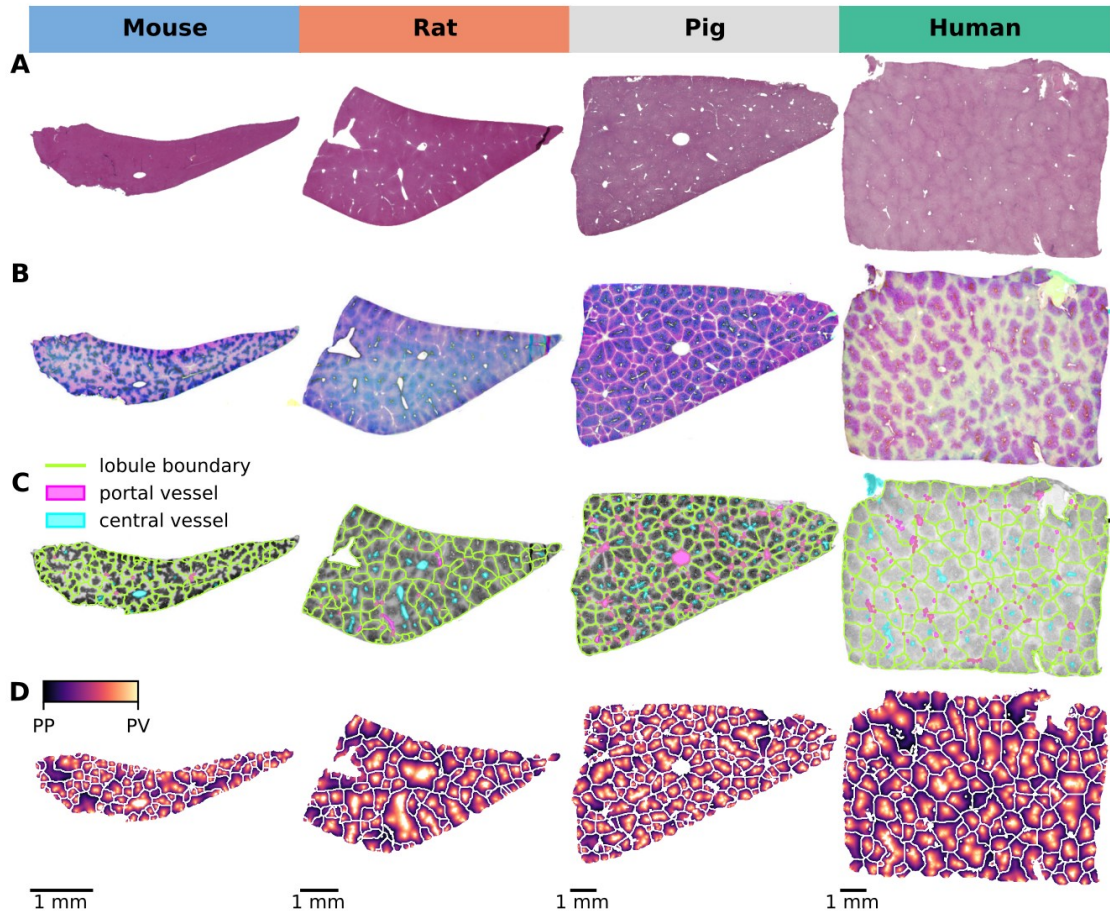
For each identified lobule, a periportal and a pericentral boundary mask were created using the lobule boundary and vessel polygons. In the pericentral boundary mask, all pixels were set to false if they were situated within a central vessel or if the weighted intensity across all channels fell within the 99th percentile. This was necessary because not every detected lobule contained a detected central vessel. In the portal boundary mask, all pixels were set as false if they were situated outside of the lobule boundary or if they were located within a portal vessel. The portal and central distances were calculated by applying the OpenCV distance transform function to the aforementioned masks. This function calculates the distance for each pixel to the nearest background pixel.

For each protein, the intensity was background corrected and normalized by

$$I_N(x, y) = \frac{I(x, y) - I_{bg}}{I_{max} - I_{bg}} \in [0, 1] \quad (3.2)$$

where  $I_N(x, y)$ ,  $I_{bg}$  and  $I_{max}$  denote the intensity of the pixel, the background intensity, and the maximum intensity on the slide, respectively. The maximum intensity was defined as the 99%tile of the foreground pixel intensities to avoid outliers. The background intensity was estimated by the 20th percentile of the foreground pixels of the GS slide for each subject. GS expression is limited to a small area of the lobule around the pericentral vessel, allowing for the remainder to be used as a robust background estimate. For each pixel, the normalized intensity and portality were recorded in a data frame for subsequent analysis.

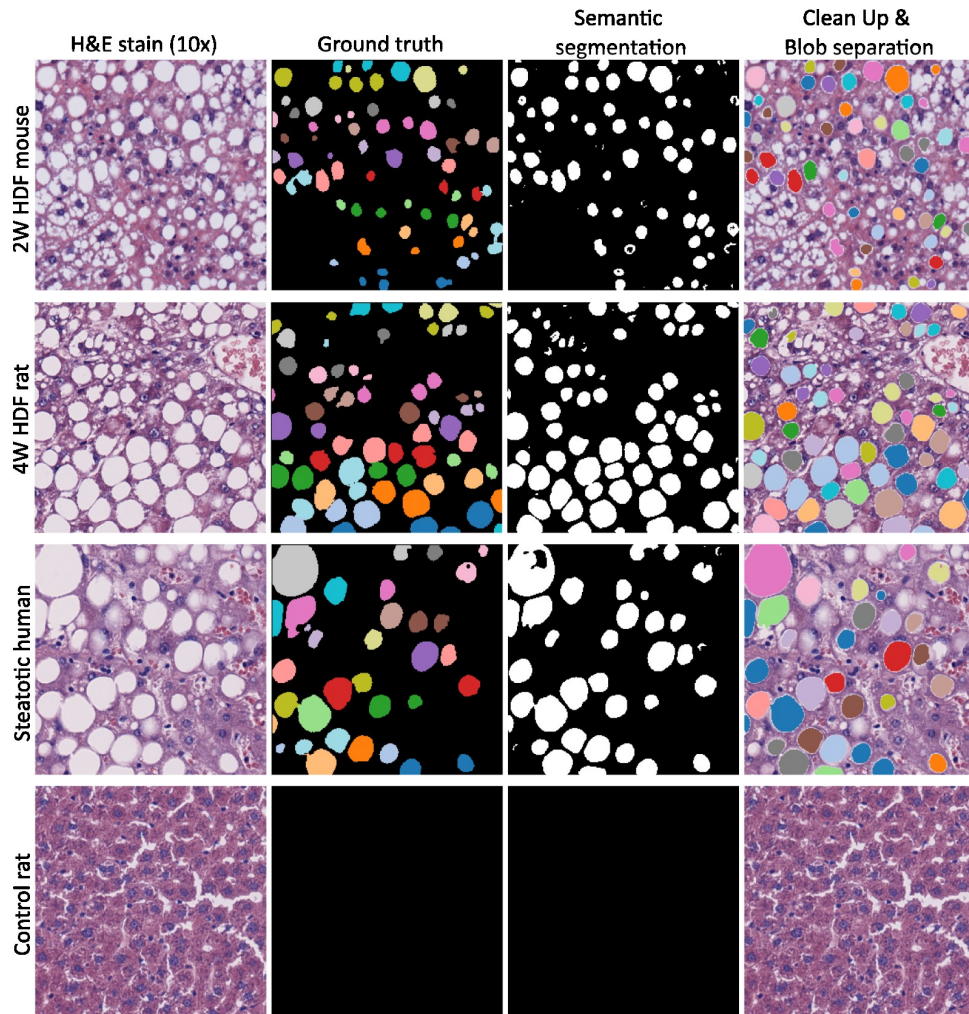
A visualization of the output of the described image analysis pipeline for exemplary subjects is presented in Figure 3.3



**Figure 3.3: Lobule boundary detection and calculation of portality:** (A) H&E staining images from normal liver tissues of all four species (mouse, rat, pig, and human) are presented, illustrating the lobular architecture specific to each species. (B) Images are normalized and color stains are separated to ensure optimal comparison across species. This step involves transforming the images to a consistent format and isolating the different stains used in the analysis. (C) The detection process clearly outlines the lobular boundaries and identifies the distribution of lobules, central vessels, and portal vessels. The example shown uses the CYP2E1 staining component to demonstrate the detection and delineation of these regions. (D) The distance between central and portal regions is mapped on each lobule. This mapping enables a quantitative analysis of the spatial arrangement within the lobules, providing insight into the zonation patterns of the lobules across different species.

## 3.2 Quantification of macrosteatosis droplets in whole-slide images (WSI)

In this work, an image analysis workflow was developed to detect macrosteatosis whole slide images (WSI) of the liver. The macrosteatosis segmentation workflow consisted of two steps: (1) binary semantic segmentation by the U-Net model into steatotic and non-steatotic pixels and (2) postprocessing and clump separation to obtain individual droplet instances. Examples of the intermediate results of the detection workflows are presented in Figure 3.4.

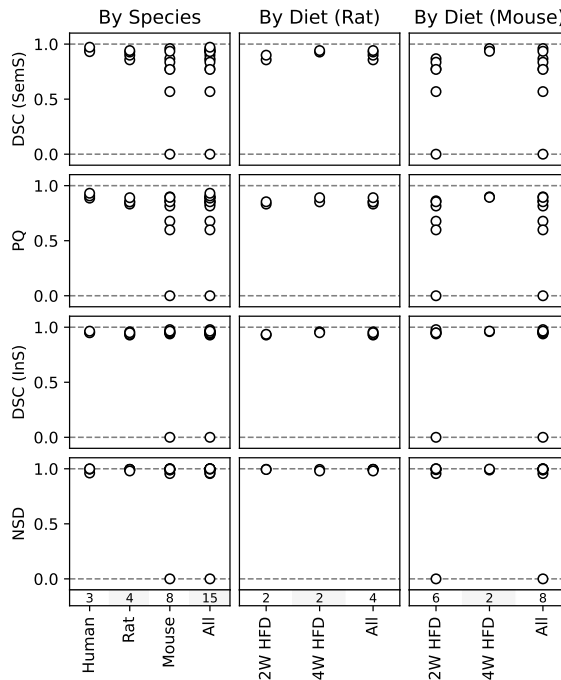


**Figure 3.4: (A) Macrosteatosis droplet detection:** Example predictions on image tiles for a 2W HFD mouse, a 4W HFD rat, a steatotic human, and a control rat sample. The columns from left to right depict (1) the stain-normalized H&E stain input image. (2) the ground truth annotations that were created from the image annotations. (3) The semantic prediction output of the U-Net model. (4) The final instance segmentation result after postprocessing and clump separation.

### 3.2.1 Steatosis Segmentation Performance

The performance of the segmentation model was evaluated on the test data set consisting of 15 images. The pixel-wise Dice-Similarity coefficient (DSC (SemS)) was computed to

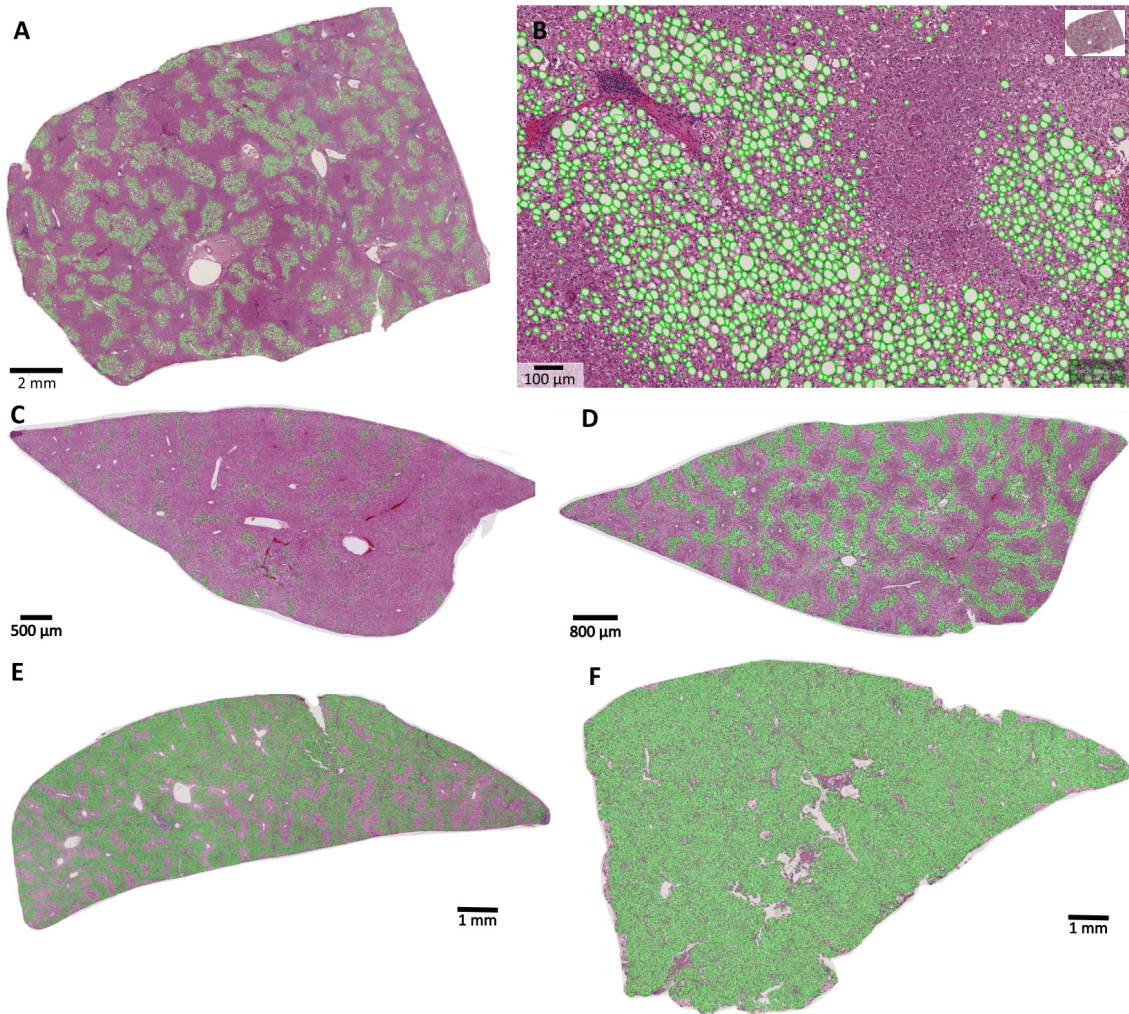
evaluate the semantic segmentation performance of the U-Net model. For evaluation of the instance segmentation by the model in combination with blob separation post-processing, the instance-wise Dice-Similarity coefficient (DSC (InS)), the Panoptic Quality PQ, and the Normalized Surface Distance (NSD) were calculated. The metrics were calculated per image and aggregated by averaging over the data set. The metrics were also analyzed per subgroup to provide a detailed view of the model performance on each subset of the image data. The results are visualized in Figure 3.5. Numerical values are reported in Table S15. The model performs well on human samples, rat samples of both 2W HFD and 4W HFD groups, and mouse 4W HFD samples. However, the metrics show lower performance for the 2W HFD group. In one instance involving a test image from the 2W HFD mouse group, the evaluation metrics returned a value of zero. Such cases can occur when the prediction set is empty while ground truth objects are present, or vice versa. Mice generally exhibited smaller droplet sizes and a smaller total area of steatosis on the slides (Figures 3.7, 3.8). These lower performance metrics suggest that the model struggles to accurately detect smaller droplets, indicating that it is not yet fully optimized for such cases.



**Figure 3.5: Validation metrics for image segmentation:** The pixel-wise (semS), and instance-wise (InS) Dice Similarity Coefficient (DSC), the Panoptic Quality (PQ), and the Normalized Surface Distance (NSD) were calculated per test image in and averaged over each presented subset.

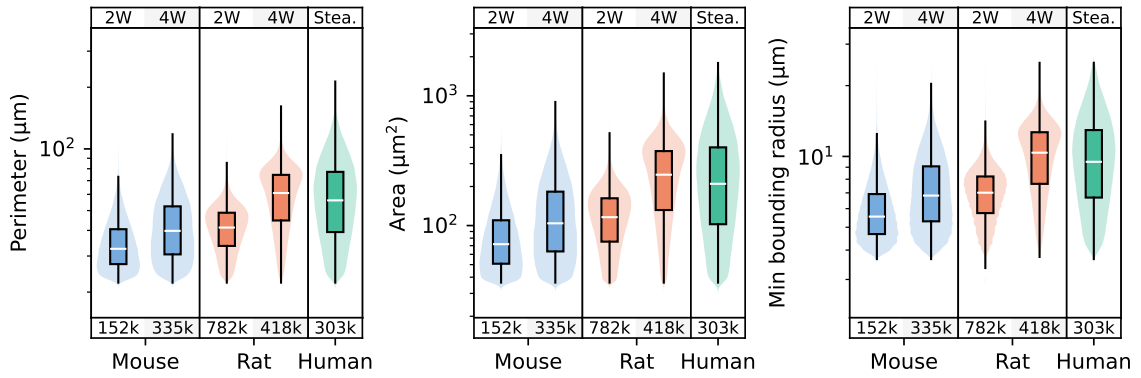
### 3.2.2 Quantification of Macrosteatosis

The steatosis detection workflow described in Section 2.5 was applied to the WSI resulting in a data set of detected macrosteatosis droplets. An example of the detection result on a human WSI is shown in Figure 3.6. The dataset included the position on the slide (centroid), and the geometrical properties perimeter, surface area, and minimum bounding circle radius for each detected droplet.



**Figure 3.6: Macrosteatosis detection on WSI:** (A) Human (B) Close-up section of the same human WSI (C) Mouse 2W HFD (D) Mouse 4W HFD (E) Rat 2W HFD (F) Rat 4W HFD. The images were obtained using QPath [8].

The macrosteatosis detection pipeline detected 152k and 335k droplets for the mouse 2W HFD, and 4W HFD groups, 782k and 418k droplets for the rat 2W HFD and 4W HFD groups, and 303k droplets for the human steatotic samples, respectively. The median (Q1, Q3) droplet sizes were  $71.9 (50.8, 109.8) \mu\text{m}^2$  and  $104.0 (63.2, 182.5) \mu\text{m}^2$  for mouse 2W HFD and 4W HFD and  $124.5 (75.2, 162.3) \mu\text{m}^2$  and  $246.5 (131.7, 375.0) \mu\text{m}^2$  for rats 2W HFD and 4W HFD, respectively. Human steatotic samples displayed a median droplet size of  $209.8 (102.4, 400.6) \mu\text{m}^2$ . The droplet statistics are displayed in Figure 3.7. Another finding was that for both mice and rats, the variability of droplets as measured by the inter-quartile range (IQR) increased from the 2W HFD (mouse  $59 \mu\text{m}^2$ , rat  $87 \mu\text{m}^2$ ) to the 4W HFD group (mouse  $119 \mu\text{m}^2$ , rat  $243 \mu\text{m}^2$ ). The human steatotic samples showed the largest variability in droplet size with an IQR of  $298 \mu\text{m}^2$ . Unlike the rat and mouse samples, which were obtained from animals that were housed and fed under controlled conditions, the human samples were obtained from patients with a variety of backgrounds and medical conditions. As a result, steatosis patterns are expected to vary considerably between subjects, which is reflected in the variability of the data.



**Figure 3.7: Geometrical properties of macrosteatosis droplets:** For all species and steatotic groups, the droplet perimeter, area, and minimum bounding radius were calculated for each droplet per section. The results are depicted using violin plots. The boxes represent the first quartile (Q1) and third quartile (Q3). The whiskers extend to the last datum less than  $Q3 + 1.5 \cdot IQR$  and the first datum greater than  $Q1 - 1.5 \cdot IQR$ , respectively. IQR denotes the interquartile range ( $Q3 - Q1$ ).

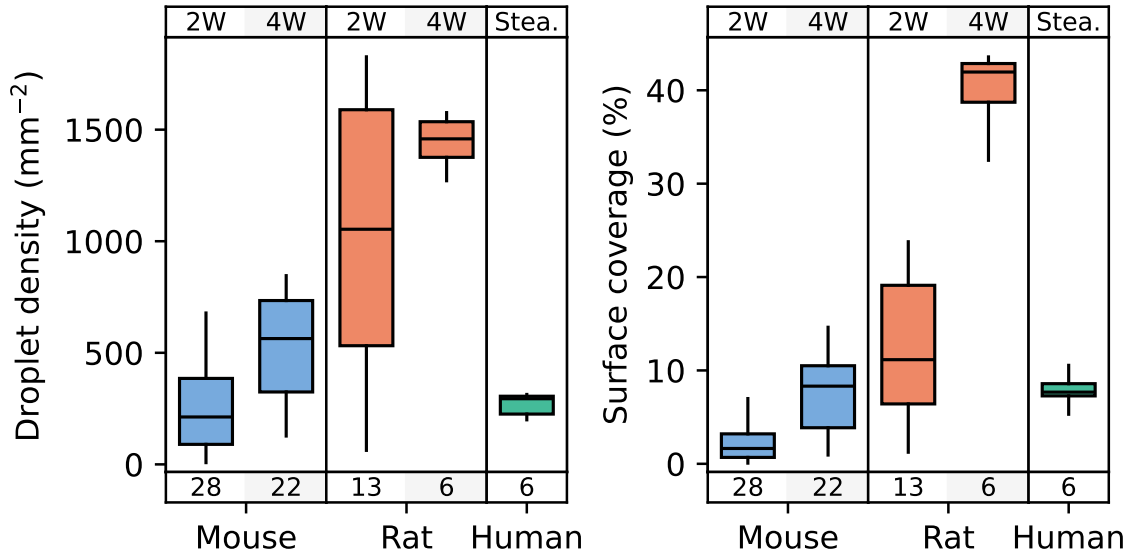
Additionally, the droplet density and surface coverage were analyzed on the WSI. Droplet density was calculated as the ratio of the droplet count on the slide to the total tissue area of the slide. The surface coverage was defined as the ratio of the summed area of all droplets on the slide to the total tissue area on the slide. The statistics are visualized in Figure 3.8.

The median (Q1, Q3) droplet density was 212 (90, 385)  $\text{mm}^{-2}$ , 564 (324, 734.2)  $\text{mm}^{-2}$  for 2W HFD and 4W HFD in mice, respectively. For rat 2W HFD and 4W HFD, droplet densities were 1054 (532, 1589)  $\text{mm}^{-2}$  and 1458 (1376, 1536)  $\text{mm}^{-2}$ , respectively. Human steatotic samples displayed droplet densities of 294 (225, 305)  $\text{mm}^{-2}$ .

The median (Q1, Q3) surface coverage was 1.6 (0.7, 3.2)%, 8.3 (3.7, 10.5)% for 2W HFD and 4W HFD in mice, 11.2 (6.4, 19.1)%, 42.0 (38.7, 42.9)% for 2W HFD and 4W HFD in rats, and 7.7 (7.3, 8.6)% for steatotic human samples, respectively.

These results show that a longer duration of the diet resulted in greater amounts of macrosteatosis in mice and rats. The results suggest that in mice and rats, a high total droplet density and a higher droplet area fraction correlated with the median droplet sizes (Figures 3.7 and 3.8). However, human samples had a median droplet size comparable to that of 4W HFD rats (Figure 3.7). The droplet density and area fraction are lower than in the same rat group. This shows that in human steatotic samples, steatosis tends to affect smaller more localized areas whereas the high-fat MCD has a more homogeneous distribution of steatosis across the tissue. This distribution is well aligned with the visual impression of the WSI (Figure 3.6).

Droplet density in 2W HFD rat samples shows extensive variability as indicated by the inter-quartile range IQR of 1058  $\text{mm}^{-2}$  with the highest values comparable to the 4W HFD group, while the surface coverage is lower than in the 4W HFD group. This finding suggests that a large portion of the droplets emerge in the first two weeks of the diet in rats and only grow during the third and fourth weeks.



**Figure 3.8: Macrosteatosis droplet density and surface coverage:** Droplet density and surface coverage on the WSI were analyzed for all species and steatotic groups. Droplet density was calculated as the fraction of the droplet count on the WSI and the total tissue area of the WSI. The surface coverage was defined as the fraction of the summed area of all droplets on the WSI and the total tissue area of the WSI. The results are depicted using box plots. The boxes represent the first quartile (Q1) and third quartile (Q3). The whiskers extend to the last datum less than  $Q3 + 1.5 \cdot IQR$  and the first datum greater than  $Q1 - 1.5 \cdot IQR$ , respectively. IQR denotes the interquartile range ( $Q3 - Q1$ ).

### 3.3 Lobular Geometry

#### 3.3.1 Species-Specific Lobular Geometry

Using the lobule segmentation algorithm described in the methods section (Section 3.1), a set of lobule polygons was obtained. Based on the segmented lobules, species-specific lobular geometry was assessed using several key geometric parameters: perimeter (circumferential length of the lobule’s outer edge), area (surface area of the lobule), compactness (ratio of the lobule’s polygonal area to that of a circle with the same perimeter), and minimum bounding radius (radius of the minimum bounding circle enclosing the lobule). Figure 3.9A illustrates the resulting geometric parameters across different species, with detailed numerical values available in Supplementary Table S1. These parameters provide insights into the structural variation of lobules between species and highlight correlations between these geometric features.

Geometric parameters were calculated for 1,530 segmented lobules in mice, 669 in rats, 698 in pigs, and 1,074 in humans. Across species, there is a noticeable increase in lobule size, with larger species exhibiting larger lobules. For example, the median (Q1, Q3) radius of murine lobular boundaries was 330 (249, 442)  $\mu\text{m}$ , while human lobules were 559 (309, 830)  $\mu\text{m}$ . The specific liver biopsy site may also influence these differences in human lobular geometry. Similarly, the perimeter of liver lobules increased with species size, approximately 1.7-fold from mice (1.94 (1.44, 2.64) mm) to humans (3.37 (1.84, 5.09) mm). As expected, the lobular area increased about three-fold across species, from 0.19 (0.11, 0.35)  $\text{mm}^2$



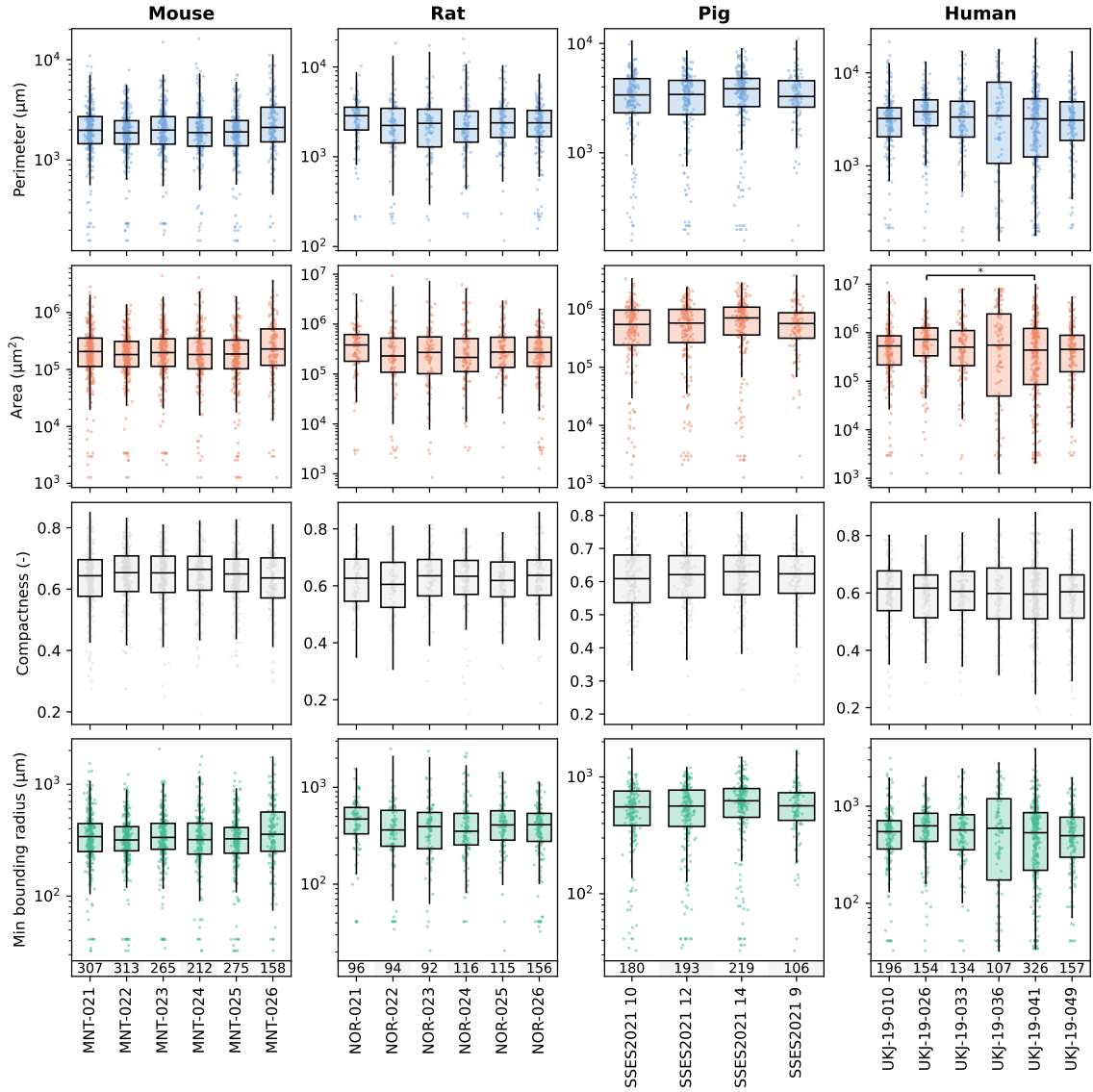
in mice to 0.53 (0.16, 1.17) mm<sup>2</sup> in humans. In contrast, compactness, which indicates lobular roundness, showed a slight decrease with species size. Compactness mean and standard deviation (SD) were highest in mice (0.64 (0.10)), followed by rats (0.62 (0.11)), pigs (0.61 (0.10)), and humans (0.59 (0.12)). This suggests that larger lobules tend to be less round across species. Statistical analysis revealed significant differences in lobular size but little variation in compactness between species. All species showed similar levels of individual variability in geometric parameters, with slightly higher variability observed in humans compared to other species. Intra-individual variability in geometric parameters was negligible across species, as indicated by consistent results within subjects and lobes (Figure 3.10). The correlation structure of geometric parameters was highly consistent across species (Figure 3.9B). Area and perimeter showed strong positive correlations ( $r = 0.99$ ) in all species, indicating that larger lobules tend to have longer perimeters. Conversely, compactness exhibited weak to moderate negative correlations with perimeter ( $r$  in the range of -0.54 to -0.63) and area ( $r$  in the range of -0.44 to -0.55), suggesting that larger lobules may appear more oval due to variable sectioning angles.

Comparison of the medians of the lobular geometry parameters across subjects within each species revealed no significant differences (Figure 3.10). This result underscores the robustness of the segmentation algorithm across different samples.

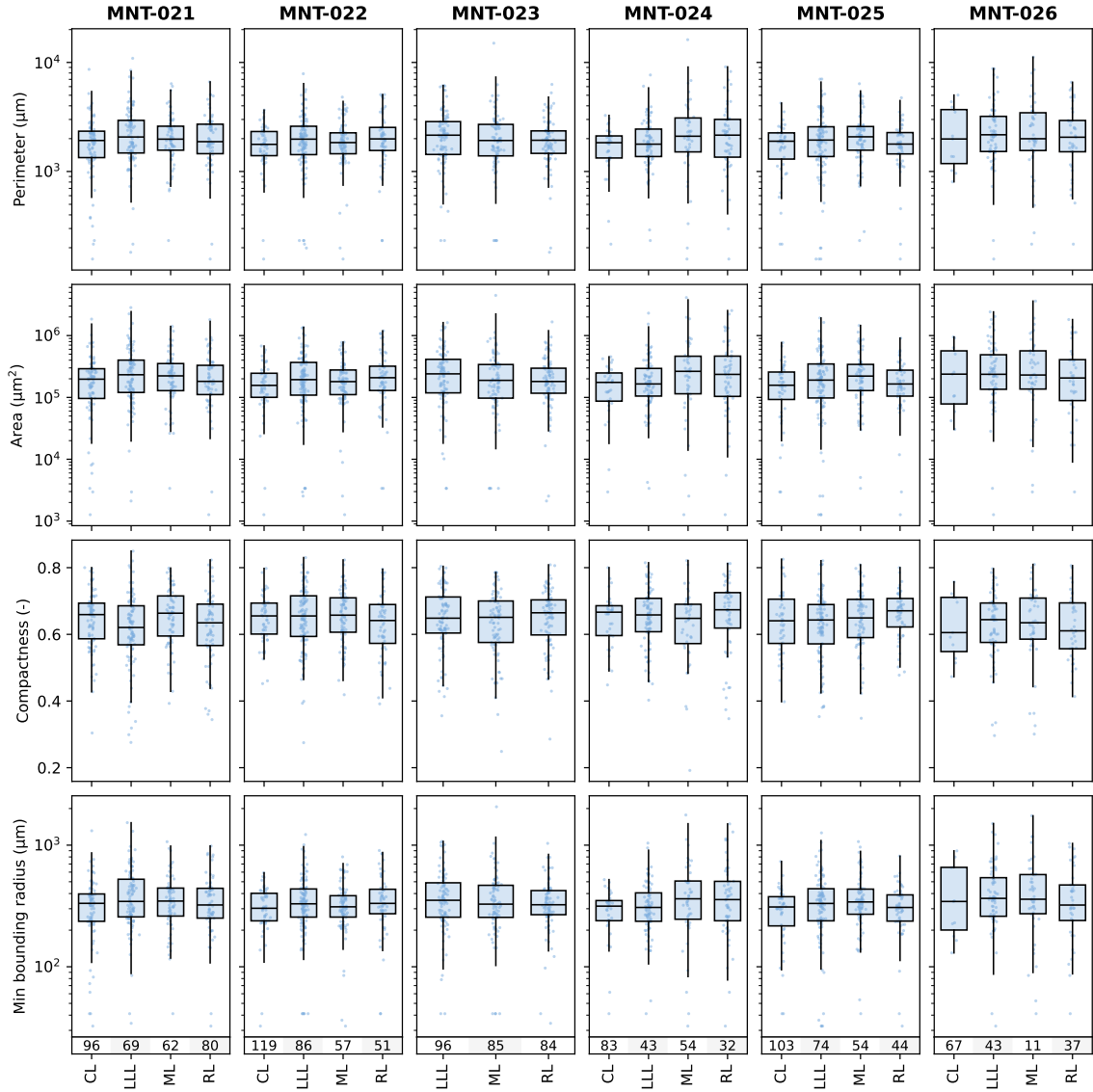
For mice, samples from different liver lobes were analyzed. Comparing the geometrical features between lobes of different mice did not reveal any statistically significant differences (Figure 3.11). Therefore, I conclude that lobular geometry is homogeneous across the entire liver in mice.

The similarity in median values and range of geometric parameters across species (Figure 3.9A), along with consistent correlation structures (Figure 3.9B), suggests a robust and comparable 3D structure of hepatic lobules across species. In conclusion, lobular geometry appears to be a stable feature with low variability within and between species, but significant variability is observed between different lobules.





**Figure 3.10: Intra- and inter-individual variability of lobular geometric parameters in human, pig, rat, and mouse** The boxes represent the first quartile (Q1) and third quartile (Q3). The whiskers extend to the last datum less than  $Q3 + 1.5 \cdot IQR$  and the first datum greater than  $Q1 - 1.5 \cdot IQR$ , respectively. IQR denotes the interquartile range ( $Q3 - Q1$ ). Significance levels are indicated as follows: \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ . Adapted from Albadyri et al. [1].



**Figure 3.11: Intra-lobe variability in lobular geometric parameters in mice** The boxes represent the first quartile (Q1) and third quartile (Q3). The whiskers extend to the last datum less than  $Q3 + 1.5 \cdot IQR$  and the first datum greater than  $Q1 - 1.5 \cdot IQR$ , respectively. IQR denotes the interquartile range ( $Q3 - Q1$ ). Significance levels are indicated as follows: \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ . Adapted from Albadyr et al. [1].

### 3.3.2 Impact of Macrosteatosis on Lobular Geometry

The lobule segmentation algorithm was applied to whole slide images (WSI) of steatotic liver samples from mice, rats, and humans. For mice and rats, two groups were analyzed: one group was fed a high-fat diet (HFD) for two weeks (2W HFD) and another for four weeks (4W HFD). As with the control samples, the geometric parameters — perimeter, area, compactness, and minimum bounding radius — were compared across the control, 2W HFD, and 4W HFD groups for both species and between the control and steatosis groups for human samples (Figure 3.12).

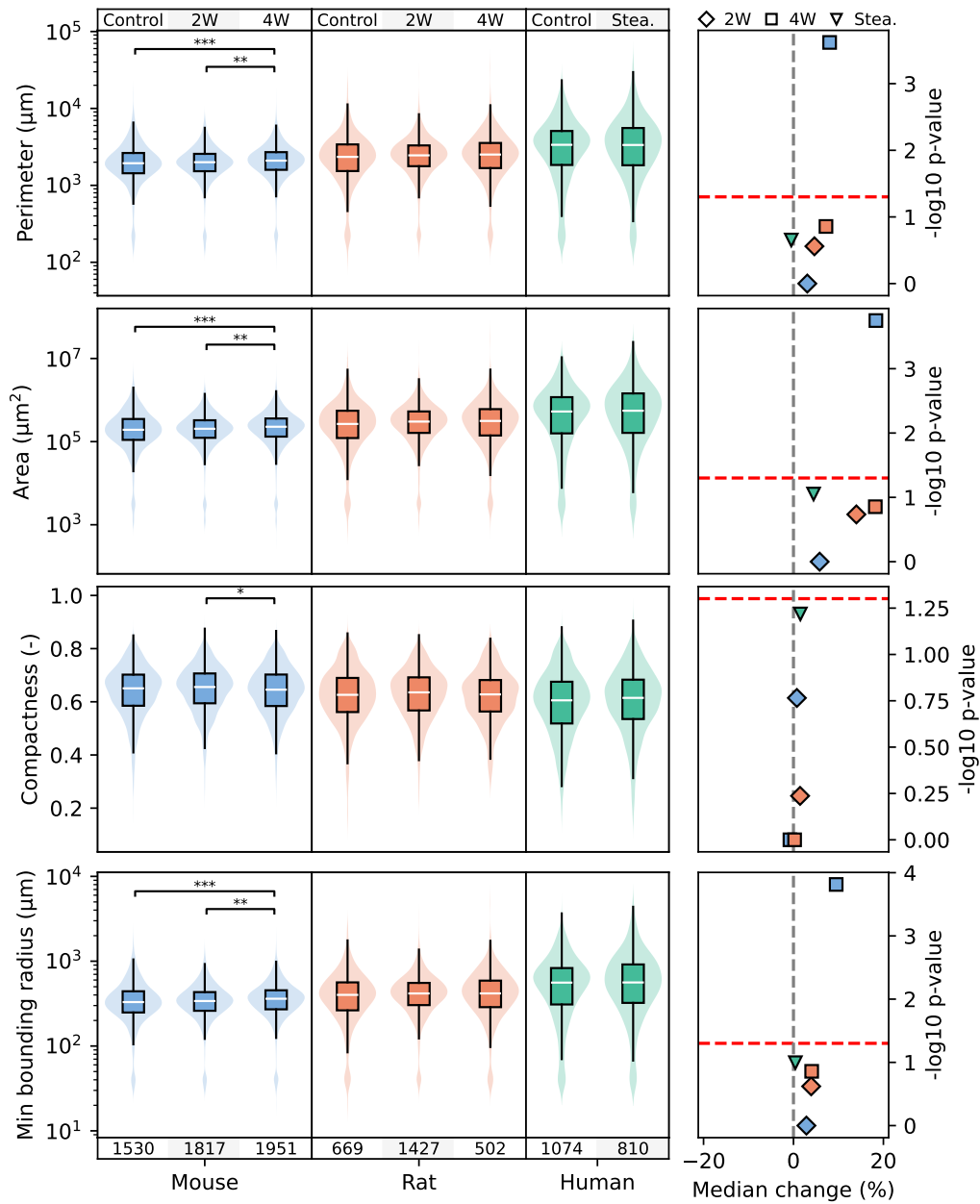
In all species, the steatotic groups showed similar distributions. A significant increase in median lobular area was observed in mice after 4 weeks of a high-fat diet (4W HFD), with an 18% increase ( $p=1.8e-4$ ). In contrast, non-significant increases in lobule area were found in 2W HFD mice (6%,  $p=1$ ), 2W HFD rats (13%,  $p=0.18$ ), 4W HFD rats (18%,  $p=0.13$ ), and steatotic human samples (4%,  $p=0.09$ ) compared to their respective controls. In addition, the median values for perimeter and minimum bounding radius showed significant increases in the 4W HFD group compared to the control group. However, no significant differences were observed in the compactness of lobules between the control and steatotic groups across all species.

The geometric parameters of the lobules showed high variability in both control and steatotic groups. Although mostly insignificant, there was a trend toward increased lobule size in the steatotic groups. However, the extent of steatosis varied considerably between individuals within the same group (Figure 3.8) and could be heterogeneously distributed across the slide, leading to different amounts of steatosis in individual lobules.

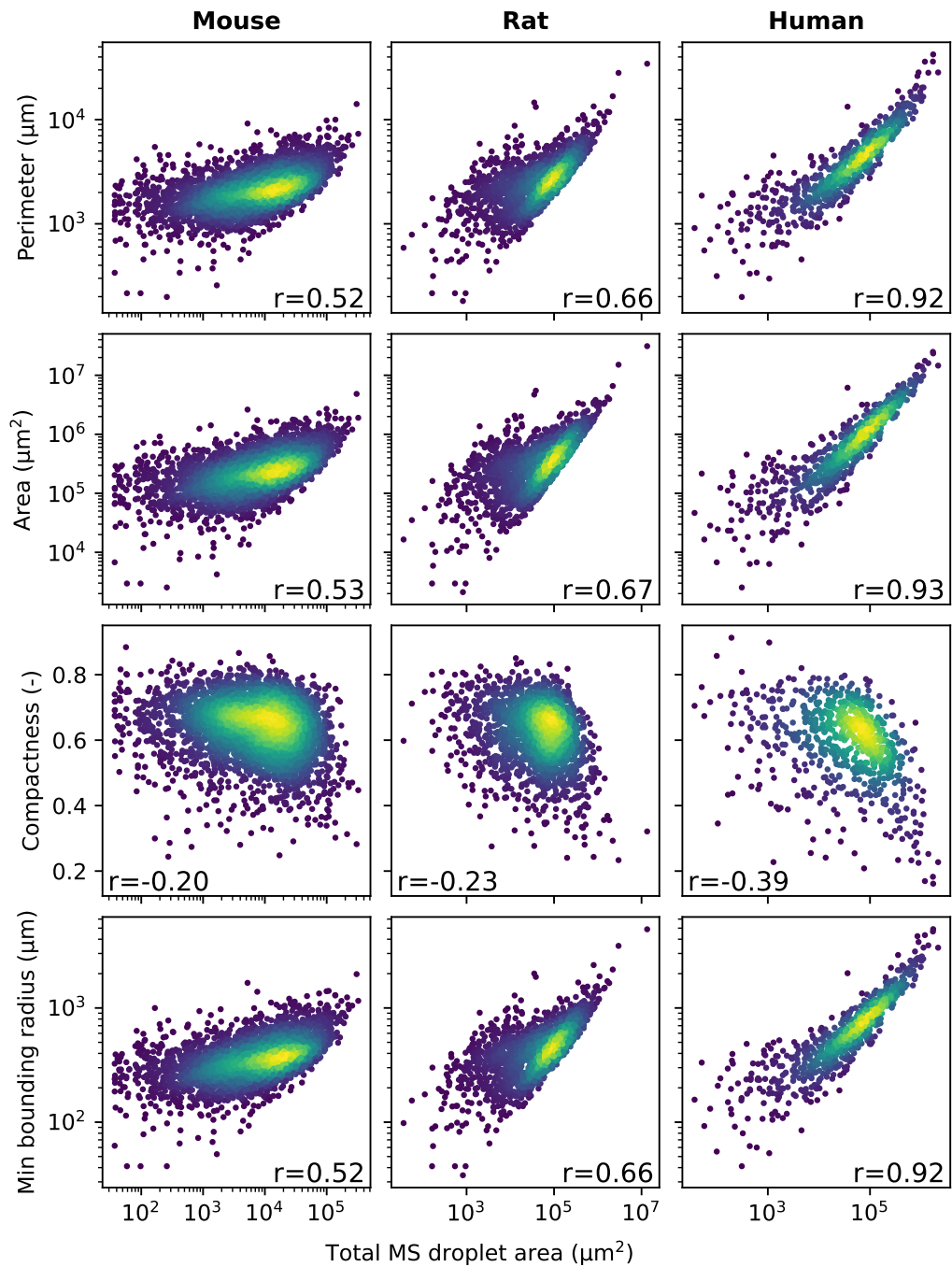
Therefore, the effect of fat content within individual lobules on lobular geometry was evaluated. The area of the macrosteatosis droplets within each lobule was summed for all detected lobules, and the geometric parameters of the lobule were plotted against the total macrosteatosis droplet area (Figure 3.13). The data shown includes all data points from the 2W HFD and 4W HFD groups for mice and rats, as well as the Steatosis group for human samples.

The area, perimeter, and minimum bounding radius showed similar correlations across all species. The strongest correlations were observed in the human data ( $r=0.92-0.93$ ), followed by rat ( $r=0.66-0.67$ ), and mouse ( $r=0.52-0.53$ ). Compactness showed a weak to moderate negative correlation with lobule fat content, with the strongest correlation again seen in the human data ( $r=-0.39$ ), followed by rat ( $r=-0.23$ ) and mouse ( $r=-0.20$ ).

These results suggest that a higher fat content increases lobular size and are consistent with a previous report [32].



**Figure 3.12: Effect of steatosis on lobular geometry:** For all species, the lobular perimeter, area, compactness, and minimum bounding radius were calculated for the control and the steatotic groups for each lobule per section. The results are depicted using violin plots. The boxes represent the first quartile (Q1) and third quartile (Q3). The whiskers extend to the last datum less than  $Q3 + 1.5 \cdot \text{IQR}$  and the first datum greater than  $Q1 - 1.5 \cdot \text{IQR}$ , respectively. IQR denotes the interquartile range ( $Q3 - Q1$ ). Significance levels are indicated as follows: \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ . The right-most column displays the fold-change of the median of the steatotic groups compared to the control group against the p-value of the group-wise post-hoc test between the groups.



**Figure 3.13: Correlation of lobular geometry and lobular fat content:** Lobular perimeter, area, compactness, and minimum bounding radius were calculated for all lobules and plotted against the total macrosteatosis droplet area within the lobule. The displayed data contains all detected lobules in the WSIs of 2W HFD and 4W HFD groups of mouse and rat, and the steatosis group of humans. Fat content was calculated by summing the area of all the macrosteatosis droplets located in each lobule. The correlation between the lobular geometric parameters was assessed using the Spearman rank correlation coefficient.

## 3.4 Hepatic Zonation

### 3.4.1 Quantification of Zonated Protein Expression

The subsequent analysis focused on quantifying the zonated expression of CYP enzymes along the portal-venous axis across entire liver lobules in four species (Figure 3.14). By leveraging the segmented lobules, the position-dependent protein expression within each species was precisely determined. Each pixel within a lobule was assigned a position, ranging from periportal (PP, 0) to perivenous (PV, 1), based on their proximity to the nearest periportal or perivenous region. This approach allowed analysis of zonation patterns of GS and CYP proteins across different lobules and species.

Analysis of the combined zonated expression of all markers revealed distinct patterns unique to each protein and species. As expected, histological staining with Hematoxylin and Eosin (HE) showed a consistent flat line across all species, as HE staining primarily delineates the morphological structure of hepatic lobules without indicating zonation differences along the sinusoid.

For clarity, the zones along the portal-venous axis are referred to as zone 1 (periportal), zone 2 (midzone), and zone 3 (pericentral) in this section.

**Glutamine Synthetase (GS)** demonstrated a consistent gradient and zonation pattern along the entire portal-venous axis across all four species, as shown by the superimposed plots of normalized staining intensity. GS showed static expression primarily in zone 3, encompassing two to three layers of pericentral hepatocytes, with no observable periportal distribution pattern in any of the species.

**Cytochrome P450 1A2 (CYP1A2)** displayed relatively similar gradient and zonation patterns in mice, rats, and humans, predominantly localized in zone 3 and extending into zone 2 within the adjacent five to six rows of pericentral hepatocytes. In contrast, pigs showed a gradient distribution of normalized intensity predominantly in zones 3 and 2, extending into zone 1 of periportal hepatocytes.

**Cytochrome P450 2D6 (CYP2D6)** exhibited a uniform and consistent zonation pattern across all species, showing a panlobular distribution within the liver lobules along the portal-venous axis. Notably, CYP2D6 did not show a distinct zonation pattern with higher protein levels in the perivenous region compared to the periportal region.

**Cytochrome P450 2E1 (CYP2E1)** displayed a linear gradient distribution of normalized intensity across liver lobules in different species, predominantly in zones 3 and 2. Rats showed a higher intensity in zone 1 with a flatter gradient compared to other species.

**Cytochrome P450 3A4 (CYP3A4)** intensity gradients were similar in mice and humans, primarily localized in zone 3 and extending into zone 2. In contrast, rats and pigs showed a distinct gradient distribution compared to mice and humans, with higher normalized intensity observed in zones 3 and 2 and extending into zone 1 periportal hepatocytes. Humans exhibited the strongest periportal to perivenous gradient among the species.



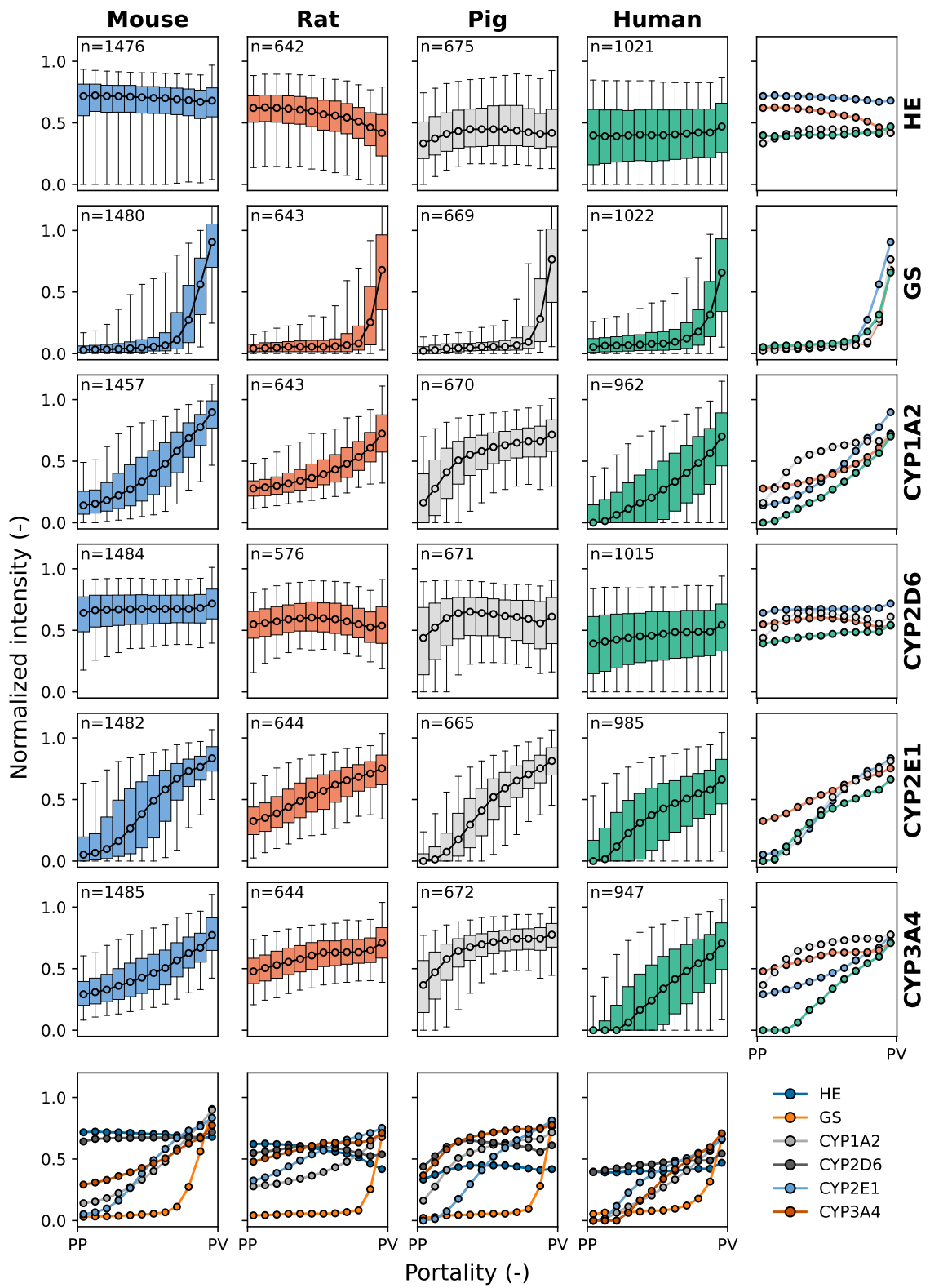


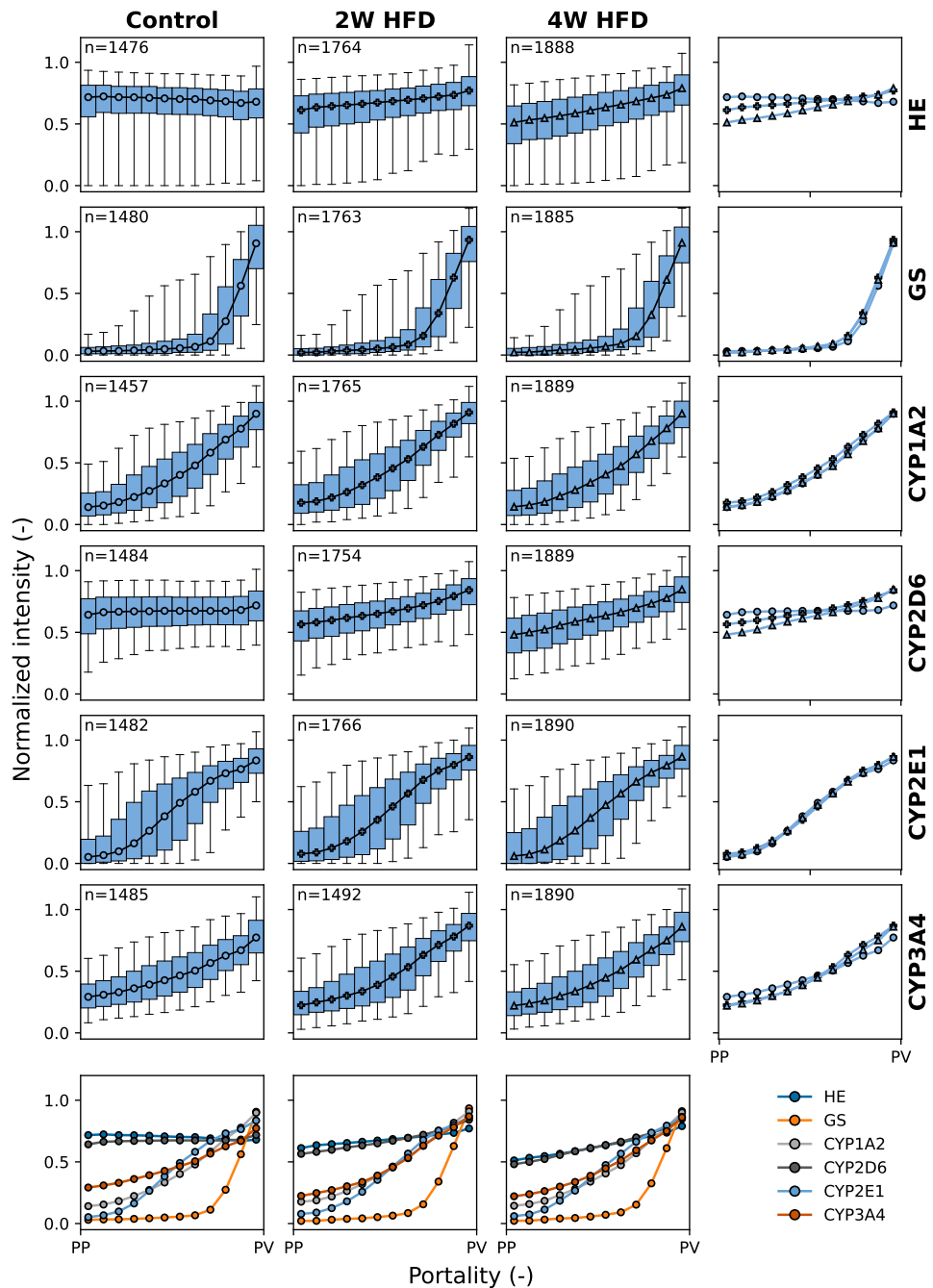
Figure 3.14: Zonated expression of liver proteins

**Figure 3.14 (cont.): Zonated expression of liver proteins:** Zonation patterns of HE, GS, CYP1A2, CYP2D6, CYP2E1, and CYP3A4 were compared among mouse (blue), rat (orange), pig (gray), and human (green) samples. Normalized staining intensity (per slide) was plotted against portality, representing the relative position between periportal (PP) and perivenous (PV) zones within each lobule. Data were divided into 12 bins ranging from PP to PV. Median values were calculated for all lobules across all individuals. Box plots show the median and interquartile range, with whiskers representing the 5th and 95th percentiles. The color coding is as follows: blue for mice, orange for rats, gray for pigs, and green for humans. The number of lobules analyzed for each species is indicated (n). Adapted from Albadry et al. [1].

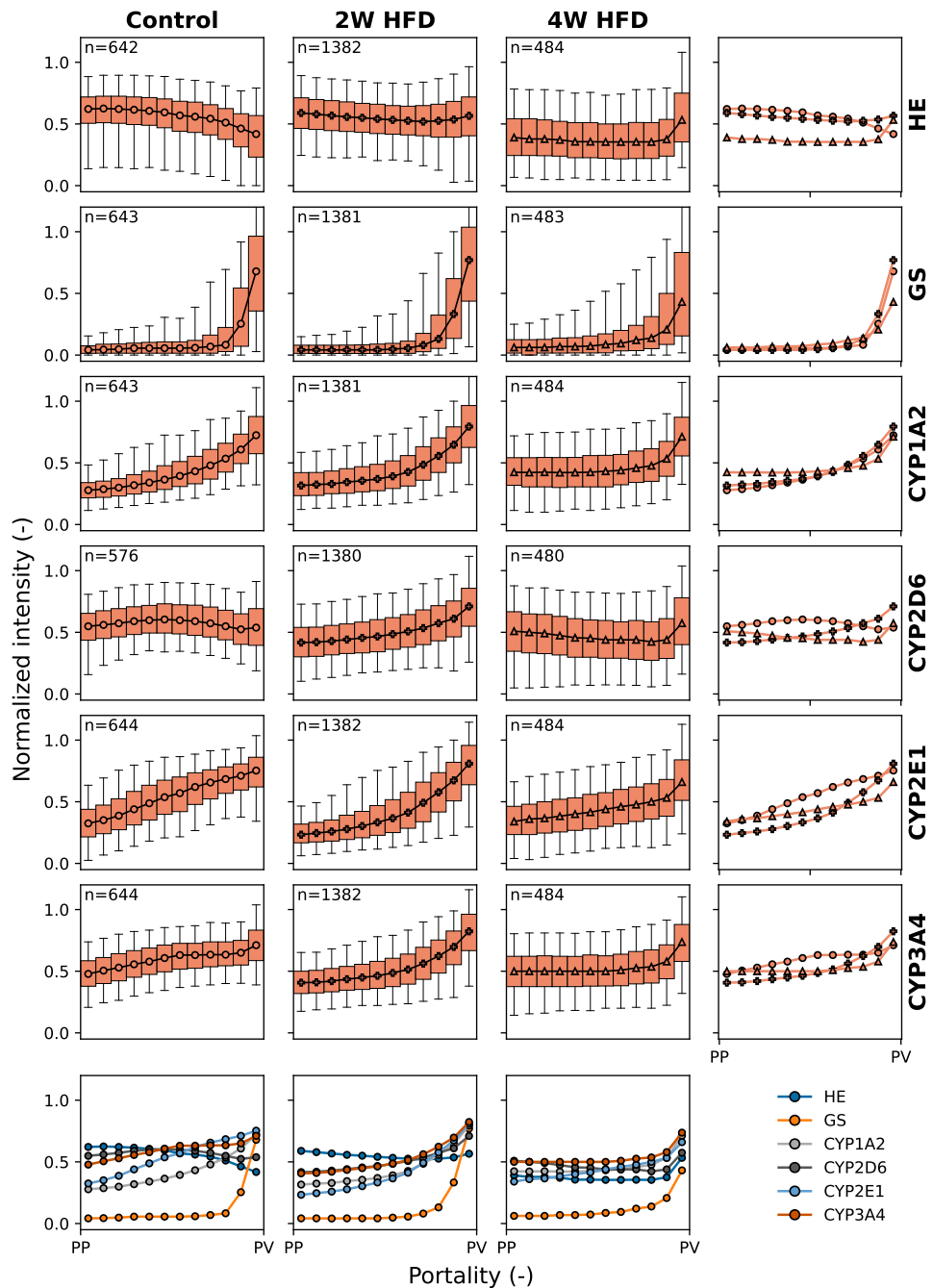
### 3.4.2 Effects of Macrosteatosis on Zonated Protein Expression

The expression gradients for the steatotic samples from the 2W HFD and 4W HFD groups in mice and rats and the steatosis group in humans were analyzed using the same approach as previously described for the control samples. Significant variability in protein signal intensity along the portal-venous axis was observed in all species and groups. In mice, no inter-group differences were detected in the gradients (Figure 3.15). In rats, a decrease in the HE signal was observed in the 4W HFD group compared to the Control and 2W HFD groups. A similar trend was observed for CYPs 2D6, 2E1, and 3A4, where the signal intensity was higher in the Control group than in the 2W HFD and 4W HFD groups.

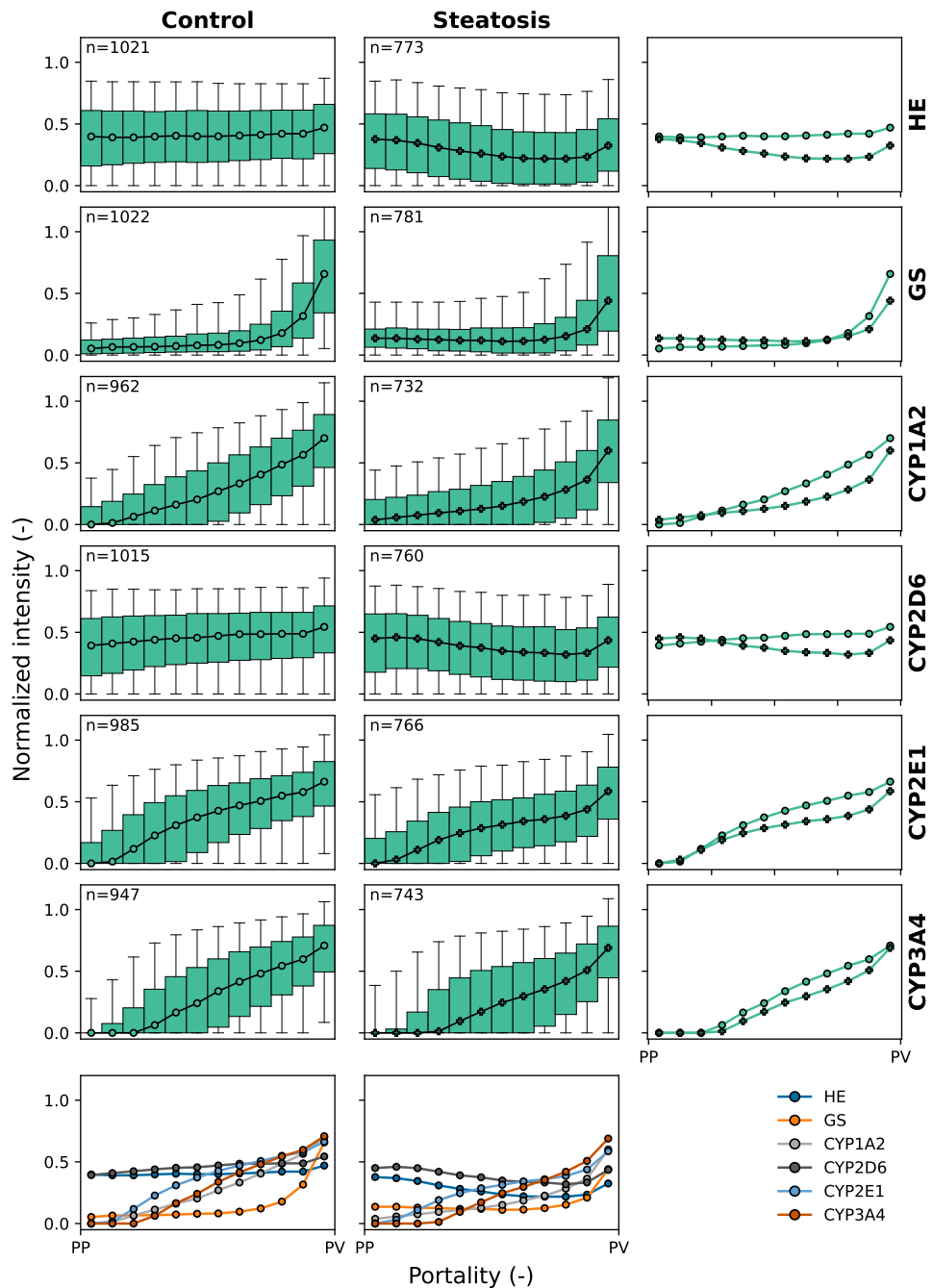
In both rats and humans, the most significant deviations from the control group were observed in the zones with the highest steatosis surface coverage (Figure 3.18). Specifically, the gradient in the rat 4W HFD group exhibited an upward trend in zone 1 (Figure 3.16). In human samples, the greatest deviations were found in zones 1 and 2 (Figure 3.17). These results suggest that protein expression patterns may be slightly altered in regions with high levels of macrosteatosis.



**Figure 3.15: Zonated expression of liver proteins in mouse control and steatosis samples:** Zonation patterns of HE, GS, CYP1A2, CYP2D6, CYP2E1, and CYP3A4 were compared among mouse control, 2W HDF, and 4W HDF samples. Normalized staining intensity (per slide) was plotted against portality, representing the relative position between periportal (PP) and perivenous (PV) zones within each lobule. Data were divided into 12 bins ranging from PP to PV. Median values were calculated for all lobules across all individuals. Box plots show the median and interquartile range, with whiskers representing the 5th and 95th percentiles. The number of lobules analyzed for each species is indicated (n).



**Figure 3.16: Zonated expression of liver proteins in rat control and steatosis samples:** Zonation patterns of HE, GS, CYP1A2, CYP2D6, CYP2E1, and CYP3A4 were compared among rat control, 2W HFD, and 4W HFD samples. Normalized staining intensity (per slide) was plotted against portality, representing the relative position between periportal (PP) and perivenous (PV) zones within each lobule. Data were divided into 12 bins ranging from PP to PV. Median values were calculated for all lobules across all individuals. Box plots show the median and interquartile range, with whiskers representing the 5th and 95th percentiles. The number of lobules analyzed for each species is indicated (n).



**Figure 3.17: Zonated expression of liver proteins in human control and steatosis samples:** Zonation patterns of HE, GS, CYP1A2, CYP2D6, CYP2E1, and CYP3A4 were compared among human control and steatosis samples. Normalized staining intensity (per slide) was plotted against portality, representing the relative position between periportal (PP) and perivenous (PV) zones within each lobule. Data were divided into 12 bins ranging from PP to PV. Median values were calculated for all lobules across all individuals. Box plots show the median and interquartile range, with whiskers representing the 5th and 95th percentiles. The number of lobules analyzed for each species is indicated (n).

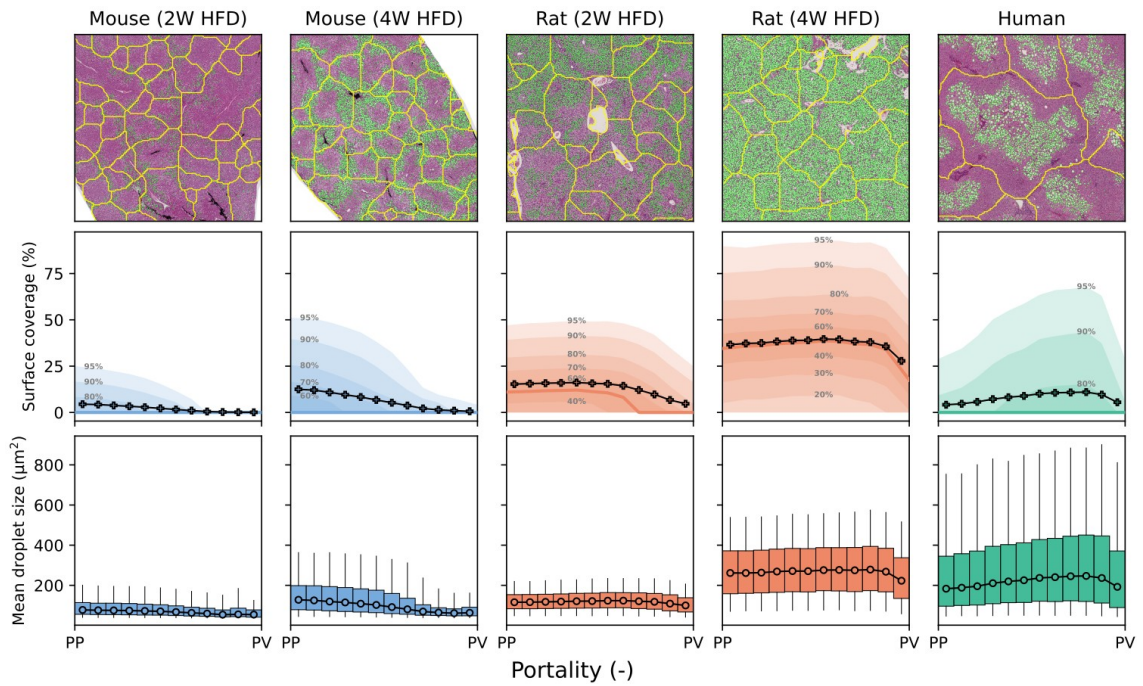
### 3.4.3 Zonated Distribution of Macrosteatosis

The previously described method was employed to detect lobules and calculate the portality for the steatotic slides. The resulting dataset contained a portality value ranging from 0 to 1 for each pixel, with a resolution of level 7, corresponding to a pixel size of  $28.16\ \mu\text{m}$ . The steatosis dataset, which included the positions and geometrical features of the steatosis droplets, was collected at the original resolution level 0, with a pixel size of  $0.22\ \mu\text{m}$ .

To combine these datasets, the high-resolution macrosteatosis data was binned into cells corresponding to the pixel size of the low-resolution portality dataset, and the droplets were mapped based on whether their centroids fell within the corresponding portality pixel. Since multiple droplets could be present within a single pixel of the portality dataset, the total droplet area and the mean droplet area were calculated for each pixel by aggregating the data from the contained droplets using sum and mean, respectively.

The resulting dataset included information about the pixel position, portality, and the mean and total areas occupied by macrosteatosis droplets within each  $28.16\ \mu\text{m} \times 28.16\ \mu\text{m}$  pixel. The total surface coverage was calculated by dividing the total droplet area within a pixel by the area of the pixel. Pixels were then binned by portality into 12 equally sized bins, enabling the visualization of the distribution of surface coverage and mean droplet area along the portal-venous axis. The surface coverages were visualized as percentile bands. Notably, even in highly steatotic samples, a significant number of pixels did not contain any droplets (Figure 3.18). This visualization with percentile bands provides insight into these two parameters.

For the further analysis of macrosteatosis zonation patterns, two quantities are introduced: steatotic fraction (SF) and top surface coverage (TSC). SF represents the percentile band that intersects zero on the surface coverage (Y-) axis, effectively measuring the proportion of pixels containing any steatosis droplets. TSC denotes the 95th percentile of the surface coverage, indicating the highest levels of steatosis within the sample.



**Figure 3.18: Zonated distribution of macrosteatosis:** The portal-venous distribution of macrosteatosis droplets (MS) droplets was analyzed. The first row displays exemplary sections of H&E stained WSI with the lobule boundaries (yellow) and the detected macrosteatosis droplets (green). The second row displays the surface area coverage of MS, calculated as the fraction of the total macrosteatosis droplet area in a pixel and the area of the pixel. The colored bands depict the percentiles for the bins. The third row shows the mean MS droplet area, representing the average droplet area for each pixel. For this plot only pixels containing at least one droplet were considered.

In mice, the 2W HFD group showed the least amount of steatosis, characterized by a periportal pattern. In this group, the steatotic fraction (SF) reached up to 25% and the total surface coverage (TSC) was 25% in the periportal zone. Both measures decreased progressively toward the midzone, resulting in less than 5% of pixels in the perivenous zones containing any macrosteatosis droplets.

Similarly, the 4-week high-fat diet (4W HFD) group also displayed a periportal steatosis pattern. In this group, the SF was 45% and the TSC was 50% in the periportal zone. However, both SF and TSC decreased sharply across the midzone, with the pericentral zone showing an SF of less than 10% and a TSC of less than 5%.

Rats exhibited more homogeneous steatosis patterns in both high-fat diet (HFD) groups compared to mice. These patterns showed higher levels of steatosis in the periportal and midzone, with a reduction in the pericentral zone. In the 2W HFD group, the steatotic fraction (SF) was 65% and remained consistent until the midzone, where it decreased to 30% in the pericentral zone. Total surface coverage (TSC) was 50%, only declining in the pericentral zone to 25%.

In the 4W HFD group, a uniform distribution of macrosteatosis was observed, with a decrease only in the last two pericentral bins. The SF remained above 80% and decreased to 70% in the most pericentral bin. Similarly, the TSC was consistently above 80%, decreasing to 75% in the pericentral bin.

In contrast to the mouse and rat samples, the human samples showed a midzone to perivenous steatosis pattern. In the periportal zone, the steatotic fraction (SF) and total surface coverage (TSC) were as low as 12% and 25%, respectively. Steatosis peaked in the third and fourth pericentral bins, with an SF of approximately 25% and a TSC nearing 70%. It is important to note that the samples analyzed were obtained from subjects with varying degrees of steatosis and varying medical conditions. Consequently, the observed patterns represent an averaged representation of the diverse steatosis distributions present in each patient's sample.



# 4 Discussion

## 4.1 Image Analysis Workflow for Lobule Segmentation

In this study, an image analysis workflow was developed to detect and segment liver lobules on histology slides, eliminating the need for manual annotations by a trained pathologist. The method relies on protein expression gradients derived from multiple stains to identify liver lobules, requiring a set of adjacent, aligned whole slide images (WSIs) stained for different proteins, with expression gradients along the periportal-perivenous axis.

For optimal results, it is crucial that the WSIs are of high quality, with minimal tissue artifacts such as ruptures and folds, and that the staining is homogeneous to avoid variability in signal levels. The precise number of stained slides required and the specific combination of protein gradients necessary for robust segmentation remain areas for further investigation. If proteins exhibit similar expression patterns, the information may be redundant, suggesting that selecting proteins with distinct spatial distributions could suffice, thereby reducing computational demands.

Most previous studies have relied on pathologist annotations to mark central veins and periportal triads. However, such annotations were not available for this study. To validate the segmentation method, I visually compared the results with pig WSIs where physiological septa naturally delineate lobular boundaries. The automatically detected lobules showed excellent agreement with those defined by the septa. However, since the septa appear as areas without protein signal, and the algorithm uses this information to place boundaries, the septa are not an ideal independent reference for validation. For a rigorous assessment, expert annotations of lobules would be necessary.

Protein intensities were normalized and background-corrected on a per-slide basis, ensuring data accuracy and reliability. While this normalization is adequate for inter-species comparisons, it limits the ability to compare absolute protein expression between steatotic and control groups. Given the inherent variability in staining procedures (e.g., differences in lab personnel, protocols, or antibody binding) and the potential non-linear relationship between protein amount and staining intensity, comparing absolute protein values is generally challenging. However, when staining is performed within the same batch, relative differences can be assessed semi-quantitatively.

The intensity analysis was conducted at the lowest resolution level (128-fold downscale of the full-resolution image). The lower resolution levels of the pyramidal images were generated by applying a Gaussian filter repeatedly and then downscaling by a factor of two.

This process may introduce bias in the observed protein intensity, particularly in highly steatotic areas, where substantial white space at higher resolutions could affect the results. Future work should explore alternative downscaling methods to quantify these effects more accurately. Understanding these potential artifacts is essential, especially when interpreting the slight relative decreases observed in expression gradients in highly steatotic regions. An alternative approach could involve retrieving intensities from higher resolution levels and aggregating them into bins matching the size used for the portality map, thereby avoiding the artifacts introduced by downscaling.

## 4.2 Quantification of Macrosteatosis

In addition, this work introduced a hybrid approach to detecting macrosteatosis droplets in whole slide images (WSI). By leveraging transfer learning, a UNET model with a pretrained VGG16 encoder was employed, enabling the development of a robust model despite the limited size of the training dataset. To enhance the accuracy of semantic segmentation, a blob separation postprocessing step was implemented, allowing for precise identification of individual steatosis droplets. The UNET segmentation was designed to operate on image tiles, ensuring that segmentation performance is not constrained by the size of the WSI. This approach also allows for scalability, as the tile size can be adjusted to optimize performance on more powerful GPU systems.

Evaluation metrics indicate that the model performs consistently well on samples with larger steatosis droplets. However, a decrease in performance was observed in samples from the 2-week high-fat diet (2W HFD) mice group, which exhibited smaller droplets and less macrosteatosis. The model was pretrained on the training dataset from [63], where human samples were annotated based on the large droplet fat definition by Neil et al. [53]. Consequently, the model may be biased towards recognizing larger droplets. Additionally, accurately distinguishing individual small droplets within microsteatotic areas can be challenging, potentially leading to inconsistent annotations. To improve the detection of macrosteatosis, it is essential to incorporate more training data specific to the analyzed dataset, with a particular focus on weighting samples containing smaller steatosis droplets.

Model training was conducted on hardware that limited the batch size to 4. Ideally, training on more advanced hardware would allow for larger batch size and the inclusion of more annotated data. Given the small training dataset, an 80/10/10 split for the training, validation, and test datasets was used, which led to the unequal representation of different groups within the test set due to random splitting.

The test data masks were generated semi-automatically. To evaluate the instance segmentation achieved through the blob separation process, the separated clumped steatosis blobs were assessed in the annotation masks using keypoint annotations provided by annotators. The boundary regions obtained through the watershed technique sometimes deviated from physiological regions, potentially impacting validation metrics. Despite these limitations, the partial automation of the image annotation task significantly reduced the workload on annotators. Rather than manually delineating macrosteatosis boundaries, annotators focused on refining the generated keypoints by adding missing ones and removing

incorrect ones.

The combination of lobule segmentation and steatosis detection provided a tool to quantify the zonal distribution of macrosteatosis, enabling a fine-grained analysis of zonal steatosis droplet coverage and size.

In conclusion, while the presented approach for macrosteatosis detection demonstrates robust performance for larger steatosis droplets, the model’s accuracy diminishes for smaller droplets, particularly in samples with low steatosis levels. Future work should prioritize improving annotation consistency, expanding the training dataset, and utilizing more powerful hardware to enhance overall model performance. The integration of lobule detection with steatosis quantification enabled automated zonal evaluation of steatosis, offering valuable insights into its distribution within the liver.

### 4.3 Lobular Geometry

I developed an image analysis pipeline for the prediction of liver lobules in liver histology WSI. This approach can segment a whole tissue section on a slide into liver lobules without the requirement of annotations by a pathologist. I demonstrated that this approach can be employed on normal and steatotic samples. Significant inter-species differences in median lobular size were observed, with the lobule area increasing from mouse, rat, human, to pig. These findings are consistent with those reported by Lau et al. [46], who employed Voronoi tessellation to quantify lobular areas. They reported mean  $\pm$  SD lobule areas of  $(0.69 \pm 0.02) \text{ mm}^2$  for pigs and  $(0.89 \pm 0.51) \text{ mm}^2$  for humans. My results demonstrated comparable mean lobule areas of  $(0.72 \pm 0.57) \text{ mm}^2$  for pigs and a higher mean of  $(0.96 \pm 1.36) \text{ mm}^2$  for humans. Segovia-Miranda et al. [67] analyzed lobule radii by finding the minimum distance between the central vessel and the closest portal tract on a stained WSI. Their median (Q1, Q3) values reported for normal 513 (427, 524)  $\mu\text{m}$  and steatotic human samples (541 (531, 586)  $\mu\text{m}$ ), are in good agreement with the values I found for normal human samples (559 (309, 830)  $\mu\text{m}$ ) and steatotic samples (562 (323, 916)  $\mu\text{m}$ ).

Schwen et al. [66] reported lobular areas in steatotic mice of  $(0.281 \pm 0.157) \text{ mm}^2$ , which is within the ranges of my findings for 2W HFD ( $(0.27 \pm 0.24) \text{ mm}^2$ ) and 4W HFD mice ( $(0.30 \pm 0.29) \text{ mm}^2$ ). For all species, I found larger variability as expressed by standard deviation and interquartile ranges compared to the literature. The distribution of the lobular geometry parameters was highly skewed, seemingly log-normally distributed. The observed high variability may be partly attributed to the varying position, size, and 3-dimensional (3-D) shape of the lobules relative to the 2-D sectioning plane. It is important to note that in this study, all structures identified through the image analysis workflow were included in the final analysis. No manual selection was performed to exclude slides with less distinct lobular patterns or lobules based on the visibility of central vessels and portal triads. This comprehensive approach may have contributed to increased variability in the resulting dataset.

Despite intra-subject variability, no significant inter-subject or inter-lobe differences in lobular geometry were observed, highlighting the robustness of the image analysis workflow presented.

The analysis of lobular geometry in steatotic samples showed a tendency towards increased lobule sizes across species. Although significant changes were only observed in the 4W HFD group of rats compared to the control group, a moderate to strong positive correlation between lobular fat content and lobular size was found, with the strength of this correlation increasing from mice to human samples. These findings are consistent with previously reported results at the lobule level [32, 67] and align with clinically observed liver enlargement in human subjects with liver steatosis [68, 15].

#### 4.4 Zonated Expression of Cytochrome P450 Isoforms

Utilizing the presented image analysis pipeline, the expression patterns of four major CYP isoforms and GS were quantified. These CYP isoforms are heavily involved in drug metabolism and have been subject to pharmacokinetic studies. It has been well established that there are considerable inter-species differences in the expression of these proteins in the liver. However, little emphasis has been laid on studying zonated expression patterns. One study investigated the panlobular expression of CYP3A4 in adult minipigs with the reported pericentral to mid-zonal expression observed in humans [74]. We found comparable expression patterns for CYP2E1, but contrastingly, CYP1A2, CYP2D6, and CYP3A4 showed considerable interspecies differences in terms of the expression pattern, as observed in this study. Additionally, we found slight relative decreases within the protein expression patterns across the control and steatotic groups. However, the normalization of the protein signal does not allow for a quantitative comparison of absolute protein expression between species or groups. [2] did not find significant differences in protein expression between the studied CYP isoforms 3A4, 1A2, 2D6, and 2E1, which indicates that the zonation pattern may not be altered. However, they reported decreased enzymatic activity for CYP3A, and CYP1A and increased activity for CYP2E1. Analyzing the protein signal without normalizing it to the maximum intensity on the slide may provide interesting insights into the zonal up or down-regulation of CYPs in steatotic samples.

Understanding the species differences in zonated expression is crucial to confidently transfer findings from pharmacologic studies in the animal model to the clinical situation. Furthermore, this knowledge is vital for understanding zonated pharmacological liver damage as observed for paracetamol overdose [7] and provides a foundation for data-driven pharmacokinetic modeling of the liver lobules. As discussed in [1], a deeper understanding of the species-specific features of drug metabolism, including differences and similarities in CYP patterns, will facilitate more accurate predictions of therapeutic efficacy and toxicity according to species. This knowledge can also inform the development of safer and more effective therapeutic plans. Furthermore, it may have a significant impact on drug testing and preclinical drug development.

## 5 Outlook

Future work will focus on enhancing the steatosis detection pipeline by systematically annotating additional training data, particularly to improve the model’s performance in detecting smaller steatosis droplets and microsteatosis patches. Since the current droplet segmentation relies on instance segmentation, exploring alternative models such as Mask R-CNN [35] may yield promising results. Mask R-CNN could enhance the segmentation of overlapping instances and potentially reduce the need for manual postprocessing routines that require fine-tuning.

The image analysis pipeline has successfully achieved lobule segmentation in both normal and steatotic immunostained liver WSIs. Because the segmentation algorithm relies on protein signals with monotonically increasing expression along the periportal-perivenous axis, this approach could be adapted for other data sources, such as multi-channel fluorescence images or spatial transcriptomics. Expanding the pipeline to accommodate different imaging modalities would broaden its applicability and validate its effectiveness across various types of data. Additionally, it could facilitate the integration of multimodal data, including transcriptomics and proteomics.

Future work could also extend the pipeline to include additional histological readouts relevant to the evaluation of liver biopsies and histologies, such as necrosis, fibrosis, cellular ballooning, and inflammation, using H&E slides or specific stainings for each readout.

In this thesis, we examined the impact of steatosis on zonation patterns and lobular geometries. Similarly, the effects of other pathophysiological conditions or surgical interventions could be studied to understand the potential zoned up- or down-regulation of protein expression.

While the segmentation of lobules in 2D planes provided valuable insights into their geometric characteristics, this approach has limitations. The sectioning plane may not always be perfectly perpendicular to the lobules, which could affect the accuracy of the analysis. To address this, future work could apply the lobule segmentation workflow to larger stacks of adjacent slides, enabling the reconstruction of the full 3D structure of entire lobules. This would offer a more comprehensive understanding of their longitudinal geometry, which is crucial for interpreting cross-sectional data. Additionally, analyzing the 3D structure could help quantify the variability in lobule geometry and protein gradients introduced by non-perpendicular sectioning planes. This would require the registration of image stacks to extend the presented analysis from 2D to 3D.

The pipeline provided detailed insights into CYP isoform expression and macrosteatosis distribution within lobules. Currently, this information is primarily descriptive. An

important next step would be to leverage this data for quantitative modeling by integrating it with computational models, such as lobular models [59, 26]. By incorporating realistic geometries and protein gradients into these models, this work could advance the development of more accurate simulations of spatially resolved liver metabolism and function. This would facilitate the study of liver zonation effects on pharmacokinetics and liver toxicity in silico.

# Acknowledgements

First and foremost, I want to express my gratitude to my supervisor, Dr. Matthias König, for his invaluable guidance and feedback throughout this project. His support was instrumental in ensuring the successful conduction of my research. I also extend my heartfelt thanks to Prof. Dr. Hans-Peter Herzel for serving as the second reviewer and providing insightful feedback on my thesis.

I would like to thank Mohamed Albadry and Utah Damen for their collaboration as they conducted the staining experiments and provided me with the image data set.

Special thanks go to Lina Schäfer for her meticulous work in annotating the macrosteatosis training dataset, which was essential for my research.

This work was supported by the Federal Ministry of Education and Research (BMBF, Germany) within LiSyM by grant number 031L0054 and ATLAS by grant number 031L0304B and by the German Research Foundation (DFG) within the Research Unit Program FOR 5151 "QuaLiPerF (Quantifying Liver Perfusion-Function Relationship in Complex Resection - A Systems Medicine Approach)" by grant number 436883643 and by grant number 465194077 (Priority Programme SPP 2311, Subproject SimLivA). This work was supported by the BMBF-funded de.NBI Cloud within the German Network for Bioinformatics Infrastructure (de.NBI) (031A537B, 031A533A, 031A538A, 031A533B, 031A535A, 031A537C, 031A534A, 031A532B).

# References

- [1] Mohamed Albadry et al. „Cross-species variability in lobular geometry and cytochrome P450 hepatic zonation: insights into CYP1A2, CYP2D6, CYP2E1 and CYP3A4“. In: *Frontiers in Pharmacology* 15 (2024), p. 1404938. DOI: 10.3389/fphar.2024.1404938.
- [2] Mohamed Albadry et al. „Periportal Steatosis in Mice Affects Distinct Parameters of Pericentral Drug Metabolism“. In: *Scientific Reports* 12.1 (2022), p. 21825. DOI: 10.1038/s41598-022-26483-6.
- [3] Michele Alves-Bezerra and David E. Cohen. „Triglyceride Metabolism in the Liver“. eng. In: *Comprehensive Physiology* 8.1 (2017), pp. 1–8. DOI: 10.1002/cphy.c170012.
- [4] M. Angelico. „Donor liver steatosis and graft selection for liver transplantation: a short review“. eng. In: *European Review for Medical and Pharmacological Sciences* 9.5 (2005), pp. 295–297.
- [5] Andreea Anghel, Milos Stanisavljevic, Sonali Andani, Nikolaos Papandreou, Jan Hendrick Rüschoff, Peter Wild, Maria Gabrani, and Haralampos Pozidis. „A High-Performance System for Robust Stain Normalization of Whole-Slide Images in Histopathology“. In: *Frontiers in Medicine* 6 (2019), p. 193. DOI: 10.3389/fmed.2019.00193.
- [6] Paul Angulo. „Obesity and Nonalcoholic Fatty Liver Disease“. en. In: *Nutrition Reviews* 65 (2008), S57–S63. DOI: 10.1111/j.1753-4887.2007.tb00329.x.
- [7] Irene Anundi, Tuula Lähteenmäki, Mats Rundgren, Peter Moldeus, and Kai O. Lindros. „Zonation of acetaminophen metabolism and cytochrome P450 2E1-mediated toxicity studied in isolated periportal and perivenous hepatocytes“. en. In: *Biochemical Pharmacology* 45.6 (1993), pp. 1251–1259. DOI: 10.1016/0006-2952(93)90277-4.
- [8] Peter Bankhead et al. „QuPath: Open Source Software for Digital Pathology Image Analysis“. In: *Scientific Reports* 7.1 (2017), p. 16878. DOI: 10.1038/s41598-017-17204-5.
- [9] J. Baron, J.A. Redick, and F.P. Guengerich. „An Immunohistochemical Study on the Localization and Distributions of Phenobarbital- and 3-Methylcholanthrene-Inducible Cytochromes P-450 within the Livers of Untreated Rats.“ In: *Journal of Biological Chemistry* 256.11 (1981), pp. 5931–5937. DOI: 10.1016/S0021-9258(19)69299-9.



- [10] Shani Ben-Moshe and Shalev Itzkovitz. „Spatial heterogeneity in the mammalian liver“. eng. In: *Nature Reviews. Gastroenterology & Hepatology* 16.7 (2019), pp. 395–410. DOI: 10.1038/s41575-019-0134-x.
- [11] Thomas D. Boyer, Michael Peter Manns, Arun J. Sanyal, and David Zakim. *Zakim and Boyer’s hepatology: a textbook of liver disease*. eng. 6th ed. Philadelphia, PA: Saunders/Elsevier, 2012.
- [12] Gregory Thomas Brown and David E. Kleiner. „Histopathology of Nonalcoholic Fatty Liver Disease and Nonalcoholic Steatohepatitis“. In: *Metabolism* 65.8 (2016), pp. 1080–1086. DOI: 10.1016/j.metabol.2015.11.008.
- [13] Rolf Bühler, Kai O. Lindros, Åsa Nordling, Inger Johansson, and Magnus Ingelman-Sundberg. „Zonation of Cytochrome *P* 450 Isozyme Expression and Induction in Rat Liver“. In: *European Journal of Biochemistry* 204.1 (1992), pp. 407–412. DOI: 10.1111/j.1432-1033.1992.tb16650.x.
- [14] Alexander Buslaev, Vladimir I. Igloukov, Eugene Khvedchenya, Alex Parinov, Mikhail Druzhinin, and Alexandr A. Kalinin. „Albumentations: Fast and Flexible Image Augmentations“. In: *Information* 11.2 (2020). DOI: 10.3390/info11020125.
- [15] Ji Young Choi, Seung Soo Lee, Na Young Kim, Hyo Jung Park, Yu Sub Sung, Yedaun Lee, Jee Seok Yoon, and Heung-Il Suk. „The effect of hepatic steatosis on liver volume determined by proton density fat fraction and deep learning–measured liver volume“. en. In: *European Radiology* 33.9 (2023), pp. 5924–5932. DOI: 10.1007/s00330-023-09603-2.
- [16] Enoch Cobbina and Fatemeh Akhlaghi. „Non-Alcoholic Fatty Liver Disease (NAFLD) – Pathogenesis, Classification, and Effect on Drug Metabolizing Enzymes and Transporters“. In: *Drug Metabolism Reviews* 49.2 (2017), pp. 197–211. DOI: 10.1080/03602532.2017.1293683.
- [17] Ivan Culjak, David Abram, Tomislav Pribanic, Hrvoje Dzapo, and Mario Cifrek. „A brief introduction to OpenCV“. In: *2012 Proceedings of the 35th International Convention MIPRO*. 2012, pp. 1725–1730.
- [18] Rory P. Cunningham and Natalie Porat-Shliom. „Liver Zonation – Revisiting Old Questions With New Technologies“. In: *Frontiers in Physiology* 12 (2021), p. 732929. DOI: 10.3389/fphys.2021.732929.
- [19] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. „ImageNet: A large-scale hierarchical image database“. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 2009, pp. 248–255. DOI: 10.1109/CVPR.2009.5206848.
- [20] Harshad Devarbhavi, Sumeet K. Asrani, Juan Pablo Arab, Yvonne Ayerki Nartey, Elisa Pose, and Patrick S. Kamath. „Global Burden of Liver Disease: 2023 Update“. In: *Journal of Hepatology* 79.2 (2023), pp. 516–537. DOI: 10.1016/j.jhep.2023.03.017.

- [21] Olive Jean Dunn. „Multiple Comparisons among Means“. In: *Journal of the American Statistical Association* 56.293 (1961), pp. 52–64. DOI: 10.1080/01621459.1961.10482090. eprint: <https://www.tandfonline.com/doi/pdf/10.1080/01621459.1961.10482090>.
- [22] „EASL–EASD–EASO Clinical Practice Guidelines for the Management of Non-Alcoholic Fatty Liver Disease“. In: *Journal of Hepatology* 64.6 (2016), pp. 1388–1402. DOI: 10.1016/j.jhep.2015.11.004.
- [23] Mohammed Eslam, Arun J. Sanyal, Jacob George, and International Consensus Panel. „MAFLD: A Consensus-Driven Proposed Nomenclature for Metabolic Associated Fatty Liver Disease“. In: *Gastroenterology* 158.7 (2020), 1999–2014.e1. DOI: 10.1053/j.gastro.2019.11.312.
- [24] Giovanna Ferraioli and Livia Beatriz Soares Monteiro. „Ultrasound-Based Techniques for the Diagnosis of Liver Steatosis“. In: *World Journal of Gastroenterology* 25.40 (2019), pp. 6053–6062. DOI: 10.3748/wjg.v25.i40.6053.
- [25] Chandler D. Gatensbee et al. „Virtual Alignment of Pathology Image Series for Multi-Gigapixel Whole Slide Images“. In: *Nature Communications* 14.1 (2023), p. 4502. DOI: 10.1038/s41467-023-40218-9.
- [26] Steffen Gerhäuser, Lena Lambers, Luis Mandl, Julian Franquinet, Tim Ricken, and Matthias König. „Simulation of zonation-function relationships in the liver using coupled multiscale models: Application to drug-induced liver injury“. In: *bioRxiv* (2024). DOI: 10.1101/2024.03.26.586870. eprint: <https://www.biorxiv.org/content/early/2024/03/29/2024.03.26.586870.full.pdf>.
- [27] Sean Gillies, Casper van der Wel, Joris Van den Bossche, Mike W. Taves, Joshua Arnott, Brendan C. Ward, et al. *Shapely*. Version 2.0.0. 2022. DOI: 10.5281/zenodo.7428463.
- [28] Cameron Gofton, Yadhavan Upendran, Ming-Hua Zheng, and Jacob George. „MAFLD: How Is It Different from NAFLD?“ In: *Clinical and Molecular Hepatology* 29.Suppl (2023), S17–S31. DOI: 10.3350/cmh.2022.0367.
- [29] Christoph Gohlke. *cgohlke/tiffFile: v2022.5.4*. Version v2022.5.4. 2022. DOI: 10.5281/zenodo.6795861.
- [30] Adam Goode, Benjamin Gilbert, Jan Harkes, Drazen Jukic, and Mahadev Satyanarayanan. „OpenSlide: A vendor-neutral software foundation for digital pathology“. In: *Journal of Pathology Informatics* 4.1 (2013), p. 27. DOI: <https://doi.org/10.4103/2153-3539.119005>.
- [31] P.E. Gooding, J. Chayen, B. Sawyer, and T.F. Slater. „Cytochrome P-450 Distribution in Rat Liver and the Effect of Sodium Phenobarbitone Administration“. In: *Chemico-Biological Interactions* 20.3 (1978), pp. 299–310. DOI: 10.1016/0009-2797(78)90108-4.

- [32] Andrew Hall, Claudia Covelli, Roberta Manuguerra, Tu Vinh Luong, Elena Buzzetti, Emmanuel Tsochatzis, Massimo Pinzani, and Amar Paul Dhillon. „Transaminase abnormalities and adaptations of the liver lobule manifest at specific cut-offs of steatosis“. en. In: *Scientific Reports* 7.1 (2017), p. 40977. DOI: 10.1038/srep40977.
- [33] Hye-Sook Han, Geon Kang, Jun Seok Kim, Byeong Hoon Choi, and Seung-Hoi Koo. „Regulation of glucose metabolism from a liver-centric perspective“. eng. In: *Experimental & Molecular Medicine* 48.3 (2016), e218. DOI: 10.1038/emm.2015.122.
- [34] Charles R. Harris et al. „Array programming with NumPy“. In: *Nature* 585.7825 (2020), pp. 357–362. DOI: 10.1038/s41586-020-2649-2.
- [35] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. „Mask R-CNN“. In: *2017 IEEE International Conference on Computer Vision (ICCV)*. 2017, pp. 2980–2988. DOI: 10.1109/ICCV.2017.322.
- [36] Jack A. Hinson, Dean W. Roberts, and Laura P. James. „Mechanisms of Acetaminophen-Induced Liver Necrosis“. In: *Adverse Drug Reactions*. Ed. by Jack Uetrecht. Vol. 196. Series Title: Handbook of Experimental Pharmacology. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 369–405. DOI: 10.1007/978-3-642-00663-0\_12.
- [37] André Homeyer, Andrea Schenk, Janine Arlt, Uta Dahmen, Olaf Dirsch, and Horst K. Hahn. „Fast and accurate identification of fat droplets in histological images“. eng. In: *Computer Methods and Programs in Biomedicine* 121.2 (2015), pp. 59–65. DOI: 10.1016/j.cmpb.2015.05.009.
- [38] Pavel Iakubovskii. *Segmentation Models*. [https://github.com/qubvel/segmentation\\_models](https://github.com/qubvel/segmentation_models). 2019.
- [39] Kamal Ishak et al. „Histological Grading and Staging of Chronic Hepatitis“. In: *Journal of Hepatology* 22.6 (1995), pp. 696–699. DOI: 10.1016/0168-8278(95)80226-6.
- [40] John G. Jones. „Hepatic glucose and lipid metabolism“. eng. In: *Diabetologia* 59.6 (2016), pp. 1098–1103. DOI: 10.1007/s00125-016-3940-5.
- [41] K. Jungermann. „Metabolic zonation of liver parenchyma“. eng. In: *Seminars in Liver Disease* 8.4 (1988), pp. 329–341. DOI: 10.1055/s-2008-1040554.
- [42] David E. Kleiner et al. „Design and validation of a histological scoring system for nonalcoholic fatty liver disease“. In: *Hepatology* 41.6 (2005), pp. 1313–1321. DOI: <https://doi.org/10.1002/hep.20701>. eprint: <https://aasldpubs.onlinelibrary.wiley.com/doi/pdf/10.1002/hep.20701>.
- [43] Matthias König, Sascha Bulik, and Hermann-Georg Holzhütter. „Quantifying the contribution of the liver to glucose homeostasis: a detailed kinetic model of human hepatic glucose metabolism“. eng. In: *PLoS computational biology* 8.6 (2012), e1002577. DOI: 10.1371/journal.pcbi.1002577.
- [44] William H. Kruskal and W. Allen Wallis. „Use of Ranks in One-Criterion Variance Analysis“. In: *Journal of the American Statistical Association* 47.260 (1952), pp. 583–

621. DOI: 10.1080/01621459.1952.10483441. eprint: <https://www.tandfonline.com/doi/pdf/10.1080/01621459.1952.10483441>.
- [45] H. W. Kuhn. „The Hungarian method for the assignment problem“. In: *Naval Research Logistics Quarterly* 2.1-2 (1955), pp. 83–97. DOI: <https://doi.org/10.1002/nav.3800020109>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/nav.3800020109>.
- [46] C. Lau, B. Kalantari, K. P. Batts, L. D. Ferrell, S. L. Nyberg, R. P. Graham, and Roger K. Moreira. „The Voronoi theory of the normal liver lobular architecture and its applicability in hepatic zonation“. eng. In: *Scientific Reports* 11.1 (2021), p. 9343. DOI: 10.1038/s41598-021-88699-2.
- [47] Marc Macenko, Marc Niethammer, J. S. Marron, David Borland, John T. Woosley, Xiaojun Guan, Charles Schmitt, and Nancy E. Thomas. „A method for normalizing histology slides for quantitative analysis“. In: *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. 2009, pp. 1107–1110. DOI: 10.1109/ISBI.2009.5193250.
- [48] Lena Maier-Hein et al. „Metrics reloaded: recommendations for image analysis validation“. en. In: *Nature Methods* 21.2 (2024), pp. 195–212. DOI: 10.1038/s41592-023-02151-z.
- [49] Rita Manco and Shalev Itzkovitz. „Liver zonation“. eng. In: *Journal of Hepatology* 74.2 (2021), pp. 466–468. DOI: 10.1016/j.jhep.2020.09.003.
- [50] Martín Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015.
- [51] Alistair Miles et al. *zarr-developers/zarr-python: v2.14.2*. Version v2.14.2. 2023. DOI: 10.5281/zenodo.7675396.
- [52] Isabelle D. Munsterman, Merijn van Erp, Gert Weijers, Carolien Bronkhorst, Chris L. de Korte, Joost P. H. Drenth, Jeroen A. W. M. van der Laak, and Eric T. T. L. Tjwa. „A Novel Automatic Digital Algorithm that Accurately Quantifies Steatosis in NAFLD on Histopathological Whole-Slide Images“. eng. In: *Cytometry. Part B, Clinical Cytometry* 96.6 (2019), pp. 521–528. DOI: 10.1002/cyto.b.21790.
- [53] Desley A. H. Neil, Marta Minervini, Maxwell L. Smith, Stefan G. Hubscher, Elizabeth M. Brunt, and A. Jake Demetris. „Banff consensus recommendations for steatosis assessment in donor livers“. en. In: *Hepatology* 75.4 (2022), pp. 1014–1025. DOI: 10.1002/hep.32208.
- [54] Teija Oinonen and O. Kai Lindros. „Zonation of hepatic cytochrome P-450 expression and regulation“. en. In: *Biochemical Journal* 329.1 (1998), pp. 17–35. DOI: 10.1042/bj3290017.
- [55] Nobuyuki Otsu. „A Threshold Selection Method from Gray-Level Histograms“. In: *IEEE Transactions on Systems, Man, and Cybernetics* 9.1 (1979), pp. 62–66. DOI: 10.1109/TSMC.1979.4310076.

- [56] Cédric Peleman et al. „Zonated quantification of immunohistochemistry in normal and steatotic livers“. eng. In: *Virchows Archiv: An International Journal of Pathology* (2023). DOI: 10.1007/s00428-023-03496-8.
- [57] Sjaak Pouwels, Nasser Sakran, Yitka Graham, Angela Leal, Tadeja Pintar, Wah Yang, Radwan Kassir, Rishi Singhal, Kamal Mahawar, and Dharmanand Ramnarain. „Non-Alcoholic Fatty Liver Disease (NAFLD): A Review of Pathophysiology, Clinical Management and Effects of Weight Loss“. In: *BMC Endocrine Disorders* 22.1 (2022), p. 63. DOI: 10.1186/s12902-022-00980-1.
- [58] Slobodan Rendic and F. Peter Guengerich. „Survey of Human Oxidoreductases and Cytochrome P450 Enzymes Involved in the Metabolism of Xenobiotic and Natural Chemicals“. In: *Chemical Research in Toxicology* 28.1 (2015), pp. 38–42. DOI: 10.1021/tx500444e.
- [59] T. Ricken, D. Werner, H. G. Holzhütter, M. König, U. Dahmen, and O. Dirsch. „Modeling function–perfusion behavior in liver lobules including tissue, blood, glucose, lactate and glycogen by use of a coupled two-scale PDE–ODE approach“. en. In: *Biomechanics and Modeling in Mechanobiology* 14.3 (2015), pp. 515–536. DOI: 10.1007/s10237-014-0619-z.
- [60] Joan Rodés. *Textbook of hepatology: from basic science to clinical practice*. eng. 3rd ed. OCLC: 1039165293. Malden, Mass: Blackwell, 2008.
- [61] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. „U-Net: Convolutional Networks for Biomedical Image Segmentation“. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Ed. by Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi. Cham: Springer International Publishing, 2015, pp. 234–241.
- [62] Mousumi Roy, Fusheng Wang, George Teodoro, Miriam B. Vos, Alton Brad Farris, and Jun Kong. „Segmentation of Overlapped Steatosis in Whole-Slide Liver Histopathology Microscopy Images“. eng. In: *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference 2018* (2018), pp. 810–813. DOI: 10.1109/EMBC.2018.8512289.
- [63] Mousumi Roy, Fusheng Wang, Hoang Vo, Dejun Teng, George Teodoro, Alton B. Farris, Eduardo Castillo-Leon, Miriam B. Vos, and Jun Kong. „Deep-learning-based accurate hepatic steatosis quantification for histological assessment of liver biopsies“. eng. In: *Laboratory Investigation; a Journal of Technical Methods and Pathology* 100.10 (2020), pp. 1367–1383. DOI: 10.1038/s41374-020-0463-y.
- [64] Massimo Salvi, Luca Molinaro, Jasna Metovic, Damiano Patrono, Renato Romagnoli, Mauro Papotti, and Filippo Molinari. „Fully automated quantitative assessment of hepatic steatosis in liver transplants“. eng. In: *Computers in Biology and Medicine* 123 (2020), p. 103836. DOI: 10.1016/j.combiomed.2020.103836.

- [65] Geoffrey F. Schau. *HEnorm\_python*. [https://github.com/schaugf/HEnorm\\_python/blob/master/normalizeStaining.py](https://github.com/schaugf/HEnorm_python/blob/master/normalizeStaining.py). 2019.
- [66] Lars Ole Schwen, André Homeyer, Michael Schwier, Uta Dahmen, Olaf Dirsch, Arne Schenk, Lars Kuepfer, Tobias Preusser, and Andrea Schenk. „Zonated quantification of steatosis in an entire mouse liver“. eng. In: *Computers in Biology and Medicine* 73 (2016), pp. 108–118. DOI: 10.1016/j.compbio.2016.04.004.
- [67] Fabián Segovia-Miranda et al. „Three-dimensional spatially resolved geometrical and functional models of human liver tissue reveal new aspects of NAFLD progression“. en. In: *Nature Medicine* 25.12 (2019), pp. 1885–1893. DOI: 10.1038/s41591-019-0660-7.
- [68] Gábor Simon, Veronika Heckmann, Dénes Tóth, Dénes Pauka, Karola Petrus, and Tamás F. Molnár. „The effect of hepatic steatosis and fibrosis on liver weight and dimensions“. en. In: *Legal Medicine* 47 (2020), p. 101781. DOI: 10.1016/j.legalmed.2020.101781.
- [69] Jitka Starekova, Diego Hernando, Perry J. Pickhardt, and Scott B. Reeder. „Quantification of Liver Fat Content with CT and MRI: State of the Art“. In: *Radiology* 301.2 (2021), pp. 250–262. DOI: 10.1148/radiol.2021204288.
- [70] Christiane Stern and Laurent Castera. „Non-Invasive Diagnosis of Hepatic Steatosis“. In: *Hepatology International* 11.1 (2017), pp. 70–78. DOI: 10.1007/s12072-016-9772-z.
- [71] The pandas development team. *pandas-dev/pandas: Pandas*. Version latest. 2020. DOI: 10.5281/zenodo.3509134.
- [72] The Matplotlib Development Team. *Matplotlib: Visualization with Python*. Version v3.9.1. 2024. DOI: 10.5281/zenodo.12652732.
- [73] Maksim Terpilowski. „scikit-posthocs: Pairwise multiple comparison tests in Python“. In: *The Journal of Open Source Software* 4.36 (2019), p. 1169. DOI: 10.21105/joss.01169.
- [74] Els Van Peer, Evy Verbueken, Moayad Saad, Christophe Casteleyn, Chris Van Ginneken, and Steven Van Cruchten. „Ontogeny of CYP3A and P-Glycoprotein in the Liver and the Small Intestine of the Göttingen Minipig: An Immunohistochemical Evaluation“. In: *Basic & Clinical Pharmacology & Toxicology* 114.5 (2014), pp. 387–394. DOI: <https://doi.org/10.1111/bcpt.12173>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/bcpt.12173>.
- [75] Pauli Virtanen et al. „SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python“. In: *Nature Methods* 17 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2.
- [76] Mingzhe Zhao et al. „Cytochrome P450 Enzymes and Drug Metabolism in Humans“. In: *International Journal of Molecular Sciences* 22.23 (2021), p. 12808. DOI: 10.3390/ijms222312808.

# Supplementary material

## S.1 Tables

**Table S1:** Descriptive statistics for lobular geometry by species

Species	Attribute	Unit	n	Median	Q1	Q3	Min	Max	Mean	SD	SE
human	area	mm <sup>2</sup>	1074	0.530	0.158	1.174	0.001	10.652	0.966	1.357	0.029
mouse	area	mm <sup>2</sup>	1530	0.193	0.110	0.350	0.001	4.463	0.299	0.360	0.008
pig	area	mm <sup>2</sup>	698	0.638	0.303	1.008	0.001	3.789	0.718	0.573	0.027
rat	area	mm <sup>2</sup>	669	0.266	0.122	0.554	0.001	9.374	0.468	0.729	0.018
human	compactness	-	1074	0.605	0.518	0.675	0.177	0.880	0.592	0.122	0.018
mouse	compactness	-	1530	0.651	0.585	0.702	0.192	0.850	0.638	0.097	0.016
pig	compactness	-	698	0.620	0.552	0.679	0.199	0.809	0.613	0.097	0.023
rat	compactness	-	669	0.626	0.561	0.690	0.149	0.857	0.620	0.105	0.024
human	min. b. radius	mm	1074	0.559	0.309	0.830	0.033	3.889	0.637	0.473	0.019
mouse	min. b. radius	mm	1530	0.330	0.249	0.442	0.033	2.068	0.375	0.216	0.010
pig	min. b. radius	mm	698	0.581	0.405	0.770	0.033	1.737	0.583	0.285	0.022
rat	min. b. radius	mm	669	0.401	0.263	0.563	0.021	2.488	0.451	0.299	0.017
human	perimeter	mm	1074	3.365	1.841	5.091	0.157	23.206	3.966	3.200	0.121
mouse	perimeter	mm	1530	1.939	1.436	2.636	0.157	16.155	2.233	1.397	0.057
pig	perimeter	mm	698	3.561	2.403	4.695	0.157	11.024	3.562	1.823	0.135
rat	perimeter	mm	669	2.345	1.534	3.407	0.116	20.607	2.734	2.052	0.106

**Table S2:** Descriptive statistics for lobular geometry for steatotic samples by species and groups

Species	Group	Attribute	Unit	n	Median	Q1	Q3	Min	Max	Mean	SD
human	Steatosis	area	mm <sup>2</sup>	810	0.554	0.162	1.446	0.001	24.909	1.293	2.323
mouse	2W HDF	area	mm <sup>2</sup>	1817	0.204	0.124	0.327	0.001	2.263	0.266	0.240
mouse	4W HDF	area	mm <sup>2</sup>	1951	0.228	0.132	0.360	0.000	4.838	0.295	0.287
rat	2W HDF	area	mm <sup>2</sup>	1427	0.303	0.162	0.532	0.001	15.098	0.438	0.607
rat	4W HDF	area	mm <sup>2</sup>	502	0.315	0.140	0.605	0.001	31.144	0.549	1.497
human	Steatosis	compactness	-	810	0.615	0.535	0.683	0.161	0.913	0.601	0.123
mouse	2W HDF	compactness	-	1817	0.656	0.594	0.707	0.244	0.884	0.645	0.090
mouse	4W HDF	compactness	-	1951	0.646	0.584	0.703	0.248	0.866	0.639	0.091
rat	2W HDF	compactness	-	1427	0.635	0.567	0.692	0.233	0.851	0.626	0.096
rat	4W HDF	compactness	-	502	0.628	0.563	0.682	0.293	0.837	0.618	0.098
human	Steatosis	min. b. radius	mm	810	0.562	0.323	0.916	0.033	4.900	0.714	0.619
mouse	2W HDF	min. b. radius	mm	1817	0.339	0.260	0.432	0.033	1.387	0.363	0.169
mouse	4W HDF	min. b. radius	mm	1951	0.361	0.271	0.454	0.021	1.978	0.382	0.183
rat	2W HDF	min. b. radius	mm	1427	0.417	0.304	0.555	0.033	3.498	0.453	0.252
rat	4W HDF	min. b. radius	mm	502	0.417	0.287	0.590	0.033	4.888	0.480	0.341
human	Steatosis	perimeter	mm	810	3.349	1.824	5.566	0.157	42.275	4.430	4.349
mouse	2W HDF	perimeter	mm	1817	1.999	1.527	2.568	0.157	8.144	2.141	1.031
mouse	4W HDF	perimeter	mm	1951	2.096	1.592	2.703	0.099	14.123	2.263	1.148
rat	2W HDF	perimeter	mm	1427	2.455	1.775	3.305	0.157	28.112	2.720	1.672
rat	4W HDF	perimeter	mm	502	2.515	1.677	3.557	0.157	34.434	2.909	2.314

**Table S3:** Descriptive statistics for lobular geometry by subject in humans

Subject	Attribute	Unit	n	Median	Q1	Q3	Min	Max	Mean	SD
UKJ-19-010	area	mm <sup>2</sup>	196	0.534	0.216	0.861	0.001	10.652	0.838	1.265
UKJ-19-026	area	mm <sup>2</sup>	154	0.729	0.333	1.253	0.001	5.047	0.879	0.754
UKJ-19-033	area	mm <sup>2</sup>	134	0.504	0.209	1.107	0.001	7.868	0.976	1.375
UKJ-19-036	area	mm <sup>2</sup>	107	0.555	0.050	2.434	0.001	8.035	1.657	2.218
UKJ-19-041	area	mm <sup>2</sup>	326	0.438	0.086	1.223	0.002	9.691	0.943	1.350
UKJ-19-049	area	mm <sup>2</sup>	157	0.453	0.157	0.885	0.003	5.399	0.782	0.954
UKJ-19-010	compactness	-	196	0.614	0.538	0.677	0.198	0.800	0.601	0.116
UKJ-19-026	compactness	-	154	0.617	0.513	0.663	0.285	0.800	0.593	0.098
UKJ-19-033	compactness	-	134	0.605	0.540	0.675	0.244	0.809	0.594	0.111
UKJ-19-036	compactness	-	107	0.598	0.510	0.687	0.316	0.857	0.589	0.129
UKJ-19-041	compactness	-	326	0.596	0.510	0.686	0.194	0.880	0.589	0.134
UKJ-19-049	compactness	-	157	0.604	0.512	0.663	0.177	0.820	0.583	0.126
UKJ-19-010	min. b. radius	mm	196	0.548	0.362	0.710	0.033	3.139	0.609	0.428
UKJ-19-026	min. b. radius	mm	154	0.628	0.434	0.846	0.033	1.963	0.652	0.334
UKJ-19-033	min. b. radius	mm	134	0.569	0.356	0.818	0.033	2.398	0.655	0.456
UKJ-19-036	min. b. radius	mm	107	0.593	0.174	1.194	0.033	2.780	0.765	0.662
UKJ-19-041	min. b. radius	mm	326	0.534	0.219	0.859	0.034	3.889	0.619	0.513
UKJ-19-049	min. b. radius	mm	157	0.497	0.298	0.771	0.041	1.946	0.588	0.394
UKJ-19-010	perimeter	mm	196	3.225	2.051	4.213	0.157	21.609	3.741	2.894
UKJ-19-026	perimeter	mm	154	3.812	2.703	5.130	0.157	12.986	4.000	2.142
UKJ-19-033	perimeter	mm	134	3.345	2.039	4.943	0.157	16.955	4.017	3.076
UKJ-19-036	perimeter	mm	107	3.451	1.065	7.934	0.157	17.593	4.940	4.560
UKJ-19-041	perimeter	mm	326	3.204	1.247	5.261	0.182	23.206	3.873	3.429
UKJ-19-049	perimeter	mm	157	3.092	1.876	4.889	0.216	16.830	3.701	2.760

**Table S4:** Descriptive statistics for lobular geometry by subject in mice

Subject	Attribute	Unit	n	Median	Q1	Q3	Min	Max	Mean	SD
MNT-021	area	mm <sup>2</sup>	307	0.207	0.113	0.356	0.001	2.834	0.301	0.336
MNT-022	area	mm <sup>2</sup>	313	0.183	0.111	0.311	0.001	1.357	0.250	0.223
MNT-023	area	mm <sup>2</sup>	265	0.198	0.113	0.346	0.002	4.463	0.300	0.374
MNT-024	area	mm <sup>2</sup>	212	0.183	0.103	0.354	0.001	4.117	0.325	0.469
MNT-025	area	mm <sup>2</sup>	275	0.187	0.103	0.328	0.001	1.924	0.262	0.262
MNT-026	area	mm <sup>2</sup>	158	0.230	0.118	0.516	0.001	3.608	0.421	0.516
MNT-021	compactness	-	307	0.644	0.577	0.696	0.276	0.850	0.630	0.102
MNT-022	compactness	-	313	0.654	0.592	0.708	0.275	0.830	0.645	0.089
MNT-023	compactness	-	265	0.653	0.589	0.708	0.249	0.809	0.640	0.097
MNT-024	compactness	-	212	0.664	0.596	0.707	0.192	0.821	0.644	0.101
MNT-025	compactness	-	275	0.649	0.592	0.698	0.348	0.825	0.638	0.092
MNT-026	compactness	-	158	0.636	0.571	0.702	0.296	0.810	0.625	0.106
MNT-021	min. b. radius	mm	307	0.340	0.250	0.445	0.033	1.532	0.380	0.210
MNT-022	min. b. radius	mm	313	0.318	0.254	0.418	0.033	1.223	0.351	0.168
MNT-023	min. b. radius	mm	265	0.335	0.262	0.447	0.034	2.068	0.380	0.218
MNT-024	min. b. radius	mm	212	0.320	0.237	0.449	0.033	1.771	0.378	0.246
MNT-025	min. b. radius	mm	275	0.325	0.242	0.411	0.033	1.262	0.352	0.184
MNT-026	min. b. radius	mm	158	0.357	0.252	0.565	0.033	1.744	0.437	0.288
MNT-021	perimeter	mm	307	1.979	1.454	2.713	0.157	10.911	2.273	1.371
MNT-022	perimeter	mm	313	1.866	1.444	2.469	0.157	7.874	2.072	1.026
MNT-023	perimeter	mm	265	1.989	1.441	2.711	0.182	15.014	2.243	1.402
MNT-024	perimeter	mm	212	1.878	1.374	2.662	0.157	16.155	2.276	1.710
MNT-025	perimeter	mm	275	1.907	1.385	2.477	0.157	7.046	2.089	1.131
MNT-026	perimeter	mm	158	2.113	1.517	3.363	0.157	11.176	2.645	1.859



**Table S5:** Descriptive statistics for lobular geometry by subject in rats

Subject	Attribute	Unit	n	Median	Q1	Q3	Min	Max	Mean	SD
NOR-021	area	mm <sup>2</sup>	96	0.382	0.180	0.619	0.003	4.007	0.507	0.563
NOR-022	area	mm <sup>2</sup>	94	0.230	0.108	0.522	0.002	9.374	0.461	1.007
NOR-023	area	mm <sup>2</sup>	92	0.271	0.101	0.554	0.001	7.836	0.507	0.926
NOR-024	area	mm <sup>2</sup>	116	0.215	0.112	0.513	0.003	6.144	0.531	0.904
NOR-025	area	mm <sup>2</sup>	115	0.275	0.136	0.545	0.001	2.883	0.478	0.566
NOR-026	area	mm <sup>2</sup>	156	0.272	0.142	0.546	0.001	1.953	0.369	0.334
NOR-021	compactness	-	96	0.627	0.546	0.694	0.350	0.816	0.624	0.093
NOR-022	compactness	-	94	0.605	0.524	0.683	0.308	0.809	0.604	0.111
NOR-023	compactness	-	92	0.635	0.565	0.693	0.264	0.813	0.623	0.102
NOR-024	compactness	-	116	0.634	0.570	0.689	0.149	0.800	0.620	0.118
NOR-025	compactness	-	115	0.619	0.561	0.684	0.310	0.786	0.613	0.097
NOR-026	compactness	-	156	0.637	0.567	0.691	0.187	0.857	0.628	0.104
NOR-021	min. b. radius	mm	96	0.470	0.331	0.619	0.041	1.553	0.486	0.255
NOR-022	min. b. radius	mm	94	0.363	0.246	0.578	0.034	2.488	0.433	0.320
NOR-023	min. b. radius	mm	92	0.394	0.232	0.550	0.021	2.369	0.446	0.329
NOR-024	min. b. radius	mm	116	0.352	0.254	0.538	0.041	2.242	0.471	0.391
NOR-025	min. b. radius	mm	115	0.409	0.285	0.573	0.021	1.412	0.467	0.270
NOR-026	min. b. radius	mm	156	0.411	0.277	0.538	0.033	1.121	0.415	0.217
NOR-021	perimeter	mm	96	2.865	1.985	3.565	0.199	10.261	2.910	1.612
NOR-022	perimeter	mm	94	2.230	1.424	3.461	0.182	18.507	2.636	2.241
NOR-023	perimeter	mm	92	2.357	1.286	3.385	0.116	17.324	2.734	2.266
NOR-024	perimeter	mm	116	2.047	1.454	3.217	0.216	20.607	2.890	2.822
NOR-025	perimeter	mm	115	2.386	1.638	3.444	0.116	10.275	2.840	1.808
NOR-026	perimeter	mm	156	2.388	1.674	3.283	0.157	8.192	2.493	1.402

**Table S6:** Descriptive statistics for lobular geometry by subject in pigs

Subject	Attribute	Unit	n	Median	Q1	Q3	Min	Max	Mean	SD
SSES2021 10	area	mm <sup>2</sup>	180	0.546	0.241	0.967	0.001	3.329	0.664	0.547
SSES2021 12	area	mm <sup>2</sup>	193	0.577	0.266	0.995	0.002	2.360	0.680	0.540
SSES2021 14	area	mm <sup>2</sup>	219	0.708	0.360	1.084	0.001	2.826	0.808	0.609
SSES2021 9	area	mm <sup>2</sup>	106	0.566	0.317	0.869	0.001	3.789	0.696	0.573
SSES2021 10	compactness	-	180	0.609	0.536	0.680	0.333	0.809	0.608	0.103
SSES2021 12	compactness	-	193	0.621	0.551	0.678	0.199	0.809	0.616	0.095
SSES2021 14	compactness	-	219	0.630	0.560	0.679	0.273	0.809	0.617	0.094
SSES2021 9	compactness	-	106	0.623	0.565	0.677	0.298	0.800	0.609	0.096
SSES2021 10	min. b. radius	mm	180	0.554	0.384	0.757	0.033	1.737	0.557	0.280
SSES2021 12	min. b. radius	mm	193	0.563	0.377	0.771	0.034	1.201	0.560	0.278
SSES2021 14	min. b. radius	mm	219	0.625	0.451	0.796	0.033	1.472	0.620	0.293
SSES2021 9	min. b. radius	mm	106	0.566	0.425	0.733	0.033	1.691	0.594	0.280
SSES2021 10	perimeter	mm	180	3.384	2.317	4.764	0.157	10.466	3.440	1.811
SSES2021 12	perimeter	mm	193	3.417	2.239	4.591	0.182	8.473	3.418	1.787
SSES2021 14	perimeter	mm	219	3.849	2.643	4.789	0.157	8.919	3.779	1.862
SSES2021 9	perimeter	mm	106	3.273	2.612	4.564	0.157	11.024	3.584	1.784

**Table S7:** Descriptive statistics for lobular geometry by mouse lobes in Subject MNT-021

Lobe	Attribute	Unit	n	Median	Q1	Q3	Min	Max	Mean	SD
CL	area	mm <sup>2</sup>	80	0.197	0.096	0.291	0.001	1.847	0.247	0.265
LLL	area	mm <sup>2</sup>	96	0.232	0.120	0.401	0.002	2.834	0.358	0.428
ML	area	mm <sup>2</sup>	69	0.223	0.129	0.353	0.003	1.406	0.299	0.282
RL	area	mm <sup>2</sup>	62	0.182	0.111	0.330	0.001	1.809	0.287	0.294
CL	compactness	-	80	0.659	0.587	0.694	0.304	0.800	0.641	0.089
LLL	compactness	-	96	0.621	0.569	0.686	0.276	0.850	0.613	0.111
ML	compactness	-	69	0.663	0.595	0.715	0.393	0.799	0.653	0.088
RL	compactness	-	62	0.635	0.566	0.691	0.344	0.823	0.617	0.110
CL	min. b. radius	mm	80	0.332	0.236	0.397	0.033	1.313	0.343	0.187
LLL	min. b. radius	mm	96	0.345	0.257	0.524	0.034	1.532	0.415	0.243
ML	min. b. radius	mm	69	0.347	0.261	0.443	0.041	1.067	0.378	0.190
RL	min. b. radius	mm	62	0.322	0.251	0.442	0.033	0.984	0.374	0.196
CL	perimeter	mm	80	1.924	1.340	2.342	0.157	8.638	2.023	1.207
LLL	perimeter	mm	96	2.071	1.478	2.944	0.182	10.911	2.517	1.661
ML	perimeter	mm	69	1.979	1.568	2.602	0.233	6.382	2.245	1.171
RL	perimeter	mm	62	1.878	1.457	2.715	0.157	6.631	2.248	1.207

**Table S8:** Descriptive statistics for lobular geometry by mouse lobes in Subject MNT-022

Lobe	Attribute	Unit	n	Median	Q1	Q3	Min	Max	Mean	SD
CL	area	mm <sup>2</sup>	51	0.157	0.100	0.246	0.001	0.677	0.198	0.145
LLL	area	mm <sup>2</sup>	119	0.194	0.109	0.368	0.003	1.357	0.270	0.247
ML	area	mm <sup>2</sup>	86	0.180	0.111	0.278	0.001	0.783	0.228	0.172
RL	area	mm <sup>2</sup>	57	0.208	0.129	0.320	0.003	1.190	0.291	0.277
CL	compactness	-	51	0.655	0.601	0.694	0.452	0.798	0.646	0.073
LLL	compactness	-	119	0.655	0.594	0.716	0.275	0.830	0.646	0.095
ML	compactness	-	86	0.658	0.606	0.710	0.419	0.823	0.652	0.081
RL	compactness	-	57	0.641	0.573	0.690	0.391	0.795	0.629	0.097
CL	min. b. radius	mm	51	0.303	0.239	0.402	0.033	0.592	0.317	0.119
LLL	min. b. radius	mm	119	0.329	0.256	0.436	0.041	1.223	0.364	0.189
ML	min. b. radius	mm	86	0.311	0.256	0.384	0.033	0.801	0.337	0.141
RL	min. b. radius	mm	57	0.332	0.273	0.434	0.041	0.904	0.376	0.186
CL	perimeter	mm	51	1.773	1.397	2.326	0.157	3.656	1.846	0.725
LLL	perimeter	mm	119	1.979	1.427	2.600	0.199	7.874	2.149	1.152
ML	perimeter	mm	86	1.840	1.454	2.262	0.157	4.810	1.991	0.857
RL	perimeter	mm	57	1.989	1.558	2.537	0.233	5.071	2.238	1.158

**Table S9:** Descriptive statistics for lobular geometry by mouse lobes in Subject MNT-023

Lobe	Attribute	Unit	n	Median	Q1	Q3	Min	Max	Mean	SD
CL	area	mm <sup>2</sup>	51	0.157	0.100	0.246	0.001	0.677	0.198	0.145
LLL	area	mm <sup>2</sup>	119	0.194	0.109	0.368	0.003	1.357	0.270	0.247
ML	area	mm <sup>2</sup>	86	0.180	0.111	0.278	0.001	0.783	0.228	0.172
RL	area	mm <sup>2</sup>	57	0.208	0.129	0.320	0.003	1.190	0.291	0.277
CL	compactness	-	51	0.655	0.601	0.694	0.452	0.798	0.646	0.073
LLL	compactness	-	119	0.655	0.594	0.716	0.275	0.830	0.646	0.095
ML	compactness	-	86	0.658	0.606	0.710	0.419	0.823	0.652	0.081
RL	compactness	-	57	0.641	0.573	0.690	0.391	0.795	0.629	0.097
CL	min. b. radius	mm	51	0.303	0.239	0.402	0.033	0.592	0.317	0.119
LLL	min. b. radius	mm	119	0.329	0.256	0.436	0.041	1.223	0.364	0.189
ML	min. b. radius	mm	86	0.311	0.256	0.384	0.033	0.801	0.337	0.141
RL	min. b. radius	mm	57	0.332	0.273	0.434	0.041	0.904	0.376	0.186
CL	perimeter	mm	51	1.773	1.397	2.326	0.157	3.656	1.846	0.725
LLL	perimeter	mm	119	1.979	1.427	2.600	0.199	7.874	2.149	1.152
ML	perimeter	mm	86	1.840	1.454	2.262	0.157	4.810	1.991	0.857
RL	perimeter	mm	57	1.989	1.558	2.537	0.233	5.071	2.238	1.158

**Table S10:** Descriptive statistics for lobular geometry by mouse lobes in Subject MNT-024

Lobe	Attribute	Unit	n	Median	Q1	Q3	Min	Max	Mean	SD
CL	area	mm <sup>2</sup>	32	0.175	0.087	0.249	0.003	0.454	0.181	0.119
LLL	area	mm <sup>2</sup>	83	0.165	0.105	0.294	0.003	2.305	0.254	0.312
ML	area	mm <sup>2</sup>	43	0.264	0.114	0.464	0.001	4.117	0.464	0.705
RL	area	mm <sup>2</sup>	54	0.236	0.104	0.465	0.001	2.536	0.410	0.516
CL	compactness	-	32	0.666	0.596	0.686	0.448	0.800	0.643	0.081
LLL	compactness	-	83	0.658	0.608	0.708	0.403	0.815	0.649	0.087
ML	compactness	-	43	0.648	0.572	0.690	0.192	0.821	0.624	0.117
RL	compactness	-	54	0.674	0.619	0.725	0.347	0.812	0.651	0.113
CL	min. b. radius	mm	32	0.315	0.239	0.350	0.041	0.521	0.301	0.111
LLL	min. b. radius	mm	83	0.307	0.237	0.406	0.041	1.040	0.341	0.168
ML	min. b. radius	mm	43	0.362	0.246	0.507	0.033	1.771	0.436	0.316
RL	min. b. radius	mm	54	0.357	0.239	0.504	0.033	1.495	0.434	0.309
CL	perimeter	mm	32	1.837	1.330	2.119	0.216	3.266	1.773	0.689
LLL	perimeter	mm	83	1.781	1.371	2.448	0.233	7.676	2.037	1.116
ML	perimeter	mm	43	2.105	1.513	3.097	0.157	16.155	2.759	2.566
RL	perimeter	mm	54	2.154	1.354	3.004	0.157	9.100	2.558	1.880

**Table S11:** Descriptive statistics for lobular geometry by mouse lobes in Subject MNT-025

Lobe	Attribute	Unit	n	Median	Q1	Q3	Min	Max	Mean	SD
CL	area	mm <sup>2</sup>	44	0.157	0.094	0.257	0.003	0.770	0.196	0.160
LLL	area	mm <sup>2</sup>	103	0.190	0.098	0.347	0.001	1.924	0.286	0.313
ML	area	mm <sup>2</sup>	74	0.221	0.129	0.342	0.003	1.455	0.301	0.277
RL	area	mm <sup>2</sup>	54	0.165	0.105	0.276	0.001	0.908	0.215	0.170
CL	compactness	-	44	0.641	0.573	0.705	0.398	0.825	0.636	0.094
LLL	compactness	-	103	0.643	0.571	0.689	0.353	0.821	0.624	0.099
ML	compactness	-	74	0.649	0.590	0.705	0.348	0.809	0.643	0.090
RL	compactness	-	54	0.671	0.622	0.708	0.478	0.800	0.659	0.071
CL	min. b. radius	mm	44	0.310	0.217	0.377	0.041	0.734	0.308	0.136
LLL	min. b. radius	mm	103	0.332	0.239	0.438	0.033	1.262	0.364	0.209
ML	min. b. radius	mm	74	0.341	0.270	0.435	0.041	1.069	0.381	0.197
RL	min. b. radius	mm	54	0.307	0.237	0.390	0.033	0.812	0.327	0.136
CL	perimeter	mm	44	1.895	1.302	2.260	0.216	4.238	1.817	0.815
LLL	perimeter	mm	103	1.941	1.371	2.572	0.157	7.046	2.181	1.300
ML	perimeter	mm	74	2.080	1.568	2.585	0.233	6.382	2.265	1.185
RL	perimeter	mm	54	1.783	1.449	2.274	0.157	4.745	1.894	0.815

**Table S12:** Descriptive statistics for lobular geometry by mouse lobes in Subject MNT-026

Lobe	Attribute	Unit	n	Median	Q1	Q3	Min	Max	Mean	SD
CL	area	mm <sup>2</sup>	11	0.238	0.082	0.570	0.030	0.948	0.349	0.331
LLL	area	mm <sup>2</sup>	67	0.236	0.135	0.489	0.001	2.414	0.416	0.484
ML	area	mm <sup>2</sup>	43	0.230	0.137	0.566	0.003	3.608	0.477	0.641
RL	area	mm <sup>2</sup>	37	0.206	0.089	0.408	0.003	1.815	0.387	0.440
CL	compactness	-	11	0.605	0.548	0.711	0.473	0.757	0.625	0.094
LLL	compactness	-	67	0.644	0.576	0.694	0.296	0.797	0.626	0.104
ML	compactness	-	43	0.635	0.586	0.709	0.301	0.810	0.627	0.118
RL	compactness	-	37	0.611	0.557	0.694	0.414	0.806	0.620	0.097
CL	min. b. radius	mm	11	0.344	0.202	0.657	0.130	0.898	0.421	0.266
LLL	min. b. radius	mm	67	0.365	0.260	0.540	0.033	1.504	0.442	0.275
ML	min. b. radius	mm	43	0.359	0.273	0.573	0.041	1.744	0.461	0.333
RL	min. b. radius	mm	37	0.322	0.240	0.471	0.041	1.033	0.406	0.258
CL	perimeter	mm	11	1.986	1.195	3.699	0.804	5.020	2.400	1.447
LLL	perimeter	mm	67	2.171	1.525	3.203	0.157	8.771	2.664	1.785
ML	perimeter	mm	43	1.996	1.558	3.457	0.216	11.176	2.826	2.215
RL	perimeter	mm	37	2.061	1.517	2.937	0.216	6.580	2.473	1.605

**Table S13:** Descriptive statistics for macrosteatosis droplet geometry by species and groups

Species	Group	Attribute	Unit	n	Median	Q1	Q3	Min	Max	Mean	SD
human	Steatosis	area	$\mu\text{m}^2$	303109	209.8	102.4	400.6	36.3	1790.2	292.7	257.8
mouse	2W HDF	area	$\mu\text{m}^2$	152116	71.9	50.8	109.8	36.3	1434.6	89.8	56.0
mouse	4W HDF	area	$\mu\text{m}^2$	334879	104.1	63.2	182.5	36.3	1236.8	139.4	105.8
rat	2W HDF	area	$\mu\text{m}^2$	782376	116.0	75.2	162.3	36.3	1481.7	124.5	62.4
rat	4W HDF	area	$\mu\text{m}^2$	418463	246.5	131.7	375.0	36.3	1487.5	265.9	159.1
human	Steatosis	min. b. radius	$\mu\text{m}$	303109	9.5	6.7	12.9	3.7	25.0	10.2	4.4
mouse	2W HDF	min. b. radius	$\mu\text{m}$	152116	5.6	4.7	6.9	3.7	25.0	6.0	1.9
mouse	4W HDF	min. b. radius	$\mu\text{m}$	334879	6.8	5.3	9.1	3.7	25.0	7.5	2.8
rat	2W HDF	min. b. radius	$\mu\text{m}$	782376	7.0	5.7	8.2	2.9	25.0	7.1	1.8
rat	4W HDF	min. b. radius	$\mu\text{m}$	418463	10.4	7.6	12.7	3.7	25.0	10.3	3.2
human	Steatosis	perimeter	$\mu\text{m}$	303109	56.1	39.3	77.3	22.3	262.6	60.6	26.4
mouse	2W HDF	perimeter	$\mu\text{m}$	152116	32.5	27.4	40.6	22.3	166.2	35.3	10.4
mouse	4W HDF	perimeter	$\mu\text{m}$	334879	39.8	30.6	52.4	22.3	155.0	43.2	15.6
rat	2W HDF	perimeter	$\mu\text{m}$	782376	41.3	33.6	48.8	22.3	157.7	41.7	10.6
rat	4W HDF	perimeter	$\mu\text{m}$	418463	60.9	44.7	74.7	22.3	173.4	60.1	19.2

**Table S14:** Descriptive statistics for macrosteatosis droplet density and surface coverage per WSI

Species	Group	Attribute	Unit	n	Median	Q1	Q3	Min	Max	Mean	SD
human	Steatosis	Av. droplet density	$\text{mm}^{-2}$	6	294.2	225.4	305.4	199.6	313.4	268.8	52.6
mouse	2W HDF	Av. droplet density	$\text{mm}^{-2}$	28	212.4	89.7	385.2	7.8	678.4	264.7	200.9
mouse	4W HDF	Av. droplet density	$\text{mm}^{-2}$	22	563.7	324.5	734.2	127.0	845.7	529.4	228.5
rat	2W HDF	Av. droplet density	$\text{mm}^{-2}$	13	1053.7	531.8	1589.3	62.8	1826.4	1028.6	625.4
rat	4W HDF	Av. droplet density	$\text{mm}^{-2}$	6	1458.7	1375.9	1535.6	1271.0	1576.3	1445.7	118.2
human	Steatosis	Surface coverage	%	6	7.7	7.3	8.6	5.1	10.6	7.8	1.8
mouse	2W HDF	Surface coverage	%	28	1.6	0.7	3.2	0.1	7.6	2.4	2.1
mouse	4W HDF	Surface coverage	%	22	8.3	3.9	10.5	0.9	14.6	7.4	4.3
rat	2W HDF	Surface coverage	%	13	11.2	6.4	19.1	1.2	23.8	12.5	8.4
rat	4W HDF	Surface coverage	%	6	42.0	38.7	42.9	24.7	43.6	38.8	7.2

**Table S15:** Segmentation validation metrics on the Test dataset

Grouped by	Group	Species subset	n	DSC (SemS)	PQ	DSC (InS)	NSD
species	All	all	15	0.889	0.812	0.923	0.784
species	mouse	all	8	0.837	0.714	0.866	0.7
species	rat	all	4	0.943	0.906	0.991	0.857
species	human	all	3	0.956	0.949	0.986	0.908
diet	2W HFD	mouse	6	0.795	0.636	0.824	0.634
diet	4W HFD	mouse	2	0.962	0.947	0.993	0.896
diet	2W HFD	rat	2	0.932	0.878	0.994	0.843
diet	4W HFD	rat	2	0.954	0.935	0.987	0.872