Image-based quantification of histological features as a function of spatial location using the Tissue Positioning System

Ruichen Rong,[a,e] Yonglong Wei,[b,e] Lin Li,[b] Tao Wang,[a,c] Hao Zhu,[b] Guanghua Xiao,[a,*] and Yunguan Wang[a,b,d,**]

[a]Quantitative Biomedical Research Center, Peter O'Donnell Jr. School of Public Health, University of Texas Southwestern Medical Center, Dallas, TX, 75390, USA
[b]Children's Research Institute, Departments of Pediatrics and Internal Medicine, Center for Regenerative Science and Medicine, University of Texas Southwestern Medical Center, Dallas, TX, 75390, USA
[c]Center for the Genetics of Host Defense, University of Texas Southwestern Medical Center, Dallas, TX, 75390, USA
[d]Division of Pediatric Gastroenterology, Hepatology and Nutrition, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, 45229, USA

Summary

Background Tissues such as the liver lobule, kidney nephron, and intestinal gland exhibit intricate patterns of zonated gene expression corresponding to distinct cell types and functions. To quantitatively understand zonation, it is important to measure cellular or genetic features as a function of position along a zonal axis. While it is possible to manually count, characterize, and locate features in relation to the zonal axis, it is labor-intensive and difficult to do manually while maintaining precision and accuracy.

Methods We addressed this challenge by developing a deep-learning-based quantification method called the "Tissue Positioning System" (TPS), which can automatically analyze zonation in the liver lobule as a model system.

Findings By using algorithms that identified vessels, classified vessels, and segmented zones based on the relative position along the portal vein to central vein axis, TPS was able to spatially quantify gene expression in mice with zone specific reporters.

Interpretation TPS could discern expression differences between zonal reporter strains, ages, and disease states. TPS could also reveal the zonal distribution of cells previously thought to be positioned randomly. The design principles of TPS could be generalized to other tissues to explore the biology of zonation.

Keywords: Tissue segmentation; Deep learning; Zonation; Expression pattern; Liver lobule

Introduction

Tissues such as the liver, kidney, and intestine carry out diverse biological functions. An elegant division of labor among cells within these tissues is spatially organized into "zones" that express specialized genetic programs.[1] For example, different zones in the liver carry out distinct metabolic functions. Liver zonation is dynamic and can change during development, aging, and disease, making it important to measure changes in zonal boundaries and gene expression patterns under different biological conditions. Necessary for this kind of analysis is the quantification of important features within a tissue and understanding how they are spatially organized. For any given coordinate on an image, one

*Corresponding authors.
**Corresponding author. Quantitative Biomedical Research Center, Department of Population and Data Sciences, University of Texas Southwestern Medical Center, Dallas, TX, 75390, USA.
  E-mail addresses: guanghua.xiao@utsouthwestern.edu (G. Xiao), Yunguan.Wang@cchmc.org (Y. Wang).
[e]Authors contributed equally.

Research in context

Evidence before this study
Currently, most approaches to distinguish tissue zonation in the liver was primarily performed through qualitative methods, relying on manual curation of histology images. There were only a few studies attempted to develop automatic tissue zonation segmentation tools. However, these tools were not robustly evaluated (Suneetha et al., 2021) and not able to support quantitative study of zonal expression patterns (Budelmann et al., 2022).

Added value of this study
We developed the TPS, a deep learning algorithm specifically designed to address the challenge of quantifying positional information and zonation within tissues, using the liver as the model system. We validated zonal expression patterns predicted by TPS using 14 CreER mouse strains with zonated reporter expression restricted to different zones in the liver.

Implications of all the available evidence
TPS offers a systematic and quantitative approach to measure tissue features in relation to spatial position, while at the same time addressing the limitations of qualitative methods and inference-based techniques. It allows for a more comprehensive understanding of zonation within tissues, potentially uncovering new insights into specialized genetic programs and their organization. Moreover, the successful application of TPS to liver zonation suggests its potential to be adapted for the analysis of zonated tissues in various organs. This study underscores the need for computational tools, like TPS, to enhance our understanding of tissue organization and enable more rigorous quantitative analysis in biological research.

would like to know the position with respect to anatomical landmarks that define the zonal axes. Once this "location" is obtained, phenotypes and gene expression at this position can be assessed as a function of anatomical coordinates, or vice-versa. This kind of analysis is ordinarily performed in a qualitative fashion by visual inspection of histology images. However, it would be preferable to obtain these data using the systematic and quantitative evaluation of images. Dedicated computational tools to parse zonation in an automated fashion have not been developed. The liver is the ideal setting to establish an algorithm to measure tissue features as a function of spatial position.

The lobule is the basic functional unit of liver tissue that is shaped like a submarine in three dimensions, hexagonal when cross-sectioned, and repeated throughout the liver.[1,2] Lobules are outlined by portal veins (PVs) at the vertices and one central vein (CV) in the center. Portal venous blood from the intestine intermixes with oxygenated arterial blood in the portal region, then flows through the sinusoids and collects in the CVs before returning to the heart. The notion of liver zones was first described in 1954 by Rappaport,[3] who assigned hepatocytes to three zones, periportal (zone 1), midlobular (zone 2), and perivenous (zone 3), based on their distance to the nearest CV or PV. Liver zones are associated with specific functionality: lipid β-oxidation and gluconeogenesis in zone 1 and lipogenesis, ketogenesis, and glycolysis in zone 3.[1] In general, these zonated functions have been characterized with immunostaining of metabolic enzymes rather than with functional testing. While individual genes were previously known to be zonated, single cell sequencing (scRNA-seq) technologies have expanded the molecular understanding of how genes and cell types are organized into zones.[4-6] To generate scRNA-seq data, cells

are enzymatically dissociated so their actual spatial position can only be inferred based on gene expression landmarks. In addition, technical difficulties associated with scRNA-seq such as doublets and dropouts make these inference based methods in some ways less reliable than image-based methods. Some of these limitations could be resolved with spatial transcriptomic technologies, which would combine the geographic integrity of immunostaining with the profiling depth of scRNA-seq. With spatial approaches, there will be a critical need to measure the position of cells, features, or gene expression values with respect to zonal landmarks. These technologies are not yet available.

Despite recent advances in computer vision and deep learning, automated analysis of histology images has not yet met the challenge of defining zonation. In the liver, applications have focused on the segmentation of cell nuclei and the prediction of ploidy states.[7,8] An unresolved problem is how to determine the position of any given modality (i.e. cell, clone, gene expression value) on an image with respect to important zonal landmarks. A key challenge is to be able to properly identify or segment zonal landmarks such as CVs or PVs, which appear as histologically similar vessel lumens lined by thin endothelial cells. Currently, most approaches to distinguish these entities rely on manual annotation based on expression of CV or PV-associated markers.[9-11] Among the studies that can automatically segmented liver zones, one used geometric methods to curate histologic features and built a support vector machine classifier to differentiate CVs from PVs.[12] However, the methodology was not described and was only tested on a small number of images so the robustness remains uncertain. A recent study used convolutional neural networks to identify portal fields (PF) and CV,[13]

however, this method could only report the presence of PFs and CVs without their boundaries, limiting additional downstream analysis. In addition, the performance was only validated using GS expression and thus its ability to detect zone 1 and zone 2 expression patterns remains uncertain.

Here we developed and employed a deep learning algorithm that we call the "Tissue Positioning System" (TPS), to enable quantitation of positional information, such as the zonation of liver lobules. The key rationale of TPS is that cellular features such as gene expression, proliferation, and damage, are anchored to key zonal structures in the tissue, and identification of those key structures allows for segmentation of tissue regions into biologically relevant regions. We applied TPS to liver zonation and found that it faithfully quantified the expected zonated expression patterns of the livers from 14 CreER mouse strains in an unbiased and reproducible fashion. We believe that TPS will be particularly useful for identifying tissue features that are organized into zones. In principle, the ability of TPS to define zonation around landmark structures can be adapted to any zonated tissue.

Methods

Mouse strains, breeding, and slide preparation

Please refer to Wei et al. for mouse model generation and experimental details involving histology, immunohistochemistry, and immunofluorescence.[14] Briefly, mice were euthanized and liver samples were fixed in 10% Neutral buffered formalin (NBF) for 48 h. After fixation, a liver sample was dehydrated in 30% sucrose in phosphate buffered saline (PBS) for cryosectioning. For frozen sections, slides were cut to 16 μm thickness, washed with PBS with Tween (PBST) 3x, and blocked in 5% bovine serum albumin (BSA) that contained 0.25% Triton X-100. Glutamine Synthetase (GS) antibody (Abcam ab49873) was diluted in the blocking buffer, applied to slides, and incubated overnight. Slides were imaged at 10x using an Axioscan (Zeiss) microscope and processed using Zen 2.6 software. A subset of the tissue sections and images analyzed in this paper were obtained from prior experiments.[14] Most of the images analyzed by TPS were independently performed for this paper. A total of 224 mice from 14 *CreER* strains were used in this study (Supplementary Table S1).

Imaging data preparation and annotation for deep learning model development

We developed an automated annotation pipeline workflow to 1) identify vessels from the images and 2) to classify them as either CVs or PVs using information from DAPI and GS channels.

The pipeline is based on a series of image morphological operations. The first step was to segment vessels on the slide. Briefly, dark regions on the DAPI channels

(DAPI masks) were identified by thresholding on a very small cutoff value (~10% of maximum DAPI intensity). Areas of each DAPI mask were calculated and the smallest ones were discarded because they were not likely to be major vascular structures (<1% of the total image area). Lastly, because vessels were sectioned in random orientations, i.e. cross-sectioned at different angles, or even tangentially but not through the lumen, some vessels could have been mistakenly removed using the vessel size cutoff. To mitigate this, we examined the very small DAPI masks and added them back if their associated GS intensity was above a GS cutoff, which was determined by applying Otsu's thresholding[15] on the GS channel. After these steps, the remaining DAPI masks were classified as vessel masks.

The second step was to annotate segmented vessels as CVs or PVs. We designed our pipeline based on the following features of liver lobule organization: 1) each liver lobule is centered around a CV, with surrounding PVs located on the perimeter of the lobule; 2) vessel lumens are devoid of nuclei that have high DAPI intensity; 3) CVs are lined with hepatocytes that have high GS intensity.

Specifically, intensity percentile values (lower limit = 25, higher limit = 75, interval = 10) from the GS channel were quantified for each vessel mask. Then, the mask/feature data were transformed by PCA and the masks were clustered into two groups using the KMeans algorithm. Finally, the masks with higher median GS intensity were labeled as CVs, and the others as PVs.

We applied the above heuristic imaging morphological pipeline to generate the masks and classified each vessel in 114 high-quality images to develop a deep-learning based segmentation model. The classes and masks were further visually validated by pathologists. We divided the 114 images into 61 for training, 24 for validation and 29 for testing, respectively.

Segmentation of CVs and PVs using deep learning

A deep-learning based segmentation algorithm was developed for CV and PV segmentation and classification. The algorithm was built with the UNet framework with the modifications described as follows. 1) We used the MobileNet-v2 backbone pre-trained on ImageNet to replace the standard CNN encoder in the original UNet. In comparison to a straightforward CNN encoder, MobileNet-v2 extracts more intricate high-level image features, and employing a pre-trained backbone boosts accuracy and lowers the risk of overfitting when the sample size is small. 2) To marginally improve performance, we swapped out the upsampling layer in the decoder for an inverted residual block plus a convolution transpose layer. The model structure is illustrated in Fig. 1b.

In the training process, instead of using cross-entropy loss with pixel heatmap, soft-dice loss was used to automatically balance pixel weights for different classes.
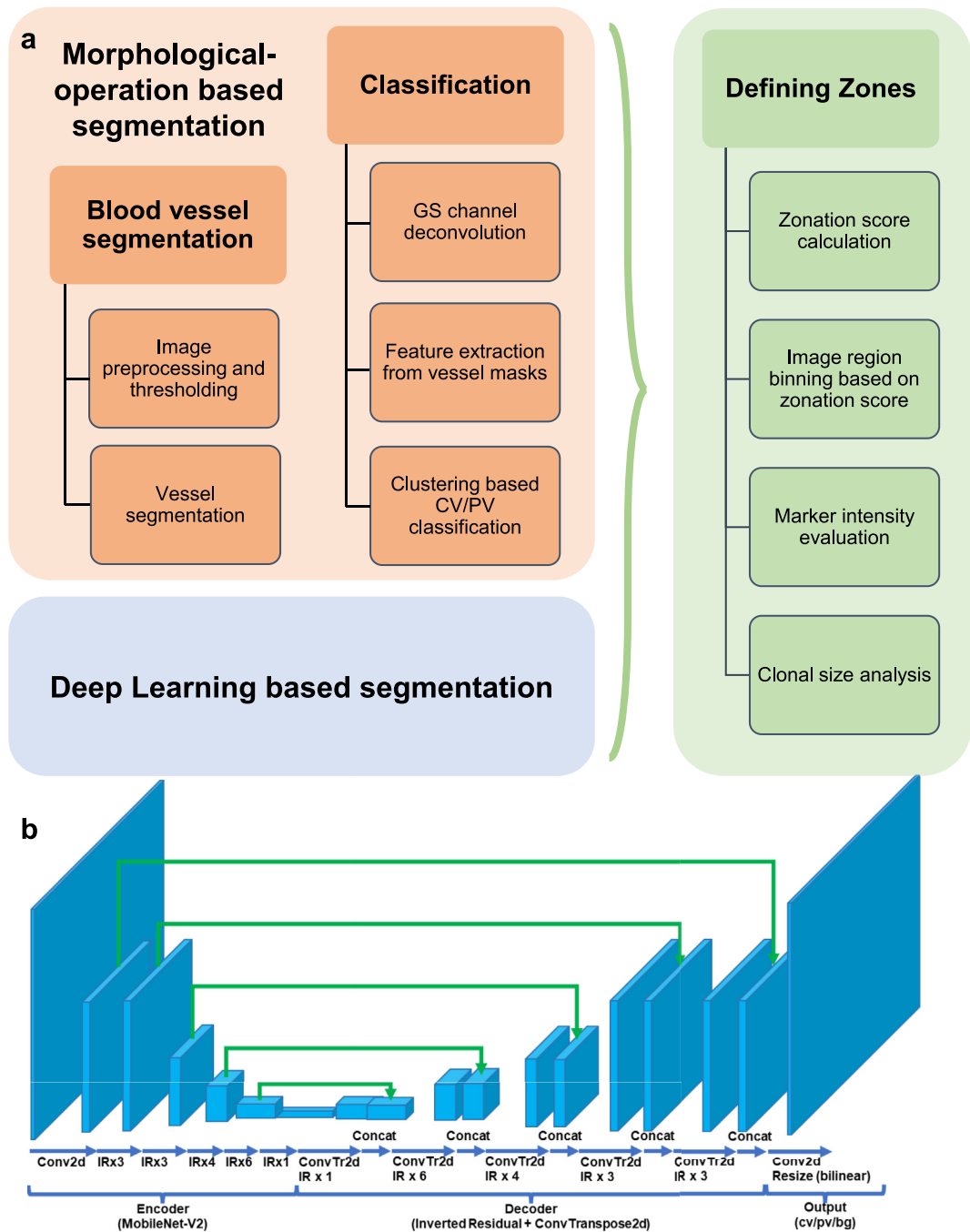
Fig. 1: Overview of the TPS workflow. a. Overview of the major steps in TPS including vessel segmentation and zone identification. b. Structure of the MobileNet-v2 + UNet deep learning model in TPS.

Briefly, a large input whole slide image was cropped into 3500*1500 image patches. Each image patch was padded into 4096*2048 images with the pixel information from the original slide. With the above approaches, there were 298 pixels overlapping on width and 274 pixels overlapping on height. This was to ensure that vessels near the edge of the images were recognized properly. Then, the vessel masks from all cropped images were stitched together and vessel classification and expression analysis were done on the full-size image using the

stitched vessel masks. In order to prevent overfitting, color intensity augmentation, random projective transformations, as well as random horizontal/vertical flipping and transposing were used. Specifically, the following parameter ranges were used in this process: contrast <0.1, brightness <0.1, HSV <0.1, rotation <30°, offset <50 pixel, scale factor <0.2, height/width ratio <0.1, shearing <10°, perspective distortion <0.0003. The above augmentations were performed with python package skimage 0.19.1 and imgaug 0.4.0.

In the validation process, mIoUs (median intersection-over-union) for CVs and PVs between ground truth and prediction were documented. The model with highest mIoU in the validation dataset was further evaluated on the testing dataset and deployed for downstream analysis.

Defining position as a function of zonal landmarks

We reasoned that lobular zones were continuous in a gradient fashion along the CV-PV axis, and thus we first calculated a positional score for each pixel based on its distance to the nearest CV or PV mask. This was performed as follows for each pixel. First, the distances to the nearest CV and PV were calculated. If the pixel's distance to the nearest CV or PV was larger than 25% of the image patch size, it was excluded from further analysis. Then, the ratio between the two distances were calculated, which reflects the position of the pixel in the context of a liver lobule. Lastly, the pixels were assigned to equal-sized groups, named TPS layers, by discretizing their distance ratios into quantile bins. We found this partition technique made the most uniform, natural looking zones within lobules compared to using raw distance ratios or projection distance along the CV-PV axis.

Zonal expression patterns in the tissue can be examined by evaluating marker expression in each layer. In this study, we used 24 equal-sized TPS layers so that each TPS layer would contain roughly 4.16% of the pixels in the image. For a typical image in this paper, each TPS layer contains approximately 400–500 cells. The selection of this number is arbitrary as the zonal expression patterns derived are not sensitive to the number of TPS layers.

Definition of Z_{max} and Z_{50}

To quantitatively describe the location and selectivity of a zonal expression pattern in the context of the classic 3-zone system, we introduced two metrics, Z_{max} and Z_{50}, respectively. Z_{max} reflects the hepatocyte layer where a marker is maximally expressed. It is defined as

$$Z_{max} = 3 * \left(1 - \frac{L_{max}}{N}\right)$$

where L_{max} is the layer number where a marker was expressed at maximal level, and N is the total number of TPS layers.

Z_{50} describes the selectivity of the distribution. It is defined as the fraction of zones where a marker is expressed below 50% of its maximal expression level L_{max}:

$$Z_{50} = \frac{1}{N} \sum_{1}^{N} [L_i < 0.5 * L_{max}]$$

where L_i is the expression level of marker at TPS layer i.

Measurement of clone size

The zonal expression pattern of markers can be divided into two broad categories: widely expressed in contiguous regions, or sparsely expressed in disparate regions. For sparsely expressed reporter strains, we could measure the clone number and size. Size was calculated as the marker expressing clone area or the number of nuclei in each clone. The number of nuclei in each clone was calculated using the watershed[16,17] segmentation algorithm on the DAPI masks derived from Otsu[15] thresholding.

Analyzing whole slide images

The TPS algorithm can also analyze whole liver section images as inputs, and this was done by parallel processing of cropped images from whole liver section images. Briefly, a large input whole slide image was cropped into 3500*1500 image patches. Each image patch was padded into 4096*2048 images with the pixel information from the original slide. With the above approaches, there are 298 pixels overlapping on width and 274 pixels overlapping on height. This ensured that vessels near the edge of the images were recognized properly. Then, the vessel masks from all cropped images were stitched together and vessel classification and expression analysis were done on the full-size image using the stitched vessel masks.

Lobule detection

We used the watershed algorithm to detect lobules in the hepatocyte image. Briefly, distances from each pixel to the closest CV mask were calculated. Then the local maxima of the distances with each CV mask were determined. Finally, the watershed transformation was applied using the distances and local maxima to find the boundaries that separated areas around CV masks into different regions, which reflected the lobular structure in the liver.

Statistics

A two-tailed Student's t test was used to evaluate if the percentage of reporter positive area of cells was different between TPS layers. A value of $p < 0.05$ was considered significant.

Ethics

All mice were handled in accordance with the guidelines of the Institutional Animal Care and Use Committee (Ref Number: 2015-101118) at UT Southwestern.

Role of funding source

The funders had no role in the study design, data collection, data analyses, interpretation, or writing of report, and the decision of paper submission.

Results

Image-based quantification of tissue features as a function of zonal location

"Tissue Positioning System" (TPS) is software that allows automatic evaluation of tissue features as a function of zonal position within histology sections. Examples of a "tissue feature" could be any phenotypic measurement that can be assigned to a particular coordinate: i.e. gene expression, proliferation, cell size, nuclear size, clone size, lipid droplets, inflammatory cells, etc. TPS operates through two major components (see the workflow in Fig. 1). The first component of TPS is a deep learning model that uses the nucleus channel of an image to segment CV and PV masks. In our experiment, the model could accurately identify zonal position in the 29 test slides with 0.8117 mIoU for CV and 0.8908 mIoU for PV, respectively (see Supplementary Table S2a for the segmentation performance on training, validation, and testing datasets; see Supplementary Table S2b for detection performance of vessels in the 29 testing images). In the second component of TPS, positional scores proportional to the ratio between the distances of a pixel to its nearest CV and PV masks were calculated, and then pixels emanating from each concentric CV or PV were divided into equidistant bins based on this score. These bins are referred to as TPS layers. In this study, we tested the performance of TPS in determining the zonation of features such as reporter gene expression and cellular clone size.

TPS extracts positional expression patterns from zonal reporter mice

We first tested TPS on liver sections from fluorescent reporter mice that express the Tomato fluorescent reporter in known zonal patterns. These mice (*GS-CreER*, *Cyp1a2-CreER*, and *Gls2-CreER*) were previously characterized in a qualitative fashion.[14] The accurate discernment of zone-specific reporter expression by TPS would validate the algorithm's ability to accurately assign positional information to all pixels on an image. Each *CreER; Rosa-LSL-Tomato* mouse was given tamoxifen to activate Tomato expression prior to liver sectioning. The Tomato expression patterns for these mice are highly specific to particular zones (Fig. 2a and b). As mentioned previously, Tomato in *GS-CreER* mice

is expressed in a 1–3 hepatocyte thick ring around all CVs. Tomato in *Cyp1a2-CreER* mice is expressed in all zone 3 and about 50% of zone 2 hepatocytes emanating from the CV. Tomato in *Gls2-CreER* mice is expressed in all zone 1 hepatocytes emanating from the PV. TPS identified CV and PV regions accurately without mistaking smaller gaps between cells, likely representing sinusoids, as large vessels (Fig. 2c). To further evaluate the accuracy of predicting CVs and PVs, we predicted liver lobule boundaries associated with each CV. The lobule boundaries were very close to the center of PV masks (Fig. 2d), indicating that the segmented CVs and PVs were reflective of true lobular structures. These results showed that TPS was able to accurately identify the landmark vessels that define liver zones in an unsupervised manner, thus providing a strong rationale for the downstream assessment of tissue features.

After vessel segmentation, positional scores for all pixels were calculated and used to divide each image into 24 TPS layers reflecting zonal positions along PV-CV axes. We chose 24 for this study because at this level, zonal features of liver lobules can be adequately resolved. Pixels that were too far from any CV or PV were excluded because this likely meant that the actual nearest CV or PV was not correctly located. Consistent with lobular zonation patterns, TPS layers were organized into ring-shaped patterns around CVs and PVs (Fig. 2e). TPS correctly showed that GS and CYP1A2 associated Tomato expression was limited to zone 3 (TPS layer 1–6), while GLS2 associated Tomato expression was highest in zone 1 (TPS layer 19–24), and reduced in zone 2 (Fig. 2f). These results showed that TPS was able to reliably identify vessels, define zonal axes, and accurately quantify expression patterns in a probabilistic fashion.

TPS discerns differences in the spatial expression patterns of genes with high resolution

Next, we evaluated TPS's performance on each of the eleven *CreER* reporter mice generated by our group and the three other *CreER* mice available to the community. These fourteen zonal markers were separated into five broad categories based on their expression patterns. These include pan-zonal, all hepatocyte expressing strains (*Apoc4-CreER*, *Pklr-CreER*), biliary-centric strains (*Krt19-CreER*, *Sox9-CreER*), zone 1-centric strains (*Gls2-CreER*, *Arg1.1-CreER*, *Arg1.2-CreER*), zone 3-centric strains (*Axin2-CreER*, *Oat-CreER*, *Cyp1a2-CreER*, *GS-CreER*), and sparsely expressing strains (*Hamp2-CreER*, *Tert-CreER*, *Mup3-CreER*) (See schema in Fig. 3a). TPS was able to accurately quantify the differences among all of these expression patterns.

In addition, TPS allowed us to resolve expression patterns between reporter mice that have different expression patterns within the same zones. For example, the expression patterns of the three zone 3-
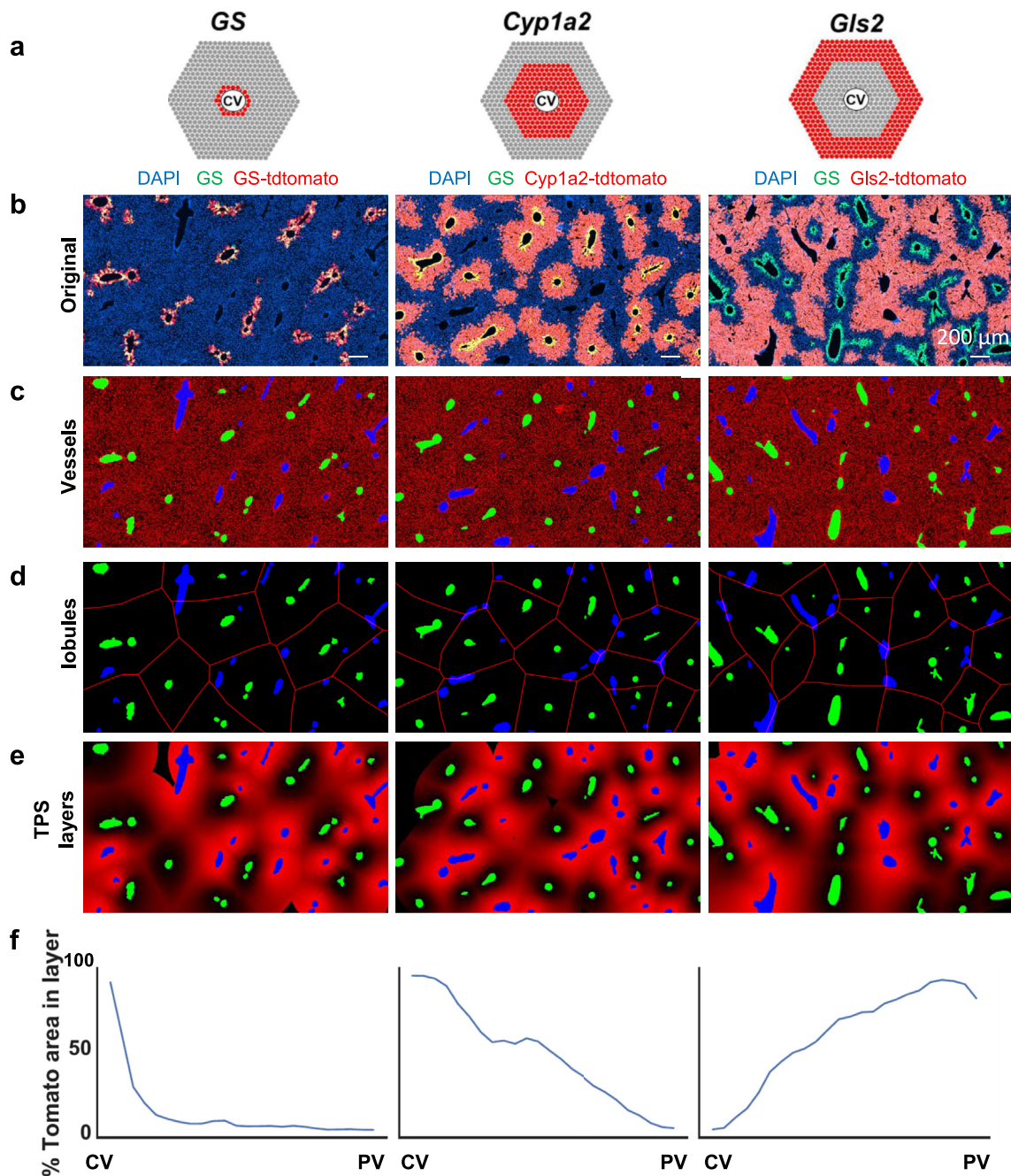
Fig. 2: TPS characterization of Tomato expression patterns within *GS-CreER, Cyp1a2-CreER,* and *Gls2-CreER* mice. a. Schema of expected Tomato labeling in *GS-CreER, Cyp1a2-CreER,* and *Gls2-CreER* mice. b. Original three-channel immunofluorescence images of livers from CreER lineage tracing mice. Frozen sections were stained with anti-GS antibody and then scanned by a ZEISS Axio Scan. Z1 using a 20X objective (blue = DAPI, green = GS, red = Tomato). Scale bars, 100 µm. c. Segmented and classified CVs and PVs, shown in green and blue, respectively, for this and below panels. DAPI stains were colored red in these images. d. Predicted liver lobule boundaries, shown with red lines. e. 24 TPS layers predicted based on the pixel wise distance along the CV-PV axis. TPS layers are shown in red, and those closer to PVs are brighter. f. Line graphs showing the percent Tomato positive area/total area in each TPS layer. TPS layers were organized from CV to PV going from left to right. Data from one mouse is shown here.
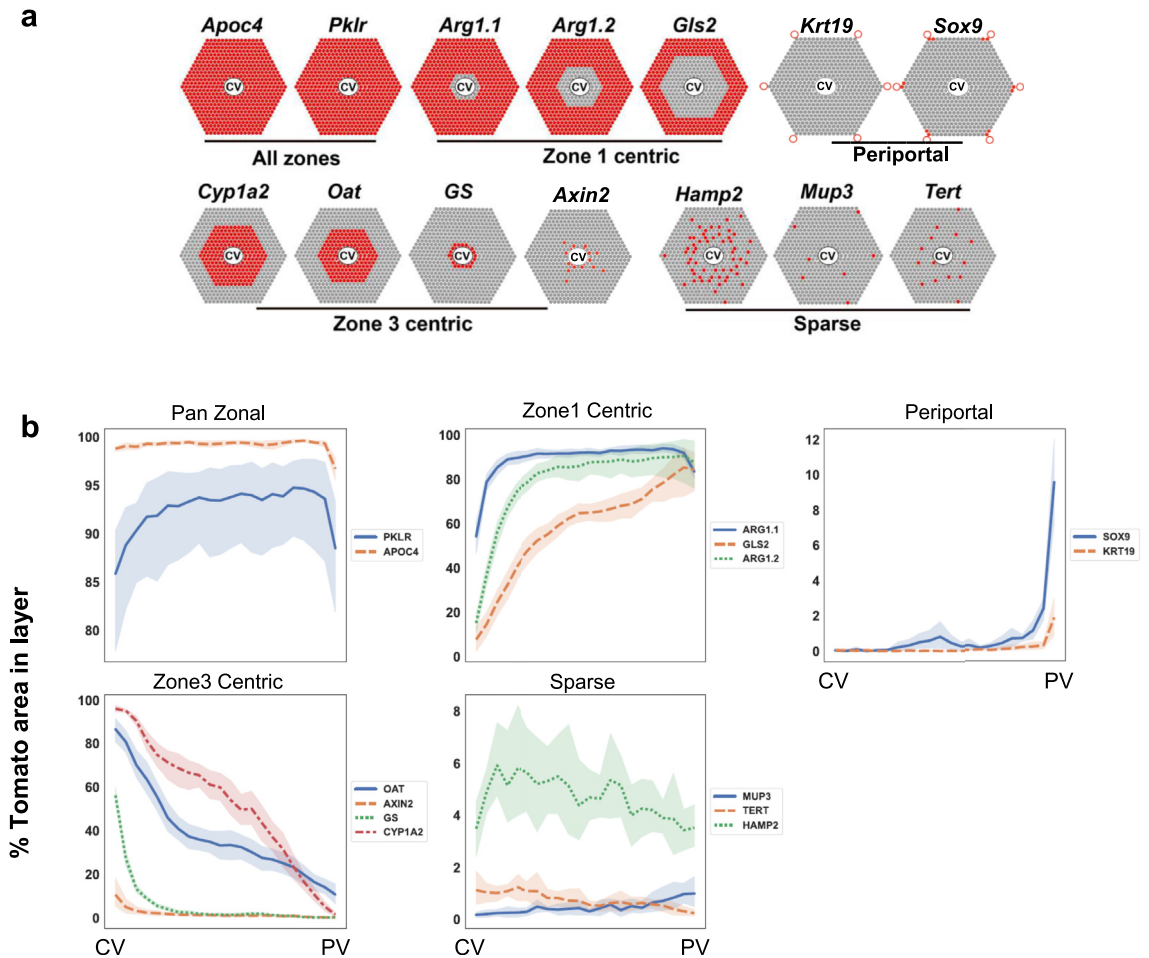
Fig. 3: TPS is able to discern differences between qualitatively similar zonal expression patterns. a. Five categories of zonal expression patterns from different CreER lineage tracing mice are schematized here. All zones: *Apoc4-CreER, Pklr-CreER*. Zone 1 centric: *Arg1.1-CreER, Arg1.2-CreER, Gls2-CreER*. Periportal: *Krt19-CreER, Sox9-CreER*. Zone 3 centric: *Cyp1a2-CreER, Oat-CreER, GS-CreER, Axin2-CreER*. Sparse or zone 2 centric: *Hamp2-CreER, Mup3-CreER, Tert-Cre*ER. b. Line graphs showing the percentage of tomato positive area/total area in each TPS layer across the CV-PV axis. Data were collected from CreER mice 1 week after tamoxifen treatment. The 95% confidence intervals for each TPS layer are shown as shaded areas.

centric strains (*Cyp1a2-CreER, Oat-CreER, GS-CreER*) could be clearly distinguished from one another using TPS (Fig. 3b). To quantitatively compare zonal reporters within the same category, we introduced two metrics to characterize zonal gene expression. The zone score, Z_{max}, describes the zonal location of peak expression, ranging from 1 to 3, corresponding to current 3-zone classification. For example, Z_{max} of GS and GLS2 are 2.92 and 1.25, respectively. A second metric, Z_{50}, describes the selectivity of the distribution, which measures the number of zones where a particular marker is expressed below 50% of its maximal expression level. Z_{50} values range from 0 to 1, where 0 indicates the expression is uniform and 1 indicates the expression is restricted only to one TPS layer. For example, Z_{50} of GS is 0.96, while Z_{50} of Apoc4 is 0. Among all zone 3

centric markers, Z_{50} is highest for GS and lower for Oat (0.54) and Cyp1a2 (0.46) (Table 1). These results showed that TPS could faithfully report zonal expression patterns and allows a more quantitative classification of expression patterns.

TPS is able to define the spatial patterns of sparsely expressed cells

If TPS was only able to quantitatively define the zonal expression patterns of reporter mice with obvious zone-specific expression patterns, then the algorithm would not be very useful. It would be essential for TPS to be able to determine if there is zonal specificity for features that have less obvious patterns. For instance, TPS would be useful for the characterization of reporter strains that label cells in a sparse manner because visual discernment

Marker	Z_{max}	Z_{50}
APOC4	2.5	0
ARG1.1	1.5	0.125
ARG1.2	1.75	0.21
AXIN2	2.92	0.96
CYP1A2	2.92	0.46
GLS2	1.25	0.33
GS	2.92	0.96
HAMP2	2.08	0.21
KRT19	1	0.98
MUP3	1	0.78
OAT	2.92	0.54
PKLR	1.92	0.04
SOX9	1	0.96

Table 1: Z_{max} and Z_{50} for all zonal reporters.

of their zonal distribution is nearly impossible. For example, the spatial distribution of Tomato positive cells in *Hamp2-CreER* mice cannot be assessed accurately by eye because of two key reasons (Supplementary Fig. S1a). First, the area occupied by the three zones is different, and thus the total number of cells in each zone declines significantly from the PV to CV. Without proper normalization for this difference in cell numbers, a perceived enrichment of labeled cells in zones that occupy larger areas would arise. Second, manual measurements are labor intensive and difficult to perform for more than a few hundred cells. For markers that express even more sparsely than HAMP2, such as MUP3 and TERT (Supplementary Fig. S1b and c), collecting zonal information from a few hundred cells would require analyzing dozens of large liver sections.

The first problem was solved by TPS naturally as it finds liver zones automatically. We assessed the *Hamp2-CreER* and *Mup3-CreER* reporter mice that label a small subset of hepatocytes in the liver. TPS analysis of sections from *Hamp2-CreER* mice showed that cells from any zone could express Tomato, but that midlobular zone 2 cells were labeled with the highest probability. TPS analysis of sections from *Mup3-CreER* mice also showed that cells from any zone could express Tomato, but zone 1 cells were labeled with the highest probability, followed by zones 2 and 3 (Fig. 3b). Then we analyzed *Tert-CreER* mice created in our lab. TERT high cells were reported to be randomly distributed throughout the lobule based on visual inspection.[10] TPS analysis of sections from *Tert-CreER* mice showed that cells from any zone could express Tomato, but zone 3 cells were labeled with the highest probability (Fig. 3b). TPS was able to discern the distinct patterns of each of these two reporter strains that sparsely label 0.3–10% of cells. These results show that a zonated pattern can exist even though visual inspection may only appreciate a randomly distributed pattern of expression.

For the second problem, we developed a function in TPS that can process images from full-sized liver sections in parallel, so that a large number of cells can be easily measured with analysis of whole-liver section images within a few hours on a HPC (High Performance Computer) node. We achieved this using the following steps. Each whole liver section image was divided into fixed-sized cropped images, and TPS segmentation was performed on each image in parallel (Fig. 4a and b). Then, cropped images containing vessel masks were re-assembled and used as the vessel mask in later analyses (Fig. 4c and d). Finally, TPS layers were defined on the original image. Zonal expression patterns derived from this large image processing were identical to those that were derived from smaller images (Fig. 4e and f).

We tested this workflow using hepatocyte labeling in the *Rosa-LSL-Tomato* reporter mouse using low titer AAV-TBG-Cre injection, as is commonly used to achieve genetic recombination in a small subset of hepatocytes in a sparse but "random" fashion.[18] Indeed, TPS analysis of sections from these mice showed that hepatocytes were labeled with Tomato in an evenly distributed pattern across zones (Fig. 4f). Next, we applied this work flow on whole-liver section images derived from *Mup3-CreER* and *Tert-CreER* reporter mice and found similar zonal expression patterns compared to those from smaller cropped images (Supplementary Fig. S2).

TPS facilitates lineage tracing of cells during steady state homeostasis

TPS is also helpful in evaluating changes in labeling patterns during development, aging, or disease. To examine hepatocyte repopulation from different zones during steady state homeostasis, we applied TPS to all of the above *CreER* mice that were traced for over 26 weeks, or 6 months.[14] The change in the percentage area labeled by Tomato reflects the expansion or shrinkage of domains that were labeled after tamoxifen induction. TPS analysis confirmed an expansion of cells in *Cyp1a2-CreER* and *Oat-CreER* mice, as well as a contraction of cells in *Gls2-CreER* mice (Fig. 5a). The most significant increase occurred in zone 2 cells labeled by *Hamp2-CreER* mice. The percentage of cellular area labeled by Tomato in *Hamp2-CreER* mice increased by more than 3-fold after 6 months of tracing (from 4.63% to 14.51%; p-value <0.0001) across all zones, but the most enriched region of expression was found near zone 2 (Fig. 5a). Using TPS, we again found that zone 2 had a large increase in the percentage area labeled by Tomato between time points, suggesting greater expansion of cells from this midlobular zone. These results showed that TPS can extract aspects of zonal expression patterns that are very difficult to quantify using traditional methods.
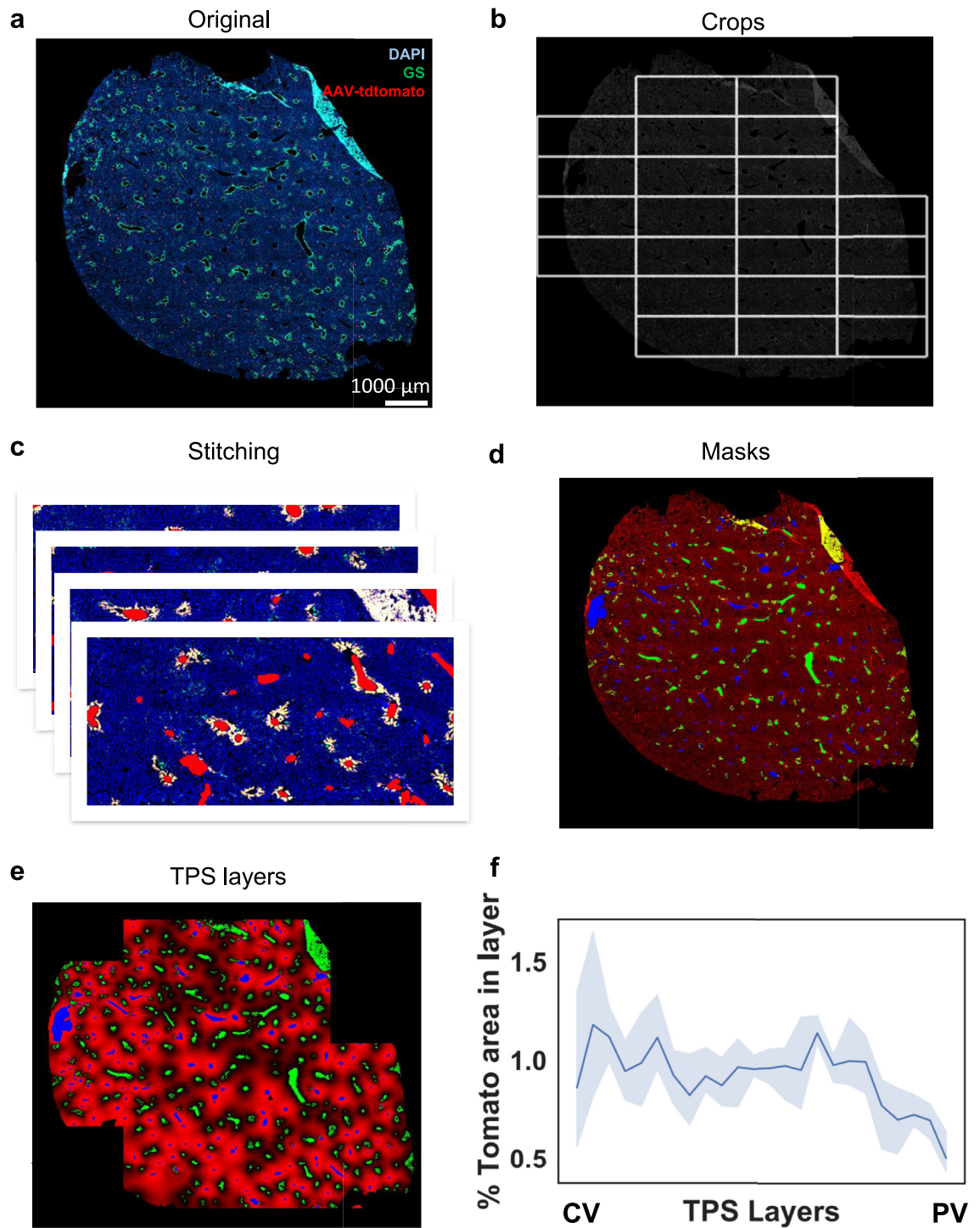
Fig. 4: Processing of whole-liver section images by TPS. TPS can be scaled up to process whole-liver section images. This is especially useful for features that are rare or sparse, such as the low dose AAV-TBG-Cre labeling that is shown in this figure. This is achieved using a bottom up approach where each whole-liver section image is divided into cropped images and analyzed in a parallel fashion. Thus, CV/PV masks from each cropped image were pooled and used in the remaining TPS processes. a. Original fluorescence image obtained from Tomato reporter mice treated with low titer AAV-TBG-Cre. The liver from this AAV experiment was obtained from Wei et al. and analyzed here. b. Cropped images used in the TPS segmentation steps. Images were extracted in a sliding window pattern, and cropped images not covering significant amounts of tissue (median DAPI intensity = 0) were not processed. Each image was set to be 3500 x 1500 pixels, but during TPS segmentation, regions

TPS is able to measure clone number, size, and location within the lobule

Beyond reporter expression, TPS was able to analyze tissue features associated with cellular clones. TPS can measure the number, size, and location of clones within lobules, information that is crucial for the understanding of regenerative activities of cells as a function of location. Here, a clone is defined as a contiguous set of reporter expressing cells surrounded by reporter negative cells. Clone size can be measured in terms of pixel area or nuclei number. The number of nuclei per clone was determined by using watershed segmentation[16] of the DAPI channel. We identified, counted, then measured clone size in sparsely labeled reporter models such as *Hamp2-CreER* and *Mup3-CreER* mice. In *Hamp2-CreER* mice, the percentage of clones with more than 1 nuclei increased in all zones after 6 months of tracing (Fig. 5b). This indicated that many clones expanded over time. Although expanding clones could be identified in every zone, the highest number of large clones were found in midlobular zone 2 (Fig. 5c). Similarly, we found that Tomato labeled cells in *Mup3-CreER* mice could be found in each zone, but the largest clones (≥2 nuclei) were enriched in zone 2 (Supplementary Fig. S3 and b). These data confirmed that clonal expansion occurs more frequently in zone 2 compared to other zones. These results provided an example of how TPS-empowered zone segmentation can be used to systemically and reproducibly extract information about clonal dynamics as a function of location.

TPS revealed differential regeneration from cells in different zones during injury

We also used TPS to evaluate zonal expression patterns in damaged liver tissues. To induce chronic biliary injury, we fed reporter mice with diets containing 0.1% DDC for 6 weeks, after which we harvested the liver and examined the distribution of tdTomato labeled cells. DDC is a strong cholestatic injury that models human cholangiopathies.[19] Importantly, DDC induces an inflammation related "ductular reaction" around the PV that is also seen in human liver diseases. As expected, DDC liver injury resulted in histological artifacts that complicated the analysis. For example, dark spots observed near the portal triads (Fig. 6a) may represent cellular debris or inclusion bodies, and if coalesced together, could be mistaken as large vessels. After introducing an additional step to shrink the dark areas

in the image to mitigate this possibility, TPS segmentation was able to reliably identify and distinguish CVs and PVs, and quantify Tomato expression across the lobule (Fig. 6a–c). To examine the changes that occur in pan-zone, zone 3, and zone 1 labeled mice fed DDC, we gave tamoxifen to *ApoC4-CreER*, *GS-CreER*, and *Gls2-CreER* Tomato reporter mice, then gave them DDC food. Comparing uninjured age matched control and DDC fed livers showed that cholestatic injury caused a significant contraction of Tomato labeled zone 1 cells near the PV (TPS layers 19–24) in both the *ApoC4-CreER* and *Gls2-CreER* strains (P-value <0.0001 in both cases) (Fig. 6c and d). In agreement with this, there was an increase in the number of zone 3 cells labeled with Tomato (Fig. 6c and d). These results were consistent with the idea that DDC-induced liver damage causes loss of cells positioned in zone 1, which are in turn replaced by cells originating from zones 2 or 3. In addition, this showed that TPS was able to quantify zonal repopulation, even in the context of liver injuries that result in histological artifacts associated with tissue damage. This makes TPS a particularly useful tool for studying the cellular dynamics of livers in various settings.

Discussion

In the liver, tissue patterning or zonation is thought to be important for normal liver metabolic function,[20,21] but whether or not altered zonation might impact normal physiology and disease pathogenesis is not known. Very little is known in part because there are no computational tools to measure zonation. In addition, zonation is dynamic and can change through development and with disease, adding to the need to measure zonation over time automatically and with high precision. To our knowledge, TPS is a fully automatic, unsupervised tool that outputs zonal information directly from input images. It can efficiently process a 4000*2000 image in minutes and a 20,000 x 20,000 whole tissue section image in less than half an hour.

There are several critical advantages of TPS. Currently, the identification of zonal axes usually requires visual analysis of images and manual discrimination of vessel identities,[9–11,22] which is labor intensive and susceptible to bias. Existing computational tools for liver zonation analysis only provide limited information about CV and PV locations and their effectiveness in characterizing different zonal patterns is unknown.[12,13]

---

of 250 pixels were added to each image to allow the correct identification of vessels near cropped image boundaries. c. The results from TPS segmentation on each cropped image were stitched together and used as the overall vessel mask in the following TPS processes. d. Shown are the CVs (green) and PVs (blue) in the stitched masks derived from all individual images. DAPI staining was colored in red. e. TPS layers were shown in red, and those closer to PV were brighter. CV and PV masks were shown in green and blue, respectively. f. Line plot of the percentage of total tomato positive area by the total cellular area in each TPS layer. The TPS layers were organized from CV to PV from left to right.
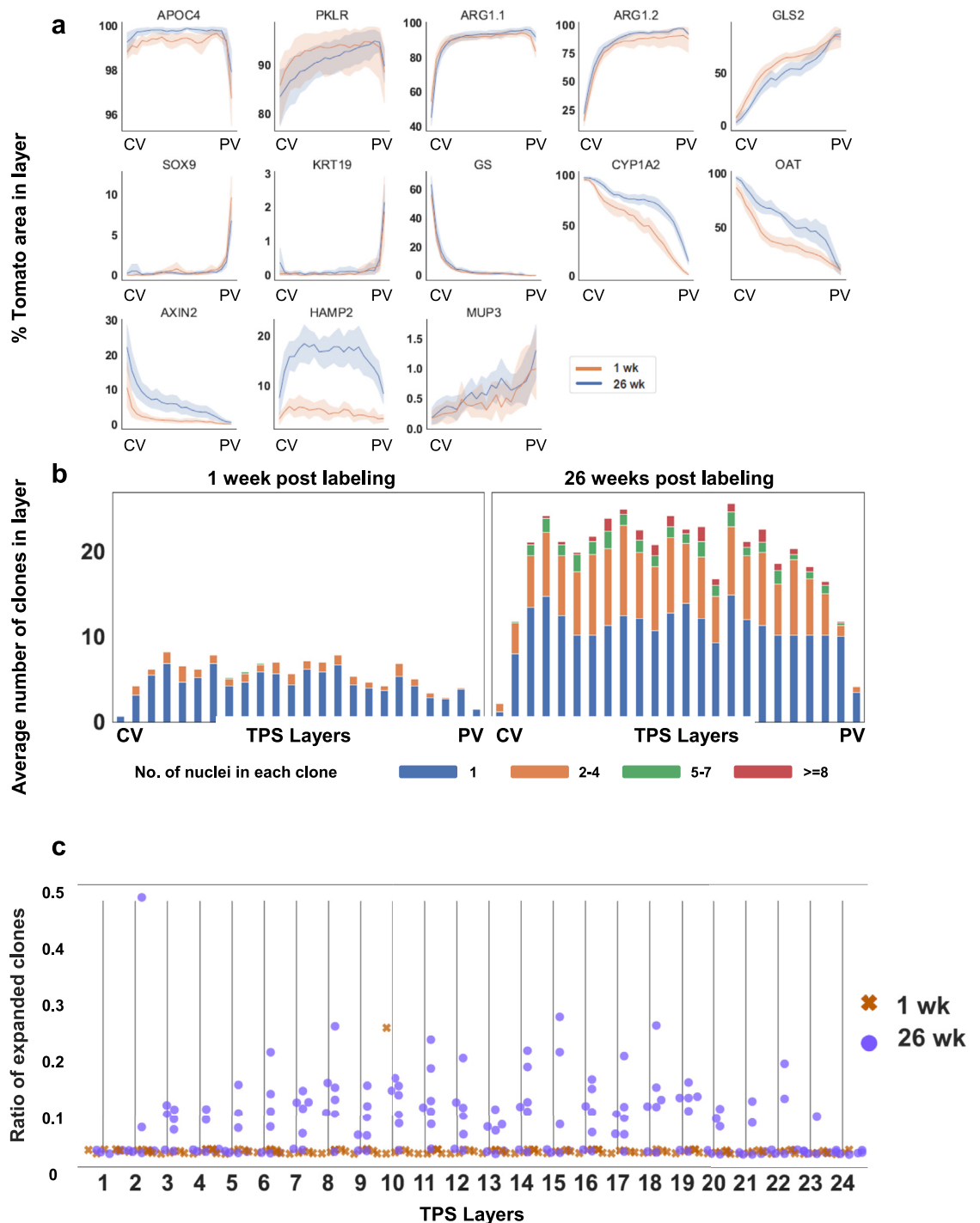
Fig. 5: TPS revealed dynamics of liver cell subpopulations from different zones at 1 and 26 weeks after reporter labeling. TPS quantifies the changes in Tomato labeling patterns during steady state homeostasis. This analysis is able to reveal differential lineage fates of cells marked within the same zones. a. Comparison of lineage tracing results of 13 CreER markers 1 and 26 weeks after tamoxifen induced Tomato labeling. TPS layers were organized along the CV-PV axis. The 95% confidence interval for each TPS layer is shown as shaded areas. b. Number of nuclei in each Tomato marked clone in liver sections from Hamp2-CreER mice. Data were grouped into four clone size categories as follows: 1 nuclei, 2–4 nuclei, 5–7 nuclei and >8 nuclei. c. Change in the spatial distribution of expanded clones (defined as clones with more than 3 nuclei) across hepatocyte zones in *Hamp2-CreER* mice examined at 1 and 26 weeks after tamoxifen.
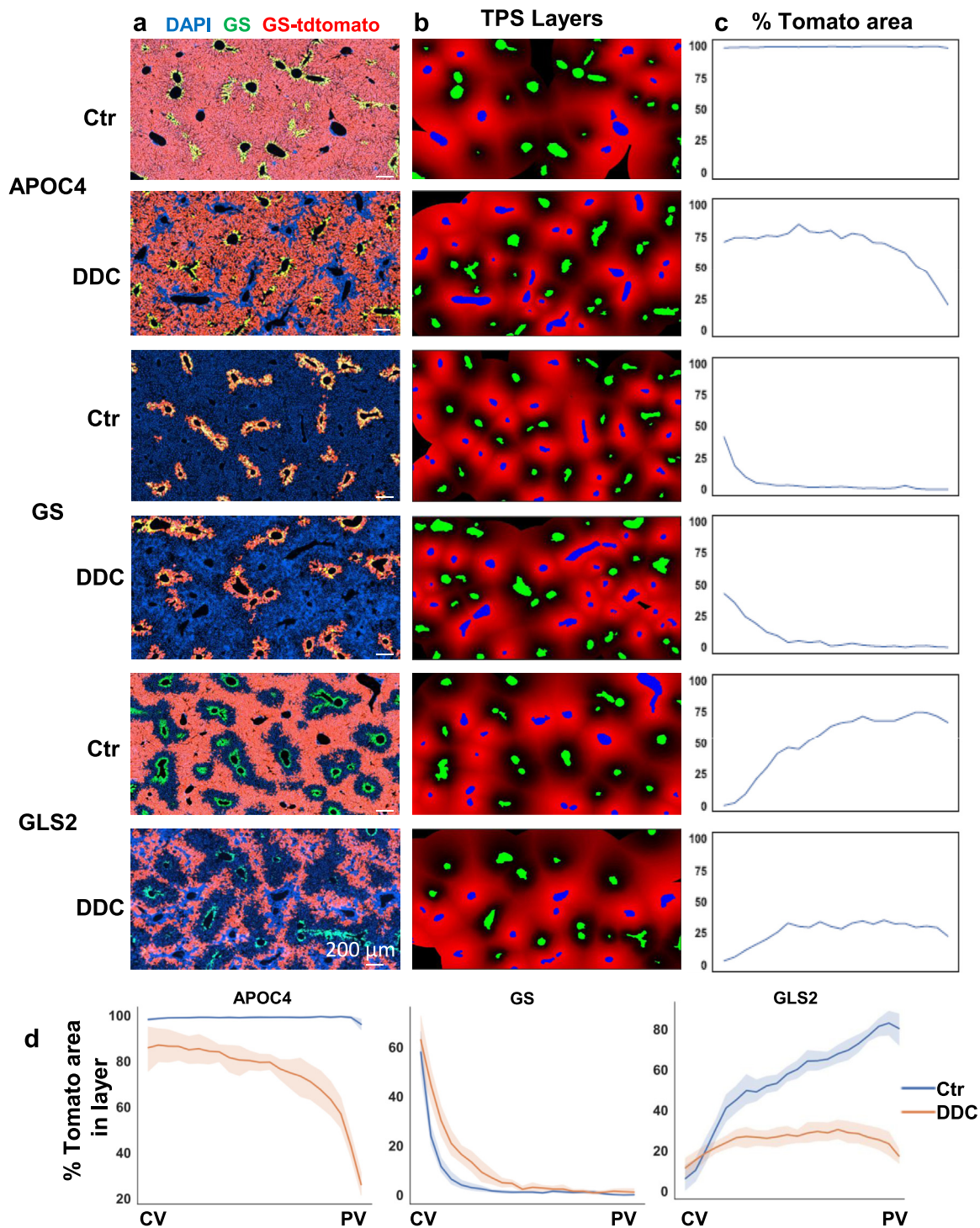
Fig. 6: TPS revealed cellular dynamics in different zones during biliary injury. TPS is able to examine changes in Tomato labeling in livers that have been exposed to DDC. a. Representative original images from control and DDC treated CreER mice. b. Predicted TPS layers from each representative image. Layers were shown in red, and those closer to PV were brighter. CV and PV masks were shown in green and blue, respectively. c. Line plot showing the percentage of total tomato positive area by the total cellular area in each TPS layer from the representative images. The TPS layers were organized from CV to PV in left to right order. d. Comparison of Tomato area percentage between control and DDC treated group. The TPS layers were organized from CV to PV in left to right order. The 95% confidence interval for each TPS layer is shown as shaded areas.

TPS is a complete, fully automatic zonation analysis tool that allows accurate identification of CVs (f1 = 0.8500, mIoU = 0.9029) and PVs (f1 = 0.8620, mIoU = 0.9170), as well as characterization of zonated expression patterns. The robustness of TPS in detecting the correct zonation pattern was demonstrated in protein markers with different zonated expression patterns. In addition, the position of tissue features such as expression are reported as continuous variables along the CV-PV axis, rather than a qualitative measure based on the arbitrarily defined 3-zone system. This makes it possible to use more robust statistical methods to compare the zonal positions of tissue features. Because of these advances, results from TPS are likely to be more accurate, reproducible, and comparable between different experiments.

TPS was able to determine the expression pattern of reporter mice with zone-specific expression, as well as mice with undefined, sparse expression patterns. We applied TPS to sections from *CreER* mice lines labeled with markers with different zonal expression patterns. TPS faithfully reported expected zonal expression patterns of all 14 of these reporter lines. In addition, TPS was able to quantitatively distinguish distinct strains that express reporters in a qualitatively similar fashion (i.e. within the same zone). TPS was also able to define the zonal distribution of sparsely labeled cells within lobules. The sparseness of labeling makes it difficult to determine the zonal expression pattern with manual visual analysis. For example, TERT expression in the liver was previously reported to be uniform across lobules,[10] however, TPS showed that TERT reporter expression had a modest location preference for zone 3. TPS can analyze whole liver sections, which increases the throughput and thus improves the accuracy of results. TPS is particularly suited for analyzing sparse markers to determine if there is a zone-specific expression pattern.

While we applied TPS to specific questions in liver biology, at its core TPS is able to analyze spatial location in the context of any positional landmarks. Any data associated with a position on a tissue section can be assessed with TPS. For example, any measurable tissue feature such as cell size, subcellular structures, inflammatory cells, lipid droplets, cell proliferation, or cell death can be integrated into TPS in a modular fashion. While we tested and challenged TPS with only a single gene expression readout (Tomato), it is possible to assess any number of expression values, as long as they are annotated with positional information. For example, spatial transcriptomic data, which will produce expression data for thousands of genes per cell/position, will be able to be positionally analyzed as a function of zones using TPS. Although TPS vessel segmentation is developed more specifically for the use case of liver, its ability to extract zonal expression patterns around the landmark structures in a tissue can be generalized to other organ systems and tissue types such as intestines, lungs, and lymph nodes. Thus, TPS can be adapted to other tissue types to facilitate research on biological processes that involve differential spatial patterning of any marker.

Appendix A. Supplementary data
Supplementary data related to this article can be found at https://doi.org/10.1016/j.ebiom.2023.104698.

References
1   Ben-Moshe S, Itzkovitz S. Spatial heterogeneity in the mammalian liver. *Nat Rev Gastroenterol Hepatol*. 2019;16:395–410. https://doi.org/10.1038/s41575-019-0134-x.
2   Birchmeier W. Orchestrating Wnt signalling for metabolic liver zonation. *Nat Cell Biol*. 2016;18:463–465. https://doi.org/10.1038/ncb3349.
3   Rappaport AM, Borowy ZJ, Lougheed WM, Lotto WN. Subdivision of hexagonal liver lobules into a structural and functional unit; role in hepatic physiology and pathology. *Anat Rec*. 1954;119:11–33. https://doi.org/10.1002/ar.1091190103.
4   MacParland SA, Liu JC, Ma XZ, et al. Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. *Nat Commun*. 2018;9:4383. https://doi.org/10.1038/s41467-018-06318-7.
5   Halpern KB, Shenhav R, Matcovitch-Natan O, et al. Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. *Nature*. 2017;542:352–356. https://doi.org/10.1038/nature21065.
6   Aizarani N, Saviano A, Sagar, et al. A human liver cell atlas reveals heterogeneity and epithelial progenitors. *Nature*. 2019;572:199–204. https://doi.org/10.1038/s41586-019-1373-2.
7   Ivanovska T, Schenk A, Homeyer A, et al. A fast and robust hepatocyte quantification algorithm including vein processing. *BMC Bioinformatics*. 2010;11:124. https://doi.org/10.1186/1471-2105-11-124.
8   Ishikawa M, Murakami Y, Ahi ST, et al. Automatic quantification of morphological features for hepatic trabeculae analysis in stained

liver specimens. *J Med Imaging (Bellingham)*. 2016;3:027502. https://doi.org/10.1117/1.JMI.3.2.027502.

9   Tanami S, Ben-Moshe S, Elkayam A, Mayo A, Bahar Halpern K, Itzkovitz S. Dynamic zonation of liver polyploidy. *Cell Tissue Res*. 2017;368:405–410. https://doi.org/10.1007/s00441-016-2427-5.

10 Lin S, Nascimento EM, Gajera CR, et al. Distributed hepatocytes expressing telomerase repopulate the liver in homeostasis and injury. *Nature*. 2018;556:244–248. https://doi.org/10.1038/s41586-018-0004-7.

11 Schwen LO, Homeyer A, Schwier M, et al. Zonated quantification of steatosis in an entire mouse liver. *Comput Biol Med*. 2016;73:108–118. https://doi.org/10.1016/j.compbiomed.2016.04.004.

12 Suneetha I, Chandra Vani D. *Automatic liver tissue section image characterization*. IJIRT; 2021.

13 Budelmann D, Laue H, Weiss N, et al. Automated detection of portal fields and central veins in whole-slide images of liver tissue. *J Pathol Inform*. 2022;13:100001. https://doi.org/10.1016/j.jpi.2022.100001.

14 Wei Y, Wang YG, Jia Y, et al. Liver homeostasis is maintained by midlobular zone 2 hepatocytes. *Science*. 2021;371:eabb1625. https://doi.org/10.1126/science.abb1625.

15 Sankur B. Survey over image thresholding techniques and quantitative performance evaluation. *J Electron Imag*. 2004;13:146. https://doi.org/10.1117/1.1631315.

16 Barnes R, Lehman C, Mulla D. Priority-flood: an optimal depression-filling and watershed-labeling algorithm for digital elevation models. *Comput Geosci*. 2014;62:117–127. https://doi.org/10.1016/j.cageo.2013.04.024.

17 Vincent L, Soille P. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Trans Pattern Anal Mach Intell*. 1991;13:583–598. https://doi.org/10.1109/34.87344.

18 Chen F, Jimenez RJ, Sharma K, et al. Broad distribution of hepatocyte proliferation in liver homeostasis and regeneration. *Cell Stem Cell*. 2020;26:27–33.e4. https://doi.org/10.1016/j.stem.2019.11.001.

19 Fickert P, Stöger U, Fuchsbichler A, et al. A new xenobiotic-induced mouse model of sclerosing cholangitis and biliary fibrosis. *Am J Pathol*. 2007;171:525–536. https://doi.org/10.2353/ajpath.2007.061133.

20 Kietzmann T. Metabolic zonation of the liver: the oxygen gradient revisited. *Redox Biol*. 2017;11:622–630. https://doi.org/10.1016/j.redox.2017.01.012.

21 Gebhardt R, Matz-Soja M. Liver zonation: novel aspects of its regulation and its impact on homeostasis. *World J Gastroenterol*. 2014;20:8491–8504. https://doi.org/10.3748/wjg.v20.i26.8491.

22 Peleman C, De Vos WH, Pintelon I, et al. Zonated quantification of immunohistochemistry in normal and steatotic livers. *Virchows Arch*. 2023;482(6):1035–1045. https://doi.org/10.1007/s00428-023-03496-8.