



MulHiST: Multiple Histological Staining for Thick Biological Samples via Unsupervised Image-to-Image Translation

Lulin Shi, Yan Zhang, Ivy H. M. Wong, Claudia T. K. Lo,
and Terence T. W. Wong^(✉)

Translational and Advanced Bioimaging Laboratory, Department of Chemical and
Biological Engineering, Hong Kong University of Science and Technology,
Hong Kong, China
ttwong@ust.hk

Abstract. The conventional histopathology paradigm can provide the gold standard for clinical diagnosis, which, however, suffers from lengthy processing time and requires costly laboratory equipment. Recent advancements made in deep learning for computational histopathology have sparked lots of efforts in achieving a rapid chemical-free staining technique. Yet, existing approaches are limited to well-prepared thin sections, and invalid in handling more than one stain. In this paper, we present a multiple histological staining model for thick tissues (MulHiST), without any laborious sample preparation, sectioning, and staining process. We use the grey-scale light-sheet microscopy image of thick tissues as model input and transfer it into different histologically stained versions, including hematoxylin and eosin (H&E), Masson's trichrome (MT), and periodic acid-Schiff (PAS). This is the first work that enables the automatic and simultaneous generation of multiple histological staining for thick biological samples. Moreover, we empirically demonstrate that the AdaIN-based generator offers an advantage over other configurations to achieve higher-quality multi-style image generation. Extensive experiments also indicated that multi-domain data fusion is conducive to the model capturing shared pathological features. We believe that the proposed MulHiST can potentially be applied in clinical rapid pathology and will significantly improve the current histological workflow.

Keywords: Virtual staining · Multi-domain image translation · Thick tissues · Light-sheet microscopy

1 Introduction

Histological staining is regarded as the standard protocol in clinical pathological examination, which is used to label biological structures and morphological changes in tissues [1]. The most frequently used histological staining is H&E

stain for the inspection of cell nuclei and the extracellular matrix, in addition to some special stains to complement specific biomarkers and particular structures, such as MT stain used for connective tissues and PAS stain used for mucopolysaccharides [2]. However, multiple tissue sections are required if special stains are desired since the same section cannot be stained several times in conventional pathology workflow. In general, pathologists need to check the H&E-stained images firstly for a basic examination, and then decide whether to prepare additional sections and perform special stains, which will increase the time for diagnosis. More importantly, the abovementioned traditional histochemical staining techniques can only be performed on thin sections of 2–10 μm . Therefore, sample preparation steps, including paraffin embedding, tissue slicing, and chemical dewaxing, will result in long turnaround times and high laboratory infrastructure demands.

The rapidly emerging field of digital virtual staining has shown great promise to revolutionize the decade-old staining workflow. Zhang et al. [3] have done pioneering works on multi-stain translation from unstained thin sections, and Yang et al. [4] also tried to achieve multiple stains generation from label-free tissue images. Both used an image registration to prepare pixel-level matched source-unstained and target-stained image pairs for supervised model training. However, obtaining such pixel-wise aligned data is not accessible for thick tissues as the traditional histological staining can only be performed on thin sections. Even though we can collect the surface cut from the thick specimen and then stain it with chemical reagents, there is still a huge morphological difference due to multiple-layer information captured by a slide-free microscope. Therefore, the virtual staining of thick tissues has to rely on unsupervised methods. There were some primary investigations on virtual staining of thick tissues that use slide-free imaging systems, such as MUSE [5], CHAMP [6], and UV-PAM [7]. However, those methods can only produce virtual H&E-stained images instead of multi-stained images. The emergency of starGAN opens new possibilities for multi-domain image translation [8], and they achieve flexible facial attribute transfer with the proposed domain label. [9, 10] employ the idea of the domain label to represent different staining for multiple histological staining generations. However, those models use H&E staining as input and transfer H&E staining into other stains, which still require laborious tissue embedding and slide sectioning process. [11] focus on the unsupervised multiple virtual staining from autofluorescence images, yet the input should be images obtained from thin slides, which is not ideal for thick tissues.

In this paper, we propose MulHiST, a novel **M**ultiple **H**istological **S**taining model for **T**hick biological tissues, which is not feasible in the traditional histochemical staining workflow. To our knowledge, this is the first attempt to achieve multiple histological staining generations for thick tissues.¹. Our key contributions are: (1) we propose MulHiST: a generative adversarial network (GAN)-based multi-domain image translation model capable of mapping a given

¹ An implementation of MulHiST is available at <https://github.com/TABLAD-HKUST/MulHiST>.

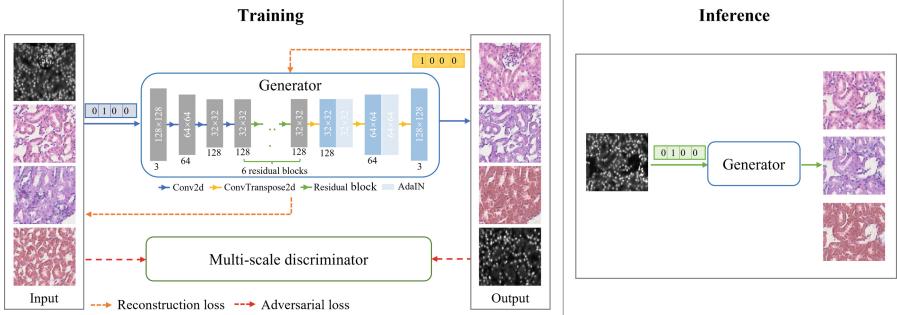


Fig. 1. Overview of the proposed **MulHiST** network.

light-sheet (LS) image of thick tissue into its histologically stained (HS) versions. We utilize unsupervised learning and do not require paired images, tailored for the virtual staining from slide-free imaging techniques; (2) we verify that the multi-domain translation can capture more reliable histopathological features for generating high-quality images, eliminating the ambiguity brought by multiple layers of information in thick tissue images; (3) both qualitative and quantitative results on H&E/PAS/MT staining generations show the superiority and efficiency of the proposed MulHiST over other baseline models.

2 Methodology

To represent different staining types, we follow the idea of the domain-specific attribute vector proposed in [8], aiming to use a one-hot vector c to indicate unstained and various stained domains. During the training, the generator aims to transfer the image style of input image x to the desired domain c while keeping the content of x : $G(x, c) \rightarrow y$, which means y has the same pathological context as x but with a different image style. When the training is finished, we use the well-trained generator to achieve multiple histological stain generations.

In general, the image of thick tissues contains several-layer information, and information from different layers interferes with each other, which will lead to ambiguity in the determination of pathological features, e.g., cell boundary and tissue content. We believe that all the image domains, i.e., LS images and H&E/PAS/MT-stained images, share some domain-invariant features that can facilitate the model training. Moreover, our generator learns the mapping between every two domains. Then, a reconstruction loss can be employed in the single-generator model (orange dashed arrows in Fig. 1). We only need to input the LS image for the inference to get its corresponding histological stained versions (right part in Fig. 1).

Unlike starGAN, we add the style code into the model with the Adaptive Instance Normalization (AdaIN) Layer [12] instead of concatenating the style code with the input image along the channel dimension. We only add the AdaIN

to the decoder module, which will not affect the image feature extraction and encoding of the model input. The parameters of AdaIN are dynamically generated by a multilayer perceptron (MLP) based on the style code. The discriminator used in this work follows the one in [13], providing adversarial feedback to guide the learning of the generator. However, our discriminator will further classify images into different domains, not limited to real or fake signals.

During the training, we incorporate multiple domain images and shuffle the ensembled dataset. A style code c can be generated randomly to indicate the target domain for the generator. The generator will transfer the input image into an image with the target style indicated by c , and the semantic content of the original input will not change. The model forward can be expressed by:

$$G(x, c) = Dec(R(Enc(x)), MLP(c)) \quad (1)$$

where c is the style code that is randomly generated during the training. R is residual block. Enc and Dec are the encoder and decoder in the generator, respectively. The AdaIN layer in the decoder can be computed as:

$$AdaIN(x, s) = \sigma(s) \frac{x - \mu(x)}{\sigma(x)} + \mu(s) \quad (2)$$

where x here is the feature map results of the previous layer, and s is the output of MLP. AdaIN aims to align the mean and variance of the input to match those of the desired style. The overall loss formulation of the generator is:

$$L_G = \lambda_1 L_{adv}^G + \lambda_2 L_{rec} \quad (3)$$

$$L_{rec} = \|x - G(G(x, c_{trg}), c_{org})\| \quad (4)$$

where the L_{adv}^G is the adversarial loss and L_{rec} is the image reconstruction loss. The c_{trg} is the generated target style code and c_{org} is the style code of the original input images.

3 Experimental Results

3.1 Dataset and Implementation Details

In this work, we prepared six thick tissue slabs and obtained scanned images of $\sim 15,000 \times 15,000$ pixels using an open-top LS microscope with an excitation wavelength of 266 nm. After imaging, the specimens were sectioned and histologically processed with the standard protocol to obtain HS images for the model training. We chose one set of scanned images as training data, and the others were used for the testing. For the training, we extracted small image patches with the size of 128×128 randomly. During testing, we divided the tested whole-slide image into 256×256 patches with 16 pixels overlap to avoid artifacts. There are 11,643 extracted image patches in the testing set.

Our model was implemented in PyTorch on a single NVIDIA GeForce RTX 3090 GPU. We trained our model with the Adam optimizer (with $\beta_1 = 0.5$ and $\beta_2 = 0.999$) [14]. The initial learning rate was set to 1×10^{-4} for both generator and discriminator with a linear decay scheduled after 50,000 iterations. The batch size was set to 16. The λ_1 in (3) was set to 5 and λ_2 was set to 10.

3.2 Evaluation Metrics

We quantitatively evaluate our model and results with Kernel Inception Distance (KID) [15] and Fréchet Inception Distance (FID) [16], which are prevalent for evaluating the quality of digitally synthesized images. Unlike natural image generation, biomedical image generation not only require these indicators that reflect the image quality but also require some more convincing metrics with clinical values. As shown in Fig. 3 (5th column, cycleGAN), the image style of generated results is similar to that of the real staining. However, the MT staining generated by cycleGAN is incorrect since the background and tissue content is reversed, which means that sometimes the model can produce target images with high fidelity, but it is difficult to keep correct semantic content or targeted biomarkers. It is not easy to identify pathological features from the LS image, therefore, we need a ground truth for a more reliable comparison. However, it is infeasible to obtain the well-matched ground truth of thick tissues.

In this work, we observed that LS images of thick tissues share a similar style as fluorescence images of thin sections i.e., the cell nuclei are highlighted with positive contrast for both thin and thick tissues with the help of fluorescent labels. In this case, we used the model trained on LS images of thick tissues to test the fluorescence images of thin sections. Then, we could prepare the ground truth of thin sections for comprehensive comparison, as well as quantitative indicators, such as mean square error (MSE), structural similarity (SSIM), and peak signal-to-noise ratio (PSNR). It is worth mentioning that when tested on the thin-section images, the model trained with thick-tissue images would underperform the model trained with thin-section images. Therefore, for virtual staining on thin slices, it is better to train on thin-section images as there are still some differences in details between data from thick tissues and thin sections. Here, we use the thin-section data only for model validation.

3.3 Quantitative and Qualitative Results

In this paper, we select cycleGAN [17], MUNIT [18], and starGAN [8] as baseline models. As shown in Fig. 2, our model surpasses all comparison methods on visual results. From the virtually stained PAS (top zoomed-in regions, yellow square), we can distinguish two different convoluted tubules according to the PAS-positive/negative patterns (yellow/green arrows) in our results. In general, the glomerular and tubular basement membrane, as well as the brush border of the proximal tubules, can be visualized by the PAS staining with pink color, whereas the interior of the distal tubule will not be stained [19]. There is something pink inside the tubules pointed by yellow arrows in our results and no pink area inside the tubules indicated by green arrows. However, from other model results, it is hard to recognize corresponding histopathological features. Here we train cycleGAN multiple times for every pair of source/target domains so that those three cycleGAN models are independent. For such single-domain image translation, it is hard to satisfy all domains with a correct transformation. Specifically, we can see that the H&E-stained images of cycleGAN are correct,

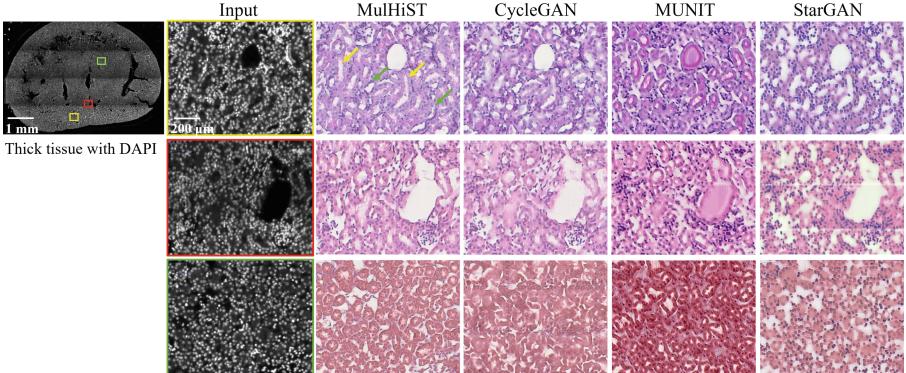


Fig. 2. Virtual generation of PAS-, H&E-, and MT-stained images of thick tissues (from the top to the bottom). Yellow arrows indicate proximal convoluted tubules (positive for PAS), and green ones mean distal convoluted tubules (negative for PAS). (Color figure online)

whereas the model reverses the background and cell nuclei in the MT domain. As shown in Fig. 2, neither MUNIT nor starGAN can perform well in this task.

Moreover, we quantitatively evaluate the model performance with FID and KID scores, as shown in Table 1 (top part). We can observe that our model outperforms the other baseline models significantly in three different image domains. Although the H&E staining of cycleGAN also achieves good FID and KID scores, the corresponding PAS and MT staining results are much worse than ours, which also agrees with the qualitative analysis shown in Fig. 2.

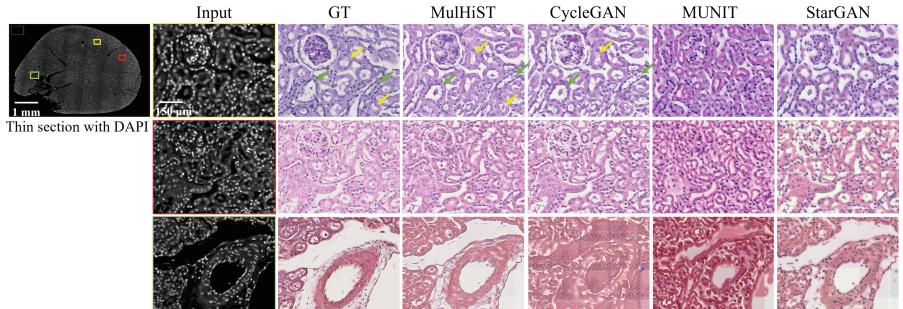
As no ground truth can be provided for the virtual staining of thick tissue. We collected 2 sets of thin mouse kidney sections for the model validation. We used the model trained with LS images of thick tissues to test the scanned images of prepared thin sections. In this situation, we can perform traditional histological staining to obtain ground truth for further comparison. Figure 3 confirms that our MulHiST can generate the correct pathological features, i.e., the PAS-positive proximal convoluted tubules indicated by yellow arrows and PAS-negative distal tubules pointed by green arrows, which are consistent with the ground truth. The same pathological representation can also be observed in the PAS result of cycleGAN, whereas the MT staining generated by cycleGAN presents an obvious error between the background and tissue content. Meanwhile, the quantitative evaluation in Table 1 (bottom part) also shows that our proposed model outperforms other baseline models in various evaluation metrics.

3.4 Ablation Analysis

In this paper, we have two main hypotheses, one is that the model can benefit from data fusion from multiple domains due to the domain-invariant features, and the other is that the AdaIN-based style transfer is better in source image feature extraction compared with channel-wise style code concatenation.

Table 1. Quantitative evaluation results on testing data. The top four rows are tested results of thick tissues, and the following ones are that of thin sections.

	H&E		PAS		MT		MSE	SSIM	PSNR
	FID	KID	FID	KID	FID	KID			
CycleGAN	68.06	0.042	111.29	0.062	150.77	0.096	–	–	–
MUNIT	150.20	0.114	131.93	0.086	128.80	0.098	–	–	–
StarGAN	165.20	0.114	207.67	0.158	208.89	0.180	–	–	–
MulHiST	65.96	0.049	63.20	0.038	67.83	0.036	–	–	–
CycleGAN	74.16	0.044	137.18	0.080	159.24	0.098	2257.73	0.5793	15.90
MUNIT	141.71	0.078	136.23	0.074	118.51	0.067	7749.29	0.4675	9.35
StarGAN	160.91	0.108	185.07	0.127	155.02	0.105	1266.57	0.6811	17.40
MulHiST	69.90	0.039	82.49	0.048	84.44	0.055	1217.36	0.6475	17.50

**Fig. 3.** Virtual staining output of thin section using different models. The 3rd column shows the ground truth (GT) obtained via the traditional histology staining protocol.

We first verify the importance of domain-invariant features shared by multiple domains. The starGAN [8] also reached a similar conclusion that different domains will share the same facial context, which is beneficial to facial expression synthesis. Similarly, there are also some shared features in our multi-domain dataset, such as cell nuclei and cytoplasm membranes. We claim that the model training can benefit from incorporating multiple data domains. From Fig. 4, the synthetic image quality improves with the increase of input data domains. For single-domain translation, the model is sensitive to different domains, where only H&E staining is correct and the other two cannot be determined. This also agrees with the performance of cycleGAN in Fig. 2, 3. The main reason is that the different models are built independently for single-domain image translation, resulting in the inability to share effective information among domains. When adding another domain, the model can correctly translate the background and tissue content. This can be attributed to the feature-sharing between different data domains. In addition, the triple-domain model is superior to the dual-domain ones, where all H&E, PAS, and MT show clear and natural structures.

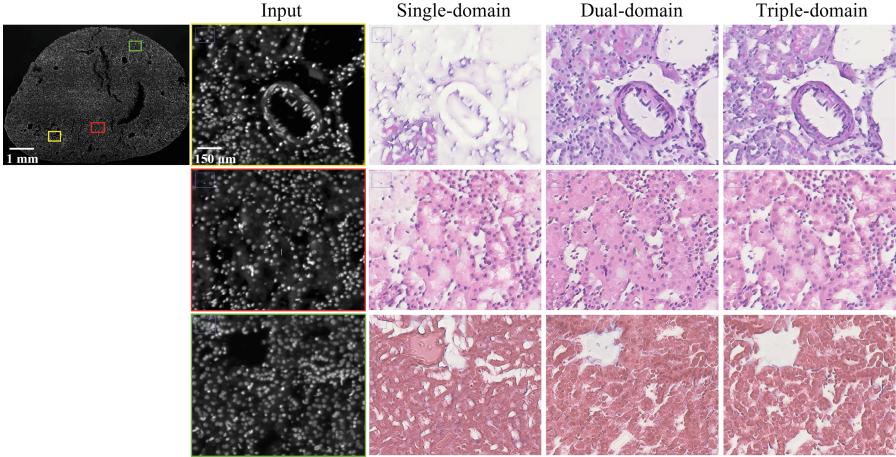


Fig. 4. Efficiency evaluation of multi-domain data fusion. Single-domain means 1-to-1 translation, dual-domain is 2-to-2 translation, and triple-domain refers to 3-to-3 one.

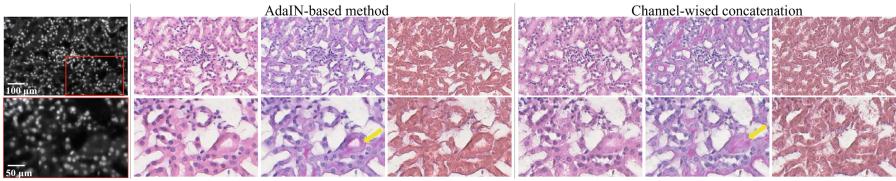


Fig. 5. Comparison of AdaIN-based and channel-wised concatenation-based generator.

Next, we compare the AdaIN-based method and channel-wise concatenation (Fig. 5). We can observe that both ways can achieve correct style transfer, but from the PAS staining, the AdaIN-based method can produce plausible pathological patterns that another model fails (yellow arrows). The PAS should be stained inside the tubules, and the results from the channel-wise concatenation method will cause ambiguity in the judgment by pathologists.

4 Conclusion

This paper proposes a multiple histological image generation model for thick biological samples. This is undesirable in the traditional histopathology workflow as the chemical histological staining should be performed on the thin sections, and the same section cannot be stained with various stains simultaneously. We use slide-free microscopy to capture the thick tissues and translate the scanned images into multiple stained-versions. The model is optimized in an unsupervised manner, fitting the issue of large morphological mismatches between the scanned thick tissue and histologically stained images. Experiment results demonstrated the superiority and the great promise of the proposed method in developing a slide-free, cost-effective, and chemical-free histopathology staining pipeline.

References

1. Gurcan, M.N., Boucheron, L.E., Can, A., Madabhushi, A., Rajpoot, N.M., Yener, B.: Histopathological image analysis: a review. *IEEE Rev. Biomed. Eng.* **2**, 147–171 (2009)
2. Alturkistani, H.A., Tashkandi, F.M., Mohammedsaleh, Z.M.: Histological stains: a literature review and case study. *Glob. J. Health Sci.* **8**(3), 72 (2016)
3. Zhang, Y., de Haan, K., Rivenson, Y., Li, J., Delis, A., Ozcan, A.: Digital synthesis of histological stains using micro-structured and multiplexed virtual staining of label-free tissue. *Light: Sci. Appl.* **9**(1), 78 (2020)
4. Yang, X., et al.: Virtual stain transfer in histology via cascaded deep neural networks. *ACS Photonics* **9**(9), 3134–3143 (2022)
5. Chen, Z., Yu, W., Wong, I.H., Wong, T.T.: Deep-learning-assisted microscopy with ultraviolet surface excitation for rapid slide-free histological imaging. *Biomed. Opt. Exp.* **12**(9), 5920–5938 (2021)
6. Zhang, Y., et al.: High-throughput, label-free and slide-free histological imaging by computational microscopy and unsupervised learning. *Adv. Sci.* **9**(2), 2102358 (2022)
7. Cao, R., et al.: Label-free intraoperative histology of bone tissue via deep-learning-assisted ultraviolet photoacoustic microscopy. *Nat. Biomed. Eng.* 1–11 (2022)
8. Choi, Y., Choi, M., Kim, M., Ha, J.W., Kim, S., Choo, J.: Stargan: unified generative adversarial networks for multi-domain image-to-image translation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8789–8797 (2018)
9. Zhang, R., et al.: Mvfstain: multiple virtual functional stain histopathology images generation based on specific domain mapping. *Med. Image Anal.* **80**, 102520 (2022)
10. Lin, Y., et al.: Unpaired multi-domain stain transfer for kidney histopathological images. *Proc. AAAI Conf. Artif. Intell.* **36**, 1630–1637 (2022)
11. Shi, L., Wong, I.H., Lo, C.T., WK, T.L., Wong, T.T.: Unsupervised multiple virtual histological staining from label-free autofluorescence images. In: *Proceedings of the IEEE International Symposium on Biomedical Imaging* (2023)
12. Huang, X., Belongie, S.: Arbitrary style transfer in real-time with adaptive instance normalization. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1501–1510 (2017)
13. Wang, T.C., Liu, M.Y., Zhu, J.Y., Tao, A., Kautz, J., Catanzaro, B.: High-resolution image synthesis and semantic manipulation with conditional gans. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8798–8807 (2018)
14. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. In: *Proceedings of International Conference on Learning Representations* (2015)
15. Bińkowski, M., Sutherland, D.J., Arbel, M., Gretton, A.: Demystifying mmd gans. *arXiv preprint arXiv:1801.01401* (2018)
16. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Adv. Neural Inf. Process. Syst.* **30** (2017)

17. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2223–2232 (2017)
18. Huang, X., Liu, M.Y., Belongie, S., Kautz, J.: Multimodal unsupervised image-to-image translation. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 172–189 (2018)
19. Kong, W., Haschler, T.N., Nürnberg, B., Krämer, S., Collasch, M., Markó, L.: Renal fibrosis, immune cell infiltration and changes of trpc channel expression after unilateral ureteral obstruction in trpc6-/-mice. *Cell. Physiol. Biochem.* **52**, 1484–1502 (2019)