Autofluorescence microscopy as a label-free tool for renal histology and glomerular segmentation

Nathan Heath Patterson[1,2], Elizabeth K. Neumann[1,2], Kavya Sharman[1,2,3], Jamie Allen[1,2], Raymond Harris[5,6], Agnes B. Fogo[7,8], Mark de Caestecker[9], Richard M. Caprioli[1,2,4], Raf Van de Plas[1,2,9], Jeffrey M. Spraggins[1,2,4,11*]

[1]Department of Biochemistry, Vanderbilt University, Nashville, TN, USA 37232
[2]Mass Spectrometry Research Center, Vanderbilt University, Nashville, TN, USA 37232.
[3]Program in Chemical and Physical Biology, Vanderbilt University School of Medicine, Nashville, TN 37232, USA
[4]Department of Chemistry, Vanderbilt University, Nashville, TN, USA 37232
[5]Division of Nephrology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN USA 37232
[6]Vanderbilt Center for Kidney Disease, Vanderbilt University Medical Center, Nashville, TN USA 37232
[7]Department of Pathology, Microbiology, and Immunology, Vanderbilt University Medical Center, Nashville, TN USA 37232
[8]Department of Medicine and Pediatrics, Vanderbilt University Medical Center, Nashville, TN USA 37232
[9]Division of Nephrology and Hypertension, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN USA 37232
[10]Delft Center for Systems and Control, Delft University of Technology, Delft, Netherlands
[11]Department of Cell & Developmental Biology, Vanderbilt University, Nashville, TN USA 37232

* Corresponding author: jeff.spraggins@vanderbilt.edu

ABSTRACT

Functional tissue units (FTUs) composed of multiple cells like the glomerulus in the kidney nephron play important roles in health and disease. Histological staining is often used for annotation or segmentation of FTUs, but chemical stains can introduce artefacts through experimental factors that influence analysis. Secondly, many molecular *-omics* techniques are incompatible with common histological stains. To enable FTU segmentation and annotation in human kidney without the need for histological staining, we detail here the use of widefield autofluorescence (AF) microscopy as a simple, label-free modality that provides detailed renal morphology comparable to periodic acid-Schiff (PAS) stained tissue in both formalin-fixed paraffin-embedded (FFPE) and fresh frozen samples and with no tissue processing beyond sectioning. We demonstrate automated deep learning-based glomerular unit recognition and segmentation on PAS and AF images of the same tissue section from 9 fresh frozen samples and 9 FFPE samples. All training comparisons were carried out using registered AF microscopy and PAS stained whole slide images originating from the same section, and the recognition models were built with the exact same training and test examples. Measures of recognition performance, such as the Dice-Sorensen coefficient, the true positive rate, and the positive predictive value differed less than 2% between standard PAS and AF microscopy for both preservation methods. These results demonstrate that AF is a potentially powerful tool to study human kidney tissue, that it can serve as a label-free source for automated and manual annotation of tissue structures.

INTRODUCTION

Recent kidney atlas programs have emerged with the goal of characterizing normal human tissues and discovering drivers of disease using cellular and molecular phenotypes spanning wide spatial scales and biomolecular classes.[1,2] Functional tissue units (FTUs), multicellular structures that perform vital functions, including components of the nephron in the kidney, such as glomeruli, proximal tubules, distal tubules, and collecting ducts, are an important intermediate between whole organ and individual cells.[3] Indeed, kidney diseases like diabetic nephropathy, acute kidney injury, and chronic kidney disease are a direct result of damage to the aforementioned FTUs.[4–6] Efficiently and accurately identifying these FTUs for exploration, and being able to perform this recognition task at scale, is critical for linking their location to *-omics* modalities and characterizing the molecular signatures specific to these FTUs.

Digitization of microscopy slides as whole slide images (WSIs) provides the histological data necessary to identify and assess FTUs in tissue sections for basic research and histopathology.[7] The proliferation of kidney WSIs in renal pathology has led to the development of computational approaches for automated analysis and interpretation.[8–10] Deep learning approaches have achieved state-of-the-art results in image analysis tasks (classification, object detection, segmentation)[11] and perform remarkably well in complex tasks like segmentation followed by diagnosis.[12]

In most studies incorporating WSIs in kidney, traditional histological stains are used.[10] While versatile and widely understood, staining can be variable due to experimental factors like reagents and protocols that introduce artefacts[13], which in turn affects reproducibility in downstream computational analysis[14] with that effect potentially magnified across multi-center studies. Further, histological stains are not always compatible with or diminish the quality of other informative techniques that probe molecular biology, e.g. transcriptomics or proteomics, as the chemical reagents used in the staining protocol may irreversibly alter the molecular content of the sample. Instead, tissue section imaging approaches that can achieve subcellular resolution, provide high quality morphology, induce minimal tissue degradation, and are also stain/label-free have particular potential to enable more reproducible WSI studies and more robust annotation results. Further, label-free techniques safe-guard the ability to perform measurements with molecular analysis techniques on the exact same tissue section that has been annotated.

We demonstrate herein the use of one such imaging modality, namely multi-channel widefield autofluorescence microscopy (AF), as a complement to standard histological morphology of periodic acid–Schiff (PAS) staining in human kidney tissue sections. We describe qualitative comparisons between AF and PAS and compare the two modalities for automated deep learning-based detection of the glomerular FTU using a cohort of fresh frozen and formalin fixed paraffin embedded (FFPE) human kidney tissue sections.

RESULTS

AF and PAS morphological comparison

Figure 1 is a side-by-side example of registered PAS and AF images from an FFPE block at low and high magnification, showing AF readily distinguishes both gross (Fig. 1 - A&B) and microscopic morphology with many FTUs qualitatively distinguishable by either method (Fig. 1 - C-H).

Comparison of performance of PAS versus AF for glomerulus unit segmentation by deep learning

Figure 2 outlines the data preparation and analysis workflow to compare the performance of PAS versus AF for glomerular detection using random, but matched-across-modalities sampling of glomerular unit annotations (fresh frozen - n=1612 , FFPE - n=2480). for training, 10 initial randomizations were made creating 10 models per subset of 1, 5, 10, 15, and 25% of the data (50 models/modality). Each subset after random initialization is cumulative, i.e., random initialization 1 with 10% of annotations includes all glomeruli in the 5% subset. The performance of the models on an independent test set (the glomeruli excluded from training) inclusive of all glomeruli not in the training set is shown in Figure 3. The segmentation accuracy was compared using the Dice-Sorensen coefficient[15], a method that computes the overlap of the ground truth annotation and the predicted segmentation masks. To assess the ability to detect and accurately segment glomerular units, predictions were considered true positives if their Dice-Sorensen coefficient was >0.75. The positive predictive value assessed the amount of false positive detections.

Glomerular segmentation in whole slide images and histological staining artefacts

Figure 4 A shows glomerular unit segmentation by CNN across a fresh frozen example WSI, showing a CNN trained on AF data's adaptability to large microscopy images typical of pathology. The AF images are collected prior to any processing of the tissue sample where artefacts may be introduced. In Figure 4 B & C, a glomerulus is shown that exhibits artefacts introduced by histological staining. The glomerulus is shown intact in the AF image (Fig 4B), which takes place pre-PAS staining, and with the tuft area partially destroyed by the staining process in the PAS image (Fig 4C).

DISCUSSION

Considered on its own merits, AF is a straightforward and label-free microscopy approach, requiring no tissue block processing beyond sectioning (and deparaffinization in the case of FFPE blocks). Gross morphology such as renal papilla, medulla, and cortex can be distinguished in AF images through changes in both fluorescence intensity and image texture (Fig. 1 A&B). Furthermore, FTUs like glomerular units and tubules can also be qualitatively distinguished by AF (Fig. 1 C-H), demonstrating broad potential for driving kidney annotation and analysis.

Using a spatially matched dataset of AF and PAS WSIs capturing identical tissue morphology and pulled from both fresh frozen and FFPE blocks (n=9 for each), we compare the performance a

CNN-driven method for automated recognition of glomerular units between using AF images versus PAS stained images as inputs. The recognition scenarios yielded less than 2% difference between the PAS and AF inputs when using at least 5% of the total data available (80 & 124 annotations in FF and FFPE cohorts, respectively), as reported by the Dice-Sorensen coefficient, the true positive rate, and the positive predictive value (Figure 3). Further, none of the differences were determined to be statistically significant using a threshold of 0.05. While similar performance was noted using a small training set, using 25% of the available annotations as a training set (leaving 75% of the ground truth glomeruli as a test set for evaluation) and using over 10 random selections of the training data, differences were <1% between PAS and AF for all metrics. AF inputs performed slighter poorer in comparison to PAS inputs when using a very limited training set of <50 annotations, which corresponds to 1% of available data. This may be due to increased morphological heterogeneity captured by AF compared to PAS, such as in schlerotic glomeruli where there is a difference in PAS, but the difference is more stark in AF as shown in Figure 5. This does indicate the need for a slightly larger training set in AF prediction versus PAS, but our modality-specific models achieved near parity at 10% of the annotations (160 and 248 training examples in FF and FFPE, respectively). Fresh frozen samples also showed a worse performance for positive predictive value, indicating more false positives in both PAS and AF; this is likely due to the relatively poorer quality of the morphology in fresh frozen tissue sections compared to FFPE sections.

Using AF to automatically map the position of each glomerulus in the WSI without altering the tissue section creates new possibilities for highly automated sample collection in microscopy-guided *-omics* analyses such as laser capture microdissection[16] or liquid extraction surface analysis[17]. As shown in Figure 4, AF has less risk of morphological artefacts that would prevent or limit analysis of FTUs with these technologies.

Conceptually similar work using virtual staining has recently been employed to predict stained images from unstained AF images[18], but only recently have virtual stainings been used for computational analysis.[19] Our work shows the possibility to employ virtual staining for qualitative evaluations where useful, but use the underlying, untransformed, and empirically measured AF data for computational analysis. This effectively shortcuts the need to first transform the data in a virtual stain and mitigates uncertainty around the transformation. However, in reiteration, as AF is label-free, both approaches can be deployed simultaneously.

Finally, AF's compatibility with traditional stains and immunohistochemistry permits facile transfer of knowledge simply through registration of the images as demonstrated herein. One example of this is that training data for tissue structure recognition models can be rapidly developed by acquiring ground truth annotations on registered and established stained image types while building machine learning models using AF measurements as inputs. This would enable easy transfer of automated annotation capability to AF images while maintaining a firm grasp on how well the performance of AF-based computational models compares to ground truth annotations on established stained images. This strategy alongside the instrumental and experimental simplicity make AF microscopy an attractive label-free modality that can complement standard histopathology and also serve as a robust and reproducible spatial map for advanced molecular experiments downstream. Further, while AF is explored herein, it may not be alone as the only

label-free microscopy modality that can serve for FTU segmentation, as many other modalities may be overlooked because the visual information is unappealing. However, computationally, these modalities may provide competitive results in analysis tasks.

With the public release of this registered AF-PAS dataset through the HuBMAP Data Portal (https://hubmapconsortium.com), we hope this serves the kidney community in assessing AF for their own purposes and opens the door to more expansive computational developments of AF for FTU segmentation.

METHODS

Sample curation and microscopy

Normal portions of renal cancer nephrectomies from adult patients were studied. Tissue blocks were collected from areas distant to the tumor.

Microscopy and image registration

Images were collected from the entire tissue section using brightfield and epi-fluorescence modes. For registration, images were imported into *wsireg* (https://github.com/nhpatterson/wsireg) for registration. WSI registration was carried out as described elsewhere.[20] Further details on microscopy settings and image processing are available in the supplemental information.

Whole slide image annotation

Registered WSIs were imported into QuPath[21] and glomerular annotations were made with the polygon tool. Annotations were reviewed for accuracy by kidney histology experts.

Deep learning

Transfer learning was employed using a Mask-RCNN (101 ResNet backbone) convolutional neural network implemented in the *detectron2* package. Additional method details for deep learning are available in the supplemental information.

Data analysis

Detection and segmentation results were evaluated using true positive rate (TPR), positive predictive value (PPV), and Dice-Sorensen Coefficient (DSC). Each metric is a value between 0 and 1 that indicates the agreement between the trained model and the ground truth glomerular unit annotation. TPR and PPV metrics describe the number (true positives), incorrect detections (false positives), and missing detections (false negatives). DSC is the pixel-wise spatial measure that computes how well the pixels belonging to the glomerular unit are segmented by the model[22]. DSC values of 0 indicate the model did not detect a glomerular unit (false negative) and 1 would indicate a perfect match between model segmentation and ground truth annotation. DSC values >0.75 were used to determine correct segmentations (true positives) in TPR and PPV calculations true positive glomeruli were well-segmented.

Data availability

The data described in this paper is available through the HuBMAP Data Portal (https://hubmapconsortium.com).

DISCLOSURES

All authors declare no competing interests.

ACKNOWLEDGEMENTS

REFERENCES

1. HuBMAP Consortium. The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. *Nature*. 2019;574(7777):187-192. doi:10.1038/s41586-019-1629-x

2. El-Achkar TM, Eadon MT, Menon R, et al. A multimodal and integrated approach to interrogate human kidney biopsies with rigor and reproducibility: guidelines from the Kidney Precision Medicine Project. *Physiological Genomics*. 2020;53(1):1-11. doi:10.1152/physiolgenomics.00104.2020

3. de Bono B, Grenon P, Baldock R, Hunter P. Functional tissue units and their primary tissue motifs in multi-scale physiology. *J Biomed Semantics*. 2013;4(1):22. doi:10.1186/2041-1480-4-22

4. Liu B-C, Tang T-T, Lv L-L, Lan H-Y. Renal tubule injury: a driving force toward chronic kidney disease. *Kidney Int*. 2018;93(3):568-579. doi:10.1016/j.kint.2017.09.033

5. Basile DP, Anderson MD, Sutton TA. Pathophysiology of Acute Kidney Injury. *Compr Physiol*. 2012;2(2):1303-1353. doi:10.1002/cphy.c110041

6. Tung C-W, Hsu Y-C, Shih Y-H, Chang P-J, Lin C-L. Glomerular mesangial cell and podocyte injuries in diabetic nephropathy. *Nephrology*. 2018;23(S4):32-37. doi:https://doi.org/10.1111/nep.13451

7. Hanna MG, Reuter VE, Hameed MR, et al. Whole slide imaging equivalency and efficiency study: experience at a large academic center. *Mod Pathol*. 2019;32(7):916-928. doi:10.1038/s41379-019-0205-0

8. Hermsen M, de Bel T, den Boer M, et al. Deep Learning-Based Histopathologic Assessment of Kidney Tissue. *J Am Soc Nephrol*. 2019;30(10):1968-1979. doi:10.1681/ASN.2019020144

9. Jayapandian CP, Chen Y, Janowczyk AR, et al. Development and evaluation of deep learning–based segmentation of histologic structures in the kidney cortex with multiple histologic stains. *Kidney International*. 2021;99(1):86-101. doi:10.1016/j.kint.2020.07.044

10. Huo Y, Deng R, Liu Q, Fogo AB, Yang H. AI applications in renal pathology. *Kidney International*. 2021;0(0). doi:10.1016/j.kint.2021.01.015

11. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-444. doi:10.1038/nature14539

12. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *JAMA*. 2017;318(22):2199-2210. doi:10.1001/jama.2017.14585

13. Rastogi V, Puri N, Arora S, Kaur G, Yadav L, Sharma R. Artefacts: a diagnostic dilemma - a review. *J Clin Diagn Res*. 2013;7(10):2408-2413. doi:10.7860/JCDR/2013/6170.3541

14. Tellez D, Litjens G, Bándi P, et al. Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. *Medical Image Analysis*. 2019;58:101544. doi:10.1016/j.media.2019.101544

15. Dice LR. Measures of the Amount of Ecologic Association Between Species. *Ecology*. 1945;26(3):297-302. doi:https://doi.org/10.2307/1932409

16. Satoskar A, Shapiro J, Bott C, et al. Characterization of Glomerular Diseases Using Proteomic Analysis of Laser Capture Microdissected Glomeruli. *Mod Pathol*. 2012;25(5):709-721. doi:10.1038/modpathol.2011.205

17. Ryan DJ, Patterson NH, Putnam NE, et al. MicroLESA: Integrating Autofluorescence Microscopy, In Situ Micro-Digestions, and Liquid Extraction Surface Analysis for High Spatial Resolution Targeted Proteomic Studies. *Anal Chem*. 2019;91(12):7578-7585. doi:10.1021/acs.analchem.8b05889

18. Rivenson Y, Wang H, Wei Z, et al. Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning. *Nat Biomed Eng*. 2019;3(6):466-477. doi:10.1038/s41551-019-0362-y

19. de Haan K, Zhang Y, Liu T, et al. Deep learning-based transformation of the H&E stain into special stains improves kidney disease diagnosis. *arXiv:200808871 [physics]*. Published online August 20, 2020. Accessed February 1, 2021. http://arxiv.org/abs/2008.08871

20. Patterson NH, Tuck M, Van de Plas R, Caprioli RM. Advanced Registration and Analysis of MALDI Imaging Mass Spectrometry Measurements through Autofluorescence Microscopy. *Anal Chem*. 2018;90(21):12395-12403. doi:10.1021/acs.analchem.8b02884

21. Bankhead P, Loughrey MB, Fernández JA, et al. QuPath: Open source software for digital pathology image analysis. *Sci Rep*. 2017;7(1):16878. doi:10.1038/s41598-017-17204-5

22. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med Imaging*. 2015;15. doi:10.1186/s12880-015-0068-x
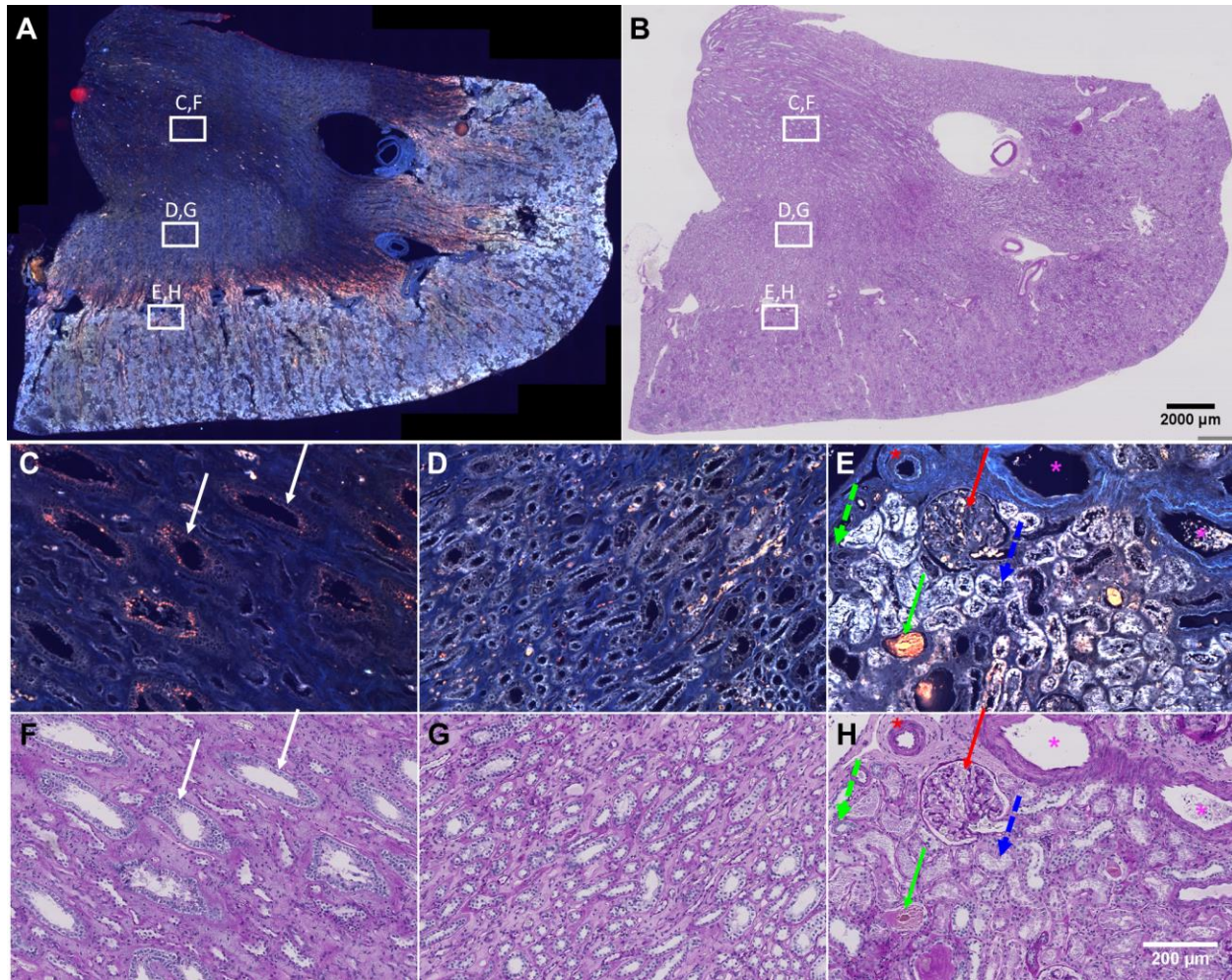
FIGURES

Figure 1. Morphology of FFPE human kidney section reported by autofluorescence and PAS stain images. A: Low magnification view of autofluorescence microscopy. B: Low magnification view of PAS staining. C, F: High magnification view of an area of the renal papilla with distal collecting ducts visible (thin white arrows). D, G: High magnification view of an area of the inner stripe of the outer medulla. E.H: High magnification view of the juxtamedullary renal cortex with a glomerulus (thin red arrow), intratubular casts (thin solid green arrow), arcuate arteries (red asterisk), arcuate veins (purple asterisks), proximal tubules (dashed blue arrow), and distal tubules (dashed thick green arrow) visible.
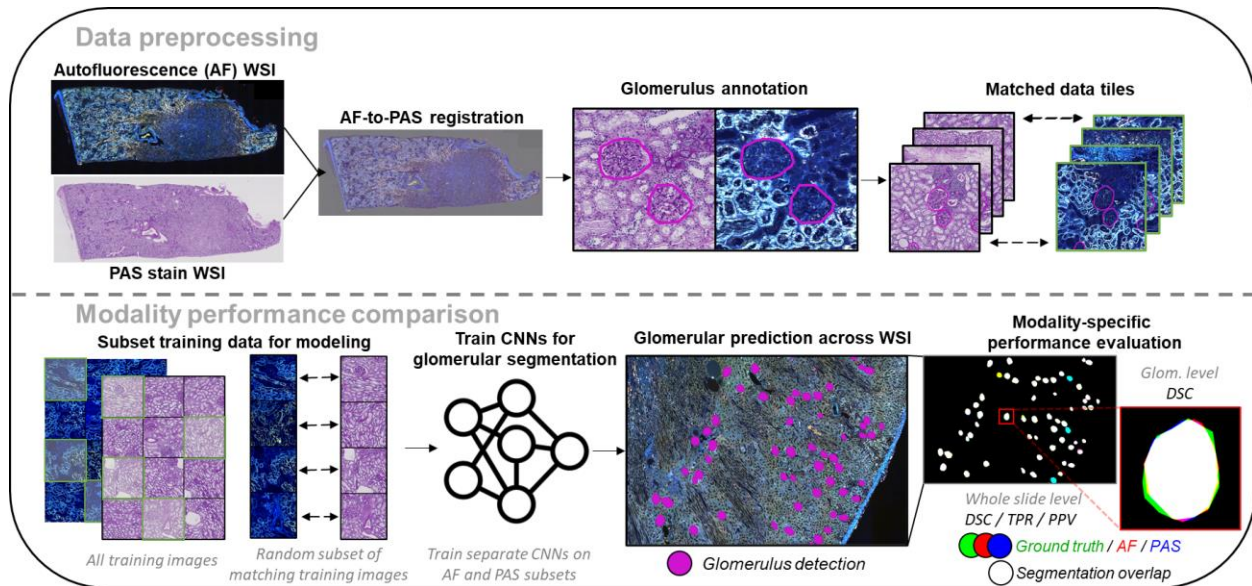
Figure 2. Workflow for performance comparison of glomerulus segmentation using spatially matched training data from PAS and AF. *Data preprocessing - Top*: Data preprocessing and preparation to sets of matched data tiles after image registration of the PAS and AF WSIs. First, WSI images are registered and then glomeruli are annotated. The WSIs and corresponding annotations are split into tiles for use in CNN modeling to fit in GPU memory. *Modality performance comparison - Bottom*: The split tiles are subset into training data with the random selection being the same for AF and PAS data to ensure modeling on the same exact set of spatially matched images and annotations. WSI detection occurs by tiling the unknown image, segmenting glomeruli, then stitching the resulting glomerular predictions back onto the WSI. In modality-specific performance evaluation, *ground truth* is the manually annotated glomerular units and *AF* and *PAS* are the CNN predictions. WSI- whole slide image. AF- autofluorescence. CNN - Convolutional neural network. DSC- Dice-Sorensen coefficient. TPR - true positive rate, PPV - positive predictive value.
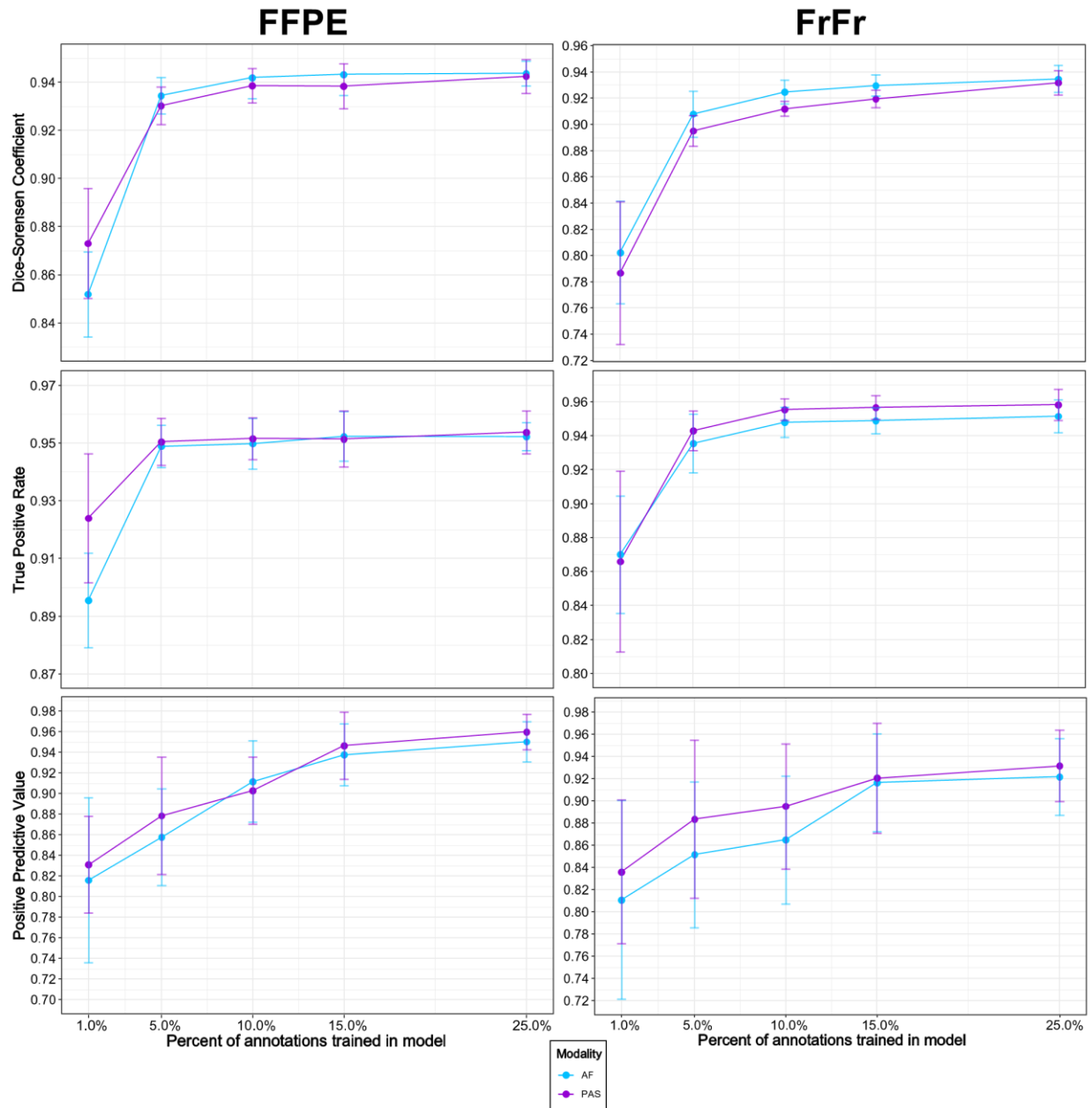
Figure 3. Performance comparison between AF and PAS images as basis for glomerular unit segmentation , reported through the Dice-Sorensen coefficient, true positive rate, and positive predictive value in function of proportion of total training data, both in FFPE and fresh frozen sample cohorts. In each case, 10 random initializations for developing subsets for training data of glomerular unit annotations (fresh frozen - n=1612 , FFPE - n=2480) were made to prevent potential bias in performance due to outlier samples. The average (point) and standard deviation (error bars) are determined from the performance of the 10 models on the proportion of the data that wasn't used in training. AF- autofluorescence. FrFr - fresh frozen. FFPE - formalin fixed paraffin embedded.
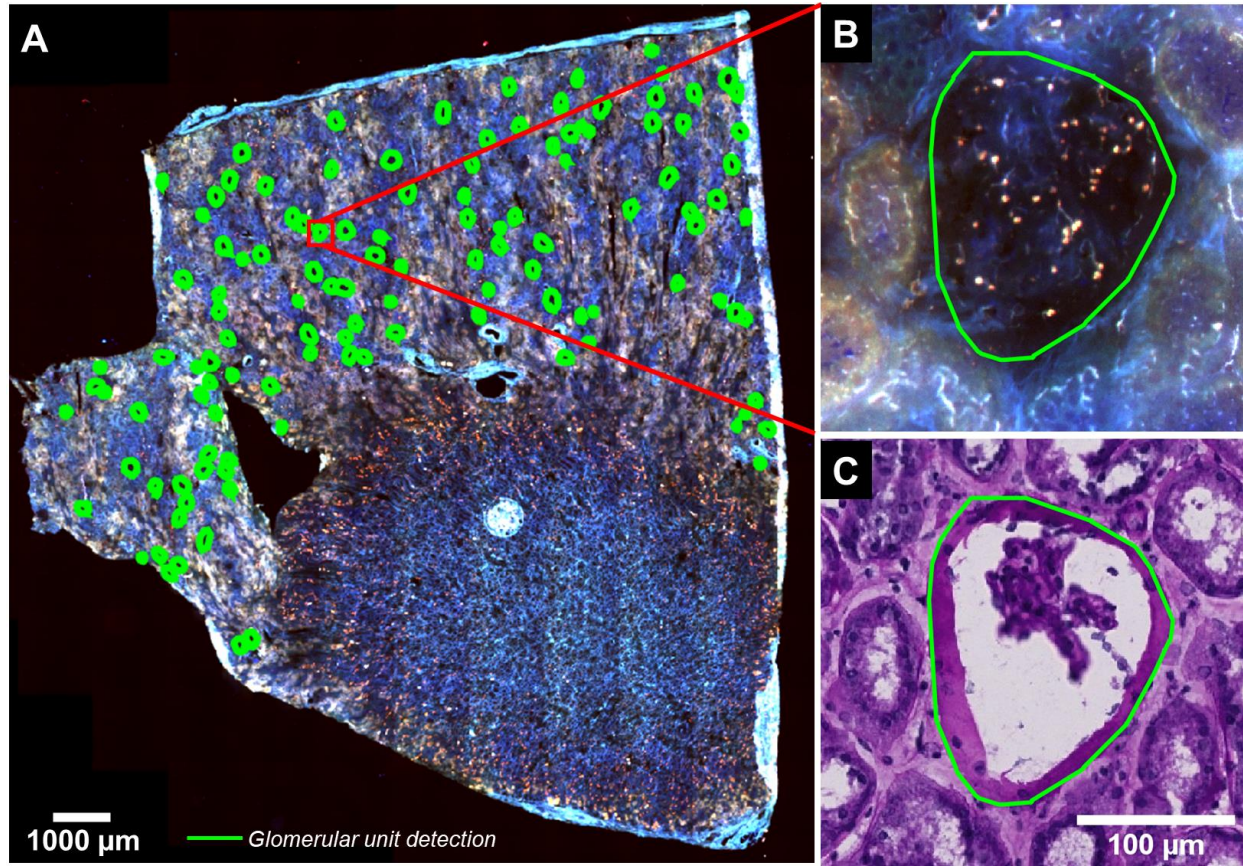
Figure 4. Glomerular unit segmentation in a fresh frozen AF whole slide image and a glomerular tuft artefact from PAS staining. A. Low magnification view of fresh frozen AF WSI of normal human kidney with green contours outlining predicted glomerular unit segmentation borders. The red box indicates the glomerulus shown in panels B and C. B. High magnification of the selected glomerulus, showing intact tuft filling the glomerular area. C. The same glomerulus after PAS staining, showing an artefact of the staining process where the glomerular tuft is lost. Artefacts such as this highlight the morphological fidelity benefits that can be obtained if staining can be avoided and annotation can be accomplished directly from a low-sample preparation modality such as AF. PAS - periodic-acid schiff stain. AF - autofluorescence. WSI - whole slide image.
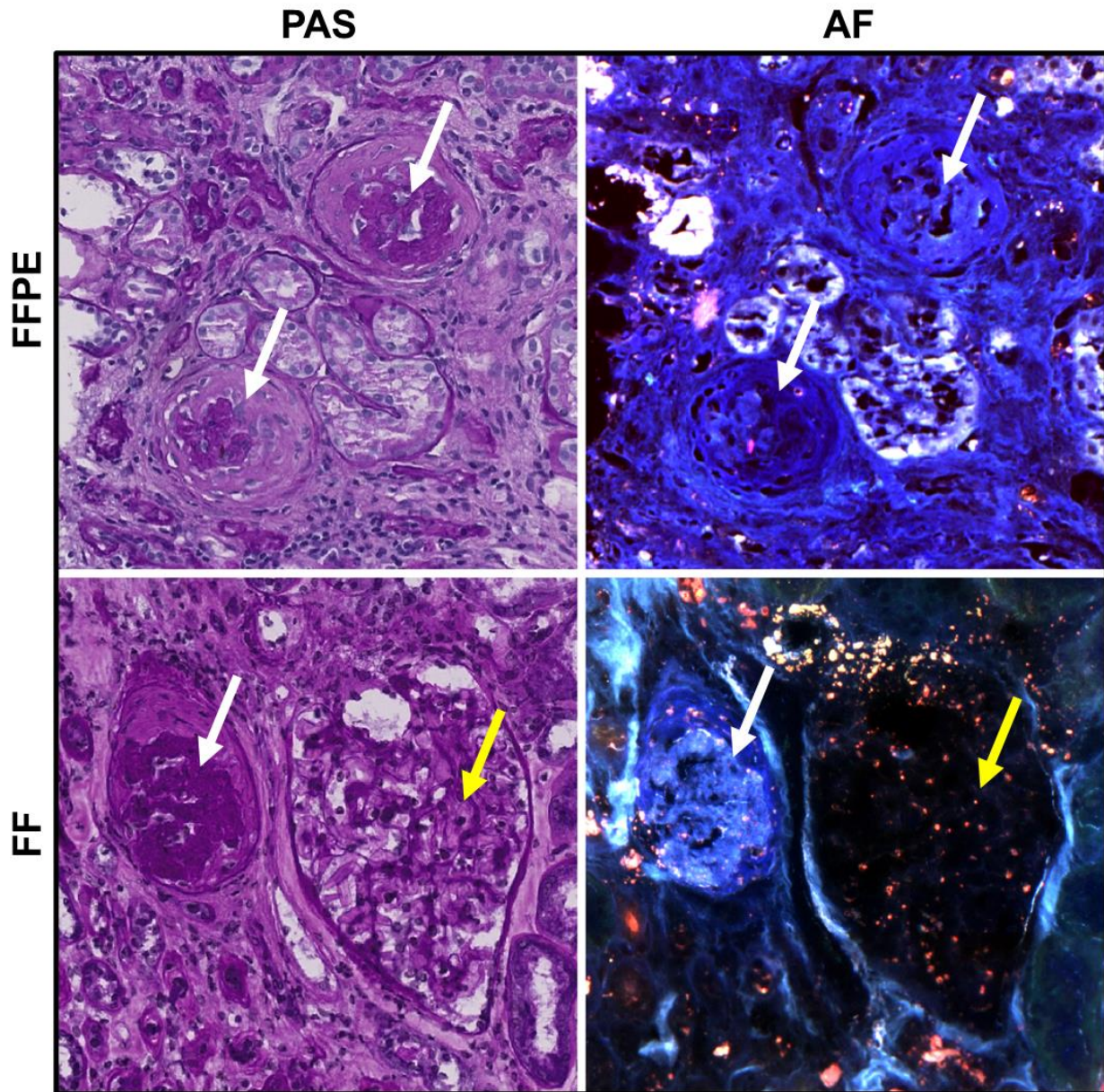
Figure 5. Differentiation of sclerotic glomeruli by autofluorescence. White arrows point to sclerotic glomeruli. The yellow arrow points to a nearby glomerulus without sclerosis in a fresh frozen tissue example. AF captures a great deal of glomerular heterogeneity in FF. The stronger blue signal in the bottom right panel corresponds to stronger staining at shorter wavelengths and represents the accumulation of elastin and collagen (both autofluorescent proteins) typical of glomerular sclerosis. FFPE - formalin fixed paraffin embedded. FF - fresh frozen. PAS - periodic-acid schiff stain. AF - autofluorescence.