PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

SHIFT: speedy histopathological-to-immunofluorescent translation of whole slide images using conditional generative adversarial networks

Erik A. Burlingame, Adam Margolin, Joe W. Gray, Young Hwan Chang

Erik A. Burlingame, Adam Margolin, Joe W. Gray, Young Hwan Chang, "SHIFT: speedy histopathological-to-immunofluorescent translation of whole slide images using conditional generative adversarial networks," Proc. SPIE 10581, Medical Imaging 2018: Digital Pathology, 1058105 (6 March 2018); doi: 10.1117/12.2293249

SPIE.

Event: SPIE Medical Imaging, 2018, Houston, Texas, United States

SHIFT: speedy histopathological-to-immunofluorescent translation of whole slide images using conditional generative adversarial networks

Erik A. Burlingame, Adam A. Margolin, Joe W. Gray, and Young Hwan Chang

Oregon Health and Science University, Portland, OR, USA

ABSTRACT

Multiplexed imaging such as multicolor immunofluorescence staining, multiplexed immunohistochemistry (mIHC) or cyclic immunofluorescence (cycIF) enables deep assessment of cellular complexity *in situ* and, in conjunction with standard histology stains like hematoxylin and eosin (H&E), can help to unravel the complex molecular relationships and spatial interdependencies that undergird disease states. However, these multiplexed imaging methods are costly and can degrade both tissue quality and antigenicity with each successive cycle of staining. In addition, computationally intensive image processing such as image registration across multiple channels is required. We have developed a novel method, speedy histopathological-to-immunofluorescent translation (SHIFT) of whole slide images (WSIs) using conditional generative adversarial networks (cGANs). This approach is rooted in the assumption that specific patterns captured in IF images by stains like DAPI, pan-cytokeratin (panCK), or α-smooth muscle actin (α-SMA) are encoded in H&E images, such that a SHIFT model can learn useful feature representations or architectural patterns in the H&E stain that help generate relevant IF stain patterns. We demonstrate that the proposed method is capable of generating realistic tumor marker IF WSIs conditioned on corresponding H&E-stained WSIs with up to 94.5% accuracy in a matter of seconds. Thus, this method has the potential to not only improve our understanding of the mapping of histological and morphological profiles into protein expression profiles, but also greatly increase the efficiency of diagnostic and prognostic decision-making.

Keywords: deep learning, conditional generative adversarial network, image translation, digital pathology

1. INTRODUCTION

The clinical management of many systemic diseases, including cancer, is informed by histopathological evaluation of biopsy tissue, wherein thin sections of the biopsy are processed to visualize tissue and cell morphologies for signs of disease. Though H&E remains the gold standard stain in such evaluations for many cancer types, additional staining by immunofluorescence or immunohistochemistry can augment pathologist interpretation, as it allows for specific targeting and visualization of clinically relevant biomolecules and cell subtypes. Moreover, the recent development of multiplexed imaging such as cycIF,[1] mIHC,[2,3] and other multiplex methods in histopathology[4–6] have greatly expanded the palette with which pathologists can visualize individual tissue sections. This allows for deep *in situ* assessment of the complexities of the tumor microenvironment, e.g. through examination of the spatial interactions and architectural organization of tumor and non-tumor cells.

However, these multiplexed imaging methods are time- and resource-intensive and suffer from technical challenges related to tissue and antigen degradation.[7] If the assumption holds that the information required to accurately infer the distribution of specific protein abundance is already encoded in an H&E-stained WSI, i.e. tissue and cell morphologies displayed in histopathological images are a function of underlying molecular drivers,[8] then it should be possible to faithfully infer an IF or IHC stain conditioned on the H&E stain. Here we propose SHIFT, a method which leverages a cGAN framework[9,10] to efficiently translate H&E WSIs into realistic IF WSIs (Figure 1), and demonstrate unit test translations of panCK, α-SMA, and DAPI, two commonly used prognostic markers and a nuclear counterstain, respectively. Furthermore, this framework can be used to test

Further author information: (Send correspondence to Y.H.C.)
E.A.B.: E-mail: burlinge@ohsu.edu
Y.H.C.: E-mail: chanyo@ohsu.edu

Medical Imaging 2018: Digital Pathology, edited by John E. Tomaszewski, Metin N. Gurcan,
Proc. of SPIE Vol. 10581, 1058105 · © 2018 SPIE · CCC code:
1605-7422/18/$18 · doi: 10.1117/12.2293249

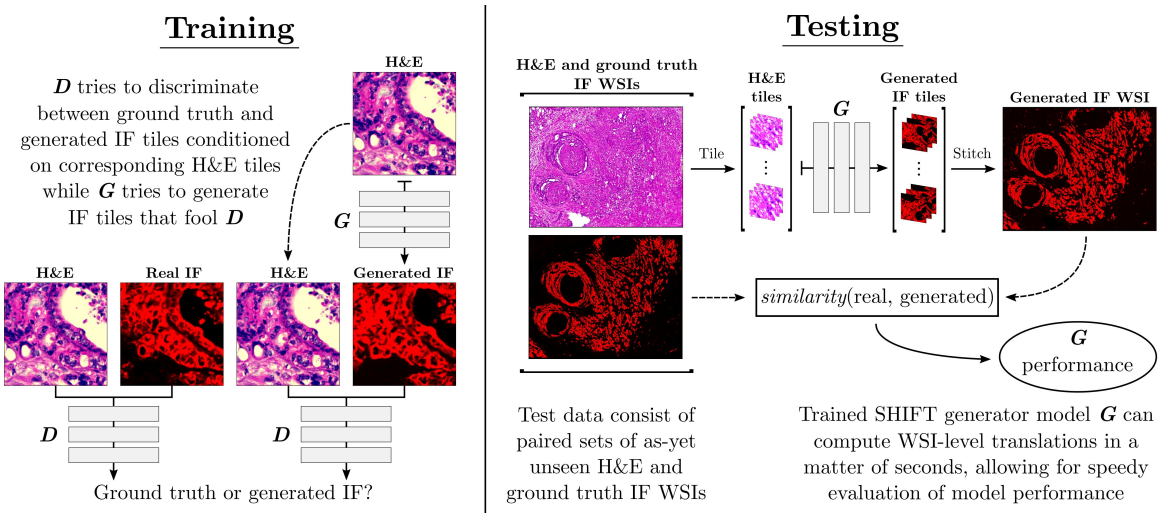

Figure 1. SHIFT: A cGAN framework that generates IF images from corresponding H&E images. During training (left), the generator attempts to generate a realistic IF image conditioned on an H&E image, while the discriminator attempts to differentiate between generated and ground truth image pairs. Once trained, a SHIFT generator can efficiently generate IF WSIs for testing purposes or downstream analysis (right). Figure adapted from Ref. 10.

whether or not tissue and cell morphologies reflect specific protein expression patterns. Thus, it can not only improve our understanding of cancer nuclear morphology, tissue architecture, and spatial patterns corresponding to cell phenotypes, but can also provide clues as to which marker is necessary in addition to H&E for appropriate digital interpretation.

The use of cGANs in medical image analysis has been proposed for many tasks, including segmentation or generation of various regions of interest,[11,12] brain lesion detection,[13] and de-noising of CT images,[14] but to the best of our knowledge, SHIFT is the first method proposed for the generation of IF WSIs by translating H&E-stained WSIs. If deployed together, a multiplicity of properly trained SHIFT models could provide a machine-driven diagnosis mere seconds after an H&E slide is mounted and scanned. As such, SHIFT could become a feasible preliminary, auxiliary, or substitute for multiplexed methods of high-dimension imaging, thus hastening workflows in histopathology, where time is so frequently of the essence.

2. METHODS

2.1 Image-to-image translation

A fundamental problem in the domain of image processing is the mapping of pixels from one representation of a scene to pixels of another representation of the same scene, i.e. image-to-image translation. To approach the problem of translating H&E-stained WSIs to their IF counterparts, we have applied the cGAN-driven algorithm *pix2pix*,[10] which benefits from its bipartite formulation. Like other methods proposed for image-to-image translation, cGANs learn a functional mapping from an input image x to translated image y, i.e. $G : x \mapsto y$, but, unique to a cGAN framework, it is the task of a generator G to generate the image y conditioned on x that fools an adversarial discriminator D, which is in turn trained to tell the difference between ground truth and generated images (Figure 1, left). What ensues from this two-network duel is a model G that generates realistic images that are difficult to distinguish from the ground truth (Figure 1, right), some GAN-generated images being sufficiently realistic to be considered as a proxy for the ground truth when labeled data are scarce or prohibitively expensive.[15]

The cGAN objective of the *pix2pix* algorithm is posed as a binary cross-entropy loss:

$$\mathcal{L}_{\text{cGAN}}(G,D) = \mathbb{E}_{x,y\sim p_{data}(x,y)}\Big[\log D(x,y)\Big] + \mathbb{E}_{x\sim p_{data}(x)}\Big[\log\big(1 - D\big(x, G(x)\big)\big)\Big] \tag{1}$$

where G seeks to minimize the objective and thus minimize the distinguishability of generated and ground truth images, while D seeks the opposite. In addition to the task of fooling D, G is also encouraged to generate images that are faithful to the ground truth through incorporation of an L1 reconstruction loss term:

$$\mathcal{L}_{\mathrm{L1}}(G) = \mathbb{E}_{x,y\sim p_{data}(x,y)}\left[\|y - G(x)\|_1\right] \tag{2}$$

The final *pix*2*pix* objective is

$$G^* = \arg\min_G \max_D \mathcal{L}_{\mathrm{cGAN}}(G,D) + \lambda\mathcal{L}_{\mathrm{L1}}(G) \tag{3}$$

where the regularization parameter $\lambda = 100$ was chosen by Ref. 10 to suit the problems of facade generation, semantic labeling and scene colorization. Though the *pix*2*pix* algorithm can proficiently translate semantic labels into a busy cityscape, its regularization strategy may be less well suited for the problem of translating sparse and ambiguous signals, e.g. low-prevalence IF stains in the training dataset.

2.2 Prevalence-based adaptive regularization

Cancer cells typically remain clustered together as shown in Figure 2 (panCK) and thus it is challenging to balance the reconstruction loss term (2) for positive/negative instances according to the stain prevalence for each training image. For instance, for low-prevalence (sparse) panCK-stained regions in ground truth WSIs, G is more likely to generate an "unstained" pattern rather than generate a sparsely localized stain pattern because the reconstruction loss is relatively small compared to the reconstruction loss for high-prevalence (dense) panCK-stained regions. In order to balance sensitivity and specificity in this context, we hypothesize that a generative model can be receptively tuned to encode sparse staining by being maximally penalized when it makes false classifications on low-prevalence ground truth tiles during training. Thus, we propose a prevalence-based adaptive regularization parameter λ' that may be more suitable for the translation of signals from H&E to IF:

$$\lambda' = \lambda\left(\epsilon + \frac{1}{n}\sum_{i=1}^{n} I_{\Omega(p_i)}\right)^{-1} \tag{4}$$

where $\epsilon = 0.1$ is chosen to offset in cases where stain prevalence is zero, n is the total number of pixels in the ground truth IF tile, and $I_{\Omega(p_i)} = \begin{cases} 1, & \text{if } p_i \text{ in } \Omega \\ 0, & \text{otherwise} \end{cases}$ where Ω represents the ground truth mask, and p_i represents the i-th pixel. Our final objective is

$$G^* = \arg\min_G \max_D \mathcal{L}_{\mathrm{cGAN}}(G,D) + \lambda'\mathcal{L}_{\mathrm{L1}}(G) \tag{5}$$

Utilization of the adaptive regularization parameter λ' maximizes the penalty for generator errors on low-prevalence ground truth tiles and minimizes the penalty for errors on high-prevalence ground truth tiles. By doing this, we can improve localization characteristics and help minimize false classification errors at a distance from true-positive pixels, as shown in Figure 2.

2.3 Ensemble approach

In the context of machine learning, aggregating several trained models can increase prediction accuracy, especially when the aggregated models capture distinct features of their shared input. Thus, we also combined the output of independently-trained models, i.e. models utilizing (3) and (5), to form an ensemble distribution, under the assumption that the training strategies put forward in (3) and (5) are complementary. By doing this, we can smoothen the final output and improve performance by reducing substantial disagreement patterns between models.

3. EXPERIMENTS: DATASET, NETWORKS, AND EVALUATION

This study utilizes a dataset[16] containing WSIs of tumorigenic pancreas tissue acquired at 20X-magnification from two adjacent thin sections: one stained with H&E and and the other co-stained with the fluorescent nuclear marker DAPI and fluorescent antibodies against panCK and α-SMA, two markers commonly used in tumor evaluation.[17,18] The paired 20X images were registered[16] and cropped into four sites, with each site image being $\sim 12,000 \times 8,000$ pixels in size. 10X WSIs were created by half-scaling 20X WSIs. Training data were created by first taking $\sim 10,000$ random 256×256 pixel H&E and IF tile pairs from three sites, then applying single operation manipulations—i.e. jitter, rotation, flipping, Poisson noise—to each tile, yielding $\sim 20,000$ total images in the augmented training data. For a given stain, we trained four leave-one-site-out SHIFT models and generated inferentially-stained WSIs for each site, i.e. each of four models were trained on random tiles from three sites and tested on non-overlapping tiles of the left-out site, which could then be stitched into cohesive WSIs. In this way, we were able to perform a fourfold cross-validation of the SHIFT method for each stain in an intra-patient context. To reduce the deleterious effects of tiling artifacts in the generated panCK WSIs, we utilized three additional test datasets of non-overlapping tiles from each site—one of each test dataset offset by 128 pixels in either x or y or both—and evaluated model performance using the jointly-scaled blend of the four generated WSIs.

The network architectures and implementations for D and G for all models are as described in the original *pix2pix* paper,[10] except where explicitly specified. Training batch size was set to 4 for all experiments and for fair comparison, we tuned the regularization setting for each model by training over a range of $\lambda : 50 - 5000$ and selected the models with optimal λ^* that yielded the best performance. Models were trained for 20 epochs at a

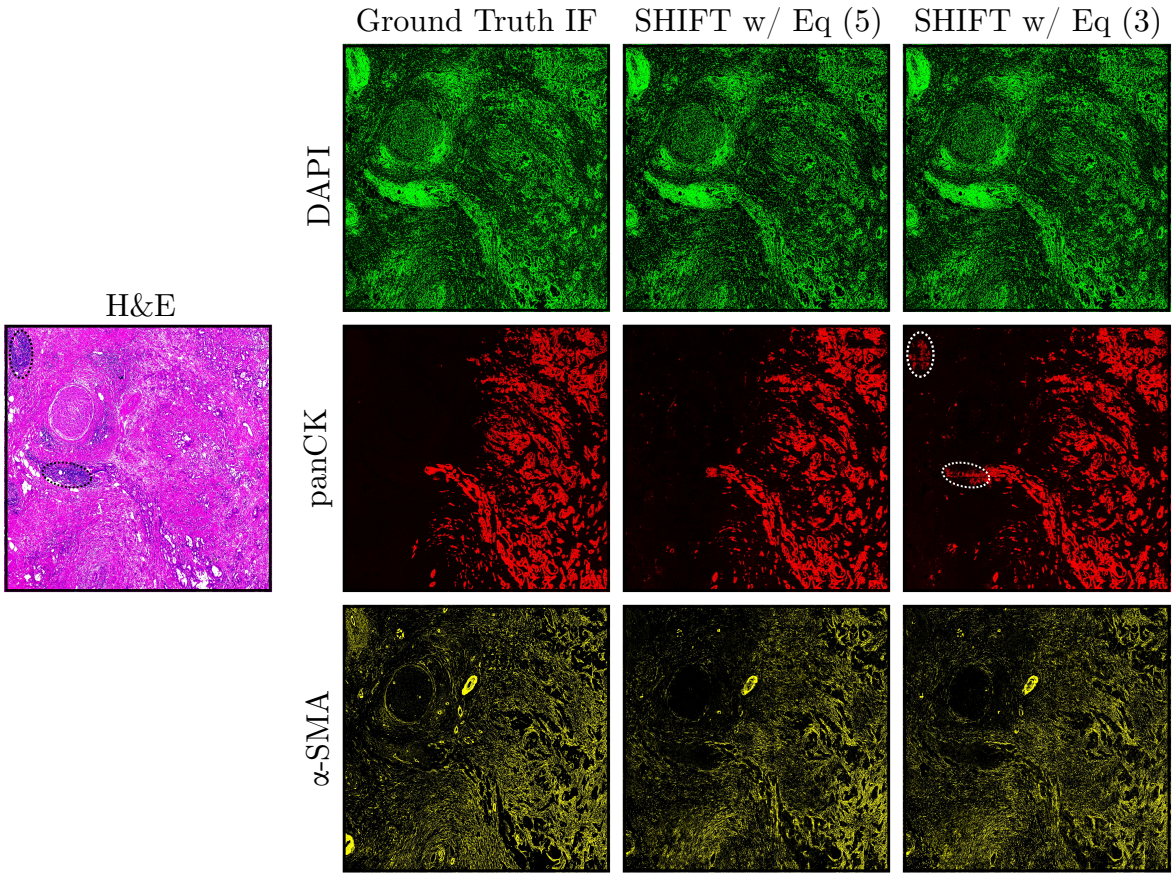

Figure 2. SHIFT model results for site 1 ($12,656 \times 10,858$ pixels at 20X magnification). Each SHIFT image represents the result for the model with optimal λ^* which yielded the best performance (Table 1). The circled dark regions in the H&E image are clusters of invading lymphocytes which the SHIFT model with fixed λ (Eq (3)) misclassified as being panCK-positive (see the corresponding circled regions in the middle-right image). The SHIFT model with adaptive λ' (Eq (5)) did not commit these errors.

fixed learning rate of 0.0002, followed by 10 epochs over which the learning rate linearly decayed to zero. Once trained, each SHIFT model was able to compute WSI-level translation in less than one minute.

For evaluation of SHIFT model performance, we measured the Matthews correlation coefficient (MCC),[19] the Dice similarity coefficient (DSC), as well as other standard classification performance metrics for comparison of the ground truth and generated IF masks produced using a global 10%-luminance threshold on the contrast-adjusted 8-bit ground truth and generated IF WSIs. We also measured the peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM)[20] between raw ground truth and raw generated IF WSIs.

4. RESULTS AND DISCUSSION

Representative results for the translations from H&E-to-DAPI (SHIFT2DAPI), H&E-to-panCK (SHIFT2panCK), and H&E-to-α-SMA (SHIFT2α-SMA) are shown in Figure 2. We performed SHIFT2DAPI experiments at both 10X- and 20X-magnification to assess whether or not SHIFT model inference is sensitive to image resolution, and found minor improvements in most metrics when models were trained on 20X tiles (Table 1, top), suggesting that localized features of the DAPI stain may be more important for SHIFT2DAPI inference than higher-level architectural features. Since hematoxylin and DAPI are both robust stains for cell nuclei, the task of a SHIFT2DAPI model is theoretically trivial—translate hematoxylin intensity into DAPI intensity—and thus provides insight into the upper limits of SHIFT performance. Note that there exists μm-scale structural differences between ground truth H&E and IF WSIs due to serial tissue acquisition. Nevertheless, the results for models utilizing (5) are consistent with those from a comparison between the DAPI mask and a cell nucleus segmentation mask derived from the H&E image (data not shown), indicating that SHIFT2DAPI achieves good performance up to the fundamental limit.

Model translation	Mag.	Site generated	G^*	λ^*	MCC	DSC	Accu.	Spec.	Prec.	Sens.	PSNR	SSIM
SHIFT2DAPI	10X	1	Eq (3)	5000	0.838	0.885	0.932	0.938	0.857	**0.916**	30.89	0.883
			Eq (5)	1000	0.845	0.890	0.936	0.951	0.881	0.898	31.40	0.887
	20X	1	Eq (3)	500	0.857	0.897	0.942	0.965	0.910	0.886	**31.53**	0.883
			Eq (5)	5000	**0.861**	**0.900**	**0.944**	**0.966**	**0.913**	0.887	31.50	**0.898**
SHIFT2panCK	10X	1	Eq (3)	1000	0.704	0.749	0.909	0.918	0.662	0.863	22.99	0.769
			Eq (5)	1000	**0.754**	**0.793**	**0.933**	**0.953**	**0.766**	0.822	22.95	**0.791**
			Ensemble	–	0.729	0.769	0.917	0.922	0.679	**0.887**	**23.19**	0.782
		2	Eq (3)	1000	0.817	0.855	0.937	0.946	0.812	**0.903**	28.21	0.819
			Eq (5)	1000	0.814	0.853	**0.939**	**0.959**	**0.845**	0.861	27.89	0.816
			Ensemble	–	**0.821**	**0.859**	0.938	0.948	0.819	**0.903**	**28.66**	**0.828**
		3	Eq (3)	1000	**0.790**	**0.822**	**0.945**	0.965	0.810	0.834	26.36	0.815
			Eq (5)	1000	0.777	0.807	**0.945**	**0.978**	**0.860**	0.760	26.16	0.818
			Ensemble	–	**0.790**	**0.822**	0.944	0.958	0.786	**0.862**	**26.69**	**0.828**
		4	Eq (3)	1000	0.812	0.849	0.940	0.967	0.865	**0.833**	26.05	0.807
			Eq (5)	1000	0.792	0.826	0.936	**0.981**	**0.908**	0.758	25.87	0.810
			Ensemble	–	**0.819**	**0.854**	**0.943**	0.972	0.881	0.828	**26.35**	**0.818**
SHIFT2α-SMA	10X	1	Eq (3)	1000	–	–	–	–	–	–	24.70	0.603
			Eq (5)	1000	–	–	–	–	–	–	24.84	0.608
			Ensemble	–	–	–	–	–	–	–	**25.09**	**0.611**
		2	Eq (3)	1000	–	–	–	–	–	–	25.69	0.634
			Eq (5)	1000	–	–	–	–	–	–	25.81	0.642
			Ensemble	–	–	–	–	–	–	–	**26.02**	**0.643**
		3	Eq (3)	1000	–	–	–	–	–	–	24.19	0.588
			Eq (5)	1000	–	–	–	–	–	–	24.41	0.598
			Ensemble	–	–	–	–	–	–	–	**24.74**	**0.606**
		4	Eq (3)	1000	–	–	–	–	–	–	25.21	0.634
			Eq (5)	1000	–	–	–	–	–	–	26.34	**0.675**
			Ensemble	–	–	–	–	–	–	–	**26.39**	0.674

Table 1. SHIFT model parameters and performances. The result for the model with the optimal λ^* that yielded the best performance (MCC for DAPI and panCK, SSIM for α-SMA) is shown for each combination of magnification and G^*. Models were trained until errors stabilized.

Given that panCK will stain only the subset of cells which are CK-positive, rather than stain a ubiquitous cytological landmark as do hematoxylin and DAPI, the translation from H&E to panCK is a more interesting but challenging task. Although SHIFT2panCK models performed less well than SHIFT2DAPI in most categories, it is difficult to visually distinguish the generated from the ground truth panCK IF WSIs, as shown in Figure 2. With one exception (the sensitivity of SHIFT2panCK for site 4), either the models utilizing the proposed method (5) alone or the ensemble approach performed as well as or better than models utilizing (3) alone, i.e. $pix2pix$. Notably, models utilizing the proposed method (5) showed better localization characteristics (Figure 2, circled misclassified regions for model utilizing (3)).

In contrast to DAPI and panCK stain patterns, the α-SMA stain pattern is sinuous and high-frequency (Figure 2, bottom). When these attributes are compounded by spatial deformity and other complications from the serial acquisition of H&E and IF WSIs, pixel-level evaluation of generated α-SMA WSIs becomes exceedingly challenging. For this reason, we excluded evaluation metrics that were contingent on α-SMA mask generation in favor of metrics which reflect the global configurations of the α-SMA IF WSIs (Table 1, bottom). While the ensemble approach performed best in both categories for most sites, all models utilizing the proposed method (5) alone outperformed the models utilizing (3) alone.

5. CONCLUSION

The results presented in this proof-of-concept study demonstrate that the proposed SHIFT method can rapidly and accurately infer the distribution of clinically relevant markers in histopathological images. Future work will focus on multiscale-image training strategies, hyperparameter tuning, translation to other IF stains, and inter-patient model evaluation.

This work covers only a small fraction of the possible applications of GANs in digital pathology. As our preliminary findings demonstrate that deep learning architectures enable the correlation of features across histopathological and IF images, we believe that SHIFT may be broadly capable of identifying crosswise mappings between different imaging modalities with shared spatial features, even when output distributions are sparse.

ACKNOWLEDGMENTS

This work was supported by SU2C American Association for Cancer Research (AACR) (SU2C-AACR-DT12-14).

REFERENCES

[1] Lin, J.-R., Fallahi-Sichani, M., and Sorger, P. K., "Highly multiplexed imaging of single cells using a high-throughput cyclic immunofluorescence method," *Nature Communications* **6**, 8390 (2015).

[2] Gerdes, M. J., Sevinsky, C. J., Sood, A., Adak, S., Bello, M. O., Bordwell, A., Can, A., Corwin, A., Dinn, S., Filkins, R. J., et al., "Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue," *Proceedings of the National Academy of Sciences* **110**(29), 11982–11987 (2013).

[3] Tsujikawa, T., Kumar, S., Borkar, R. N., Azimi, V., Thibault, G., Chang, Y. H., Balter, A., Kawashima, R., Choe, G., Sauer, D., et al., "Quantitative multiplex immunohistochemistry reveals myeloid-inflamed tumor-immune complexity associated with poor prognosis," *Cell Reports* **19**(1), 203–217 (2017).

[4] Zrazhevskiy, P., True, L. D., and Gao, X., "Multicolor multicycle molecular profiling (m3p) with quantum dots for single-cell analysis," *Nature Protocols* **8**(10), 1852 (2013).

[5] Angelo, M., Bendall, S. C., Finck, R., Hale, M. B., Hitzman, C., Borowsky, A. D., Levenson, R. M., Lowe, J. B., Liu, S. D., Zhao, S., et al., "Multiplexed ion beam imaging of human breast tumors," *Nature Medicine* **20**(4), 436–442 (2014).

[6] Giesen, C., Wang, H. A., Schapiro, D., Zivanovic, N., Jacobs, A., Hattendorf, B., Schüffler, P. J., Grolimund, D., Buhmann, J. M., Brandt, S., et al., "Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry," *Nature Methods* **11**(4), 417–422 (2014).

[7] O'hurley, G., Sjöstedt, E., Rahman, A., Li, B., Kampf, C., Pontén, F., Gallagher, W. M., and Lindskog, C., "Garbage in, garbage out: a critical evaluation of strategies used for validation of immunohistochemical biomarkers," *Molecular oncology* **8**(4), 783–798 (2014).

[8] Fuchs, T. J. and Buhmann, J. M., "Computational pathology: Challenges and promises for tissue analysis," *Computerized Medical Imaging and Graphics* **35**(7), 515–530 (2011).

[9] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y., "Generative adversarial nets," in [*Advances in Neural Information Processing Systems*], 2672–2680 (2014).

[10] Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A., "Image-to-image translation with conditional adversarial networks," in [*The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*], (July 2017).

[11] Udrea, A. and Mitra, G. D., "Generative adversarial neural networks for pigmented and non-pigmented skin lesions detection in clinical images," in [*Control Systems and Computer Science (CSCS), 2017 21st International Conference on*], 364–368, IEEE (2017).

[12] Costa, P., Galdran, A., Meyer, M. I., Mendonça, A. M., and Campilho, A., "Adversarial synthesis of retinal images from vessel trees," in [*Image Analysis and Recognition. ICIAR 2017. Lecture Notes in Computer Science*], Karray, F., Campilho, A., and Cheriet, F., eds., **10317**, Springer, Cham.

[13] Alex, V., Safwan K. P., M., Chennamsetty, S. S., and Krishnamurthi, G., "Generative adversarial networks for brain lesion detection," *Proc. SPIE* **10133**, 101330G–101330G–9 (2017).

[14] Wolterink, J. M., Leiner, T., Viergever, M. A., and Isgum, I., "Generative adversarial networks for noise reduction in low-dose CT," *IEEE Transactions on Medical Imaging* (2017).

[15] Bousmalis, K., Silberman, N., Dohan, D., Erhan, D., and Krishnan, D., "Unsupervised pixel-level domain adaptation with generative adversarial networks," in [*The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*], (July 2017).

[16] Chang, Y. H., Thibault, G., Madin, O., Azimi, V., Meyers, C., Johnson, B., Link, J., Margolin, A., and Gray, J. W., "Deep learning-based nucleus classification in pancreas histological images," in [*39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*], 672–675 (2017).

[17] Barak, V., Goike, H., Panaretakis, K. W., and Einarsson, R., "Clinical utility of cytokeratins as tumor markers," *Clinical biochemistry* **37**(7), 529–540 (2004).

[18] Sinn, M., Denkert, C., Striefler, J., Pelzer, U., Stieler, J., Bahra, M., Lohneis, P., Dörken, B., Oettle, H., Riess, H., et al., "α-Smooth muscle actin expression and desmoplastic stromal reaction in pancreatic cancer: results from the CONKO-001 study," *British journal of cancer* **111**(10), 1917–1923 (2014).

[19] Matthews, B. W., "Comparison of the predicted and observed secondary structure of T4 phage lysozyme," *Biochimica et Biophysica Acta (BBA)-Protein Structure* **405**(2), 442–451 (1975).

[20] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing* **13**(4), 600–612 (2004).