Neuro-Oncology Advances

Integrating morphologic and molecular histopathological features through whole slide image registration and deep learning

Kevin Faust[†], Michael K. Lee[†], Anglin Dent, Clare Fiala, Alessia Portante, Madhumitha Rabindranath, Noor Alsafwani, Andrew Gao, Ugljesa Djuric, and Phedias Diamandis

Department of Computer Science, University of Toronto, 40 St. George Street, Toronto, ON M5S 2E4, Canada (K.F.); Laboratory Medicine Program, Department of Pathology, University Health Network, 200 Elizabeth Street, Toronto, ON M5G 2C4, Canada (K.F., C.F., N.A., A.G., P.D.); Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON M5S 1A8, Canada (M.K.L., A.D., A.P., M.R., A.G., P.D.); Department of Pathology, College of Medicine, Imam Abdulrahman Bin Faisal University, P.O. Box.2208, Dammam, 31441, Saudi Arabia (N.A.); Princess Margaret Cancer Centre, 101 College Street, Toronto, ON M5G 1L7, Canada (M.K.L., U.D., P.D.); Department of Medical Biophysics, University of Toronto, 101 College St, Toronto, ON M5G 1L7, Canada (P.D.)

†Equal contributions.

Corresponding Author: Phedias Diamandis, MD, PhD, FRCPC, Neuropathologist, Department of Pathology, University Health Network, 12-308, Toronto Medical Discovery Tower (TMDT), 101 College St, Toronto, ON M5G 1L7, Canada (p.diamandis@mail.utoronto.ca).

Abstract

Background. Modern molecular pathology workflows in neuro-oncology heavily rely on the integration of morphologic and immunohistochemical patterns for analysis, classification, and prognostication. However, despite the recent emergence of digital pathology platforms and artificial intelligence-driven computational image analysis tools, automating the integration of histomorphologic information found across these multiple studies is challenged by large files sizes of whole slide images (WSIs) and shifts/rotations in tissue sections introduced during slide preparation.

Methods. To address this, we develop a workflow that couples different computer vision tools including scale-invariant feature transform (SIFT) and deep learning to efficiently align and integrate histopathological information found across multiple independent studies. We highlight the utility and automation potential of this workflow in the molecular subclassification and discovery of previously unappreciated spatial patterns in diffuse gliomas.

Results. First, we show how a SIFT-driven computer vision workflow was effective at automated WSI alignment in a cohort of 107 randomly selected surgical neuropathology cases (97/107 (91%) showing appropriate matches, AUC = 0.96). This alignment allows our AI-driven diagnostic workflow to not only differentiate different brain tumor types, but also integrate and carry out molecular subclassification of diffuse gliomas using relevant immunohistochemical biomarkers (IDH1-R132H, ATRX). To highlight the discovery potential of this workflow, we also examined spatial distributions of tumors showing heterogenous expression of the proliferation marker MIB1 and Olig2. This analysis helped uncover an interesting and unappreciated association of Olig2 positive and proliferative areas in some gliomas (r = 0.62).

Conclusion. This efficient neuropathologist-inspired workflow provides a generalizable approach to help automate a variety of advanced immunohistochemically compatible diagnostic and discovery exercises in surgical neuropathology and neuro-oncology.

Key Points

- We develop an automated workflow to integrate morphologic and molecular data into tumor classification.
- This approach could provide quality assurance and standardization to brain tumor classification at remote centers.

Importance of the Study

Histopathological analysis of brain tumor continues to remain an important tool for both clinical management and neuro-oncology research. The complexity of these analyses has however greatly increased in the molecular era with a growing number of immunohistochemical studies that need to be interpreted in a tandem and parallel manner. In this study, we develop and demonstrate a generalizable workflow that helps address this issue by automating alignment of multiple tissue sections generated from different immunohistochemistry-based molecular studies. We highlight how this spatial preservation of tissue coordinates and staining patterns, across these multiple studies, provides improvements to both clinical workup of molecularly defined brain tumor entities and discovery of novel associations of tumor biomarkers. Together, this automated workflow offers the potential to provide a powerful quality assurance and discovery tool for both remote and tertiary academic centers respectively.

Progress in the molecular analysis of patient tissue continues to substantially revise traditional approaches to disease classification and analysis.[1,2] As a result, most contemporary pathology workflows now rely on an integrated approach in which traditional histomorphologic patterns found on Hematoxylin and Eosin (H&E)-stained tissue sections are examined in tandem with ancillary molecular tests. These later studies are often carried out in the form of convenient immunohistochemical (IHC) assays on serial tissue sections. Many of these analytical steps however remain largely manual, qualitative, and difficult to scale across large cohorts for discovery.

Recently, breakthroughs in digital pathology have also allowed for glass histopathology slides to be digitized into high-resolution whole slide images (WSIs) and analyzed by computational tools. This has been particularly successful through the application of deep convolutional neural networks (CNNs) to H&E- and IHC-stained WSIs.[3–6] However, attempts to automate integration of histomorphologic and molecular information found on respective WSIs are scarce. One challenge in automating this modern workflow is unpredictability in tissue rotations and shifts introduced during slide preparation that alter the spatial coordinates of tissue when WSIs are generated. While human observers use visual cues of overall tissue shapes to manually align independent sections, the large files sizes of WSIs (~1-2 gigabytes) and contrasting colorization between H&E and IHC studies make this computationally difficult to streamline. Here, we show how adequate WSI alignment can be achieved by pairing the image registration tool, scale invariant feature transform (SIFT),[7] with simple image resizing and pixel intensity normalization steps (Figure 1a). We append this workflow with contemporary deep learning tools to automate quantitative integration of histopathologic data across multiple studies and resolve various known and less well appreciated patterns of tumoral heterogeneity in gliomas.

Methods

Whole Slide Image Registration

Given two WSIs, I_1 and I_2, both are downsized (bicubic) by 100×, foregoing cellular level detail but allowing the outline of tissue shapes to be retained. Images are then automatically converted to gray scale in order to match the distribution of pixel intensities found within I_2 to those found within I_1 (histogram matching[8]). In the case where I_1 is an H&E image and I_2 is an IHC image, this is particularly needed to better equalize the different pixel intensity values between the two images that hinder alignment. These gray scale images thus represent a 10,000-fold reduction in pixel count with only a single channel. It is important to note that while these down-sampled slide avatars provide manageable image sizes to carry out WSI alignment, all downstream classification steps are carried out on the full resolution versions as outlined below. Following the initial automated processing steps, we use SIFT on these small tissue representation images to find features (keypoints and feature vectors) and then match them using Fast Library for Approximate Nearest Neighbors (FLANN).[9] Bad matches are filtered out using Lowe's ratio test[10] and the following conditions are required to be met to create a match:

1. ≥ 4 matches
2. Homography can be estimated using random sample consensus (RANSAC)[11]
3. The alignment score is above the cutoff of 71.4% (described in the next section)

Because the homography H was computed on the smaller images, we scale it appropriately to work on the original images as $H' = S \times H \times S^{-1}$ where $*$ is our resize factor of 100

$$S = \begin{pmatrix} * & 0 & 0 \\ 0 & * & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

To determine where a given image patch in I_1 belongs to in I_2, we take its center coordinate and apply H'. The transformed coordinate is thus the corresponding center coordinate of the image patch in I_2. By using the center coordinate, we are able to maximize the robustness of the transformation when images are rotated and/or have other alterations. Any transformed coordinates in I_2 that end up out of bounds or in empty space are ignored.
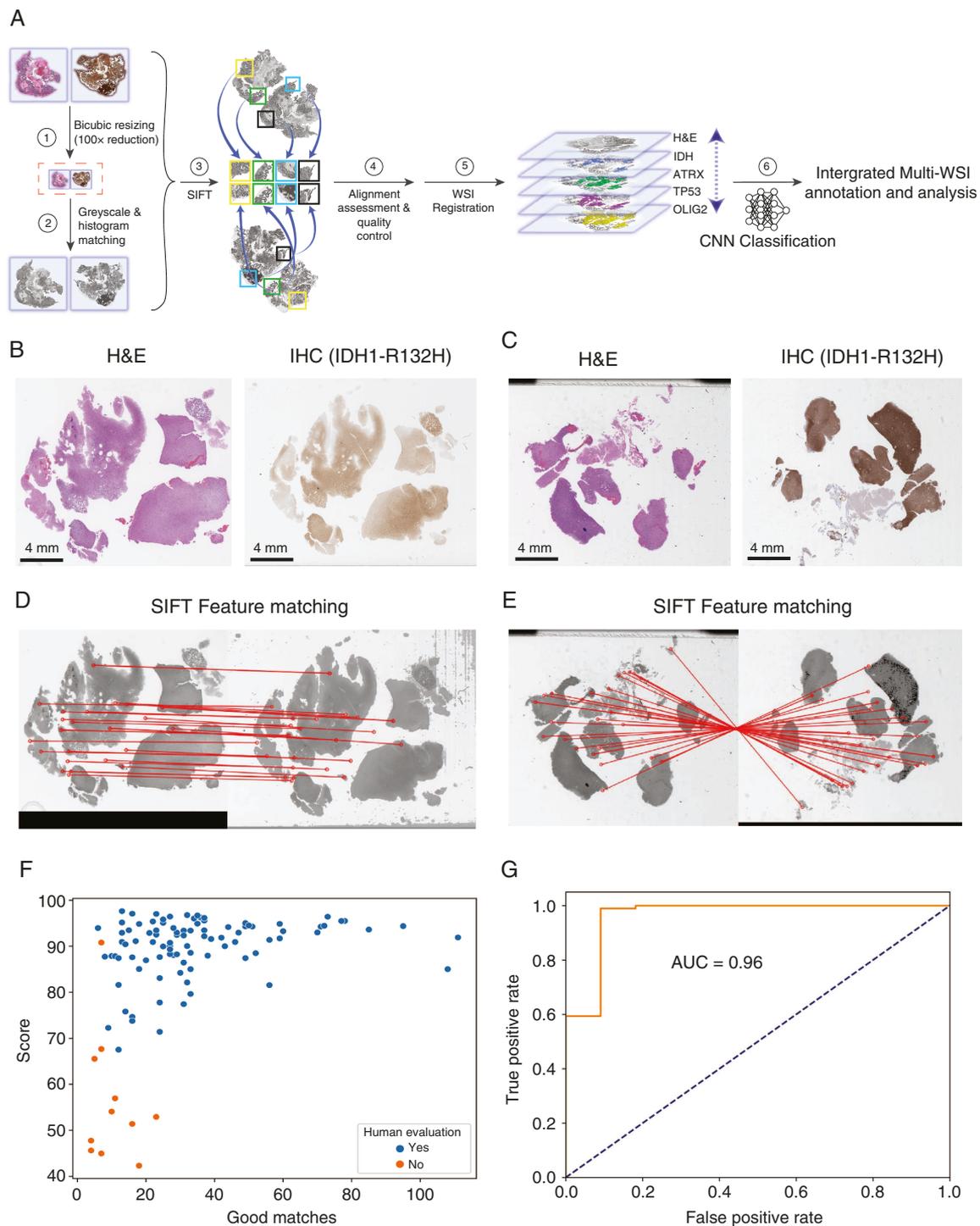
Figure 1. Automated WSI registration using scale-invariant feature transform (SIFT). **a**. Scheme illustrating automated multi-WSI alignment and analysis workflow. **b–c**. H&E and IHC (IDH1-R132H) WSI pairs of contiguous sections of diffuse gliomas. Notice 180° rotation introduced during slide preparation in panel **c** that makes alignment difficult. **d–e**. Numerous SIFT feature matchings (red lines) highlight good alignment despite these rotations/shifts. **f**. Scatter plot showing relationship between a tissue overlap alignment score, number of good matches, and human-based assessment of alignment. **g**. Receiver operating characteristic (ROC) curve illustrating the accuracy of this SIFT-based matching with varying alignment score thresholds. Optimal cutoff: 71.4% (Youden's index).

The following summarizes the variations of the alignment workflows highlighted in the manuscript:

1. Evaluating lesion-only areas on an IHC image (Figure 2): Given an H&E (I_1) and IHC (I_2), the coordinates of all lesional patches (probability score of 85%) within I_1 are transformed using H' and only these coordinates are evaluated on in I_2. The entire tissue present in I_2 would be evaluated on if the alignment conditions listed above fail to be satisfied.

2. Evaluating H&E clustered subgroups on an IHC image (Figure 3): Given an H&E (I_1) and IHC (I_2), all tissue patches within I_1 are extracted and have their deep learning feature vector (DLFV) hierarchically clustered into computer-based subgroups.[12] The coordinates of the tissue patches in a given subgroup are transformed using H' and only these coordinates are evaluated on in I_2.

3. Correlating IHC images (Figure 4): Given an IHC (I_1) and IHC (I_2), the coordinates of all tissue patches within I_1 are transformed using H'. The image patches belonging to each coordinate are iterated over concurrently, examining their immunopositivity score at each step.

If I_2 fails to be aligned to I_1 for any reason, an auxiliary WSI I_1', if available, may be used to serve as a "slide-bridge" to improve the chances for alignment assuming the following:

1. I_1' is the same specimen as I_1 and I_2
2. I_1' is able to be aligned with I_1
3. I_2 is able to be aligned with I_1'

in which case we would have two resulting homographies H^* and H^{**}. Coordinates would be transformed sequentially as $(T_{H^{**}} \circ T_{H^*})(x)$. The typical tissue thickness of H&E and IHC at our institute are 4 and 3 micrometers, respectively. This "bridging" step is therefore especially useful in the cases in which multiple IHC studies were carried on a given tissue block and alignment is desired between two distant H&E/IHC studies in which the overall tissue patterns may have changed sufficiently to compromise alignment. As all our slides were anonymized prior to the commencement of this study, the effect of slide position on the slide alignment performance could not be formally evaluated and is likely also affected by other tissue factors (eg variation in the distinctiveness of tissue shapes across specimens). Code can be found at https://bitbucket.org/diamandislabii/faust-alignment-2021

Performance Evaluation of Whole Slide Image Registration

To evaluate the accuracy of our alignment (tissue overlap alignment score) between two whole slide images I_1 and I_2, we first extract the coordinates of every possible 100 × 100 pixel image patch belonging to I_1. Assuming the homography H' was able to successfully be computed, we transform these coordinates to get the corresponding

coordinates in I_2. Both coordinate lists are iterated over concurrently, observing one of the following results at each step: (i) Both p_1 and p_2 are blank, (ii) Both p_1 and p_2 are tissue or (iii) p_1 is blank but p_2 is tissue (or vice versa), where p_1 and p_2 are the image patches located at the current coordinate in I_1 and I_2 respectively. Using these conditions, we compute an alignment score as the percentage:

$$\text{Tissue alignment score : } \frac{(i) + (ii)}{(i) + (ii) + (iii)} \times 100 \text{ \%}$$

The image patch size of 100 × 100 for this was chosen as we found it small enough to ensure patches were either wholly blank or wholly tissue while having an acceptable computational runtime. Using a cohort of 107 image pairs (H&E and IHC) (Supplementary Table 1), the alignment score algorithm took an average of 8 min 40 s to complete for each slide and resulted in an AUC of 0.96 (Figure 1g) when compared to the ground truth (manual human assessment). Youden's index[13] was used to find a cutoff value for our alignment score; calculated to be 71.4%. We used tissue alignment score for the purpose of paper, but SIFT-based alignment only takes ~1 second on average per WSI pair.

CNN Optimization and Development of Image Training Sets

For the development of the histomorphologic classifier, we optimized the widely available VGG19 CNN using transfer learning to fine-tune the existing ImageNet-based weight matrices towards histological patterns of 16 common tissue and brain tumour classes/patterns encountered in surgical neuropathology practice. Training images for this exercise were retrieved from a previous study and consisted of 172,712 pathologist-annotated image patches generated from 973 WSIs.[14,15] Closely related tumor types (eg meningioma subtypes) or those often requiring molecular analysis to subtype (eg gliomas) were grouped together for simplicity for this H&E model. This model reached a validation accuracy of 97.6% among training images after 50 epochs.

To complement the histomorphologic analysis, IHC classifiers were also optimized using transfer learning on VGG19 CNN. We used retrospective WSI images for training the ATRX and IDH1 classifiers that did not overlap with the testing set. Each IHC classifier was trained to recognize IHC stains of specific molecular markers (ATRX, IDH1-R132H, and MIB-1) relevant for molecular subclassification of glioma.[16] Training images for these various classifiers consisted of 4,248 ATRX and 8,259 IDH1-R132H. Our binary ATRX classifier (retained/lost) reached a validation accuracy of 97.8% among training images after 250 epochs. The binary IDH1-R132H classifier (mutated/negative) reached a validation accuracy of 99.3% among training images after 250 epochs. For this specific study, our classifiers were meant to quantify the overall staining pattern in areas defined as diagnostic lesioned tissue. While we felt this was sufficient for this specific application,
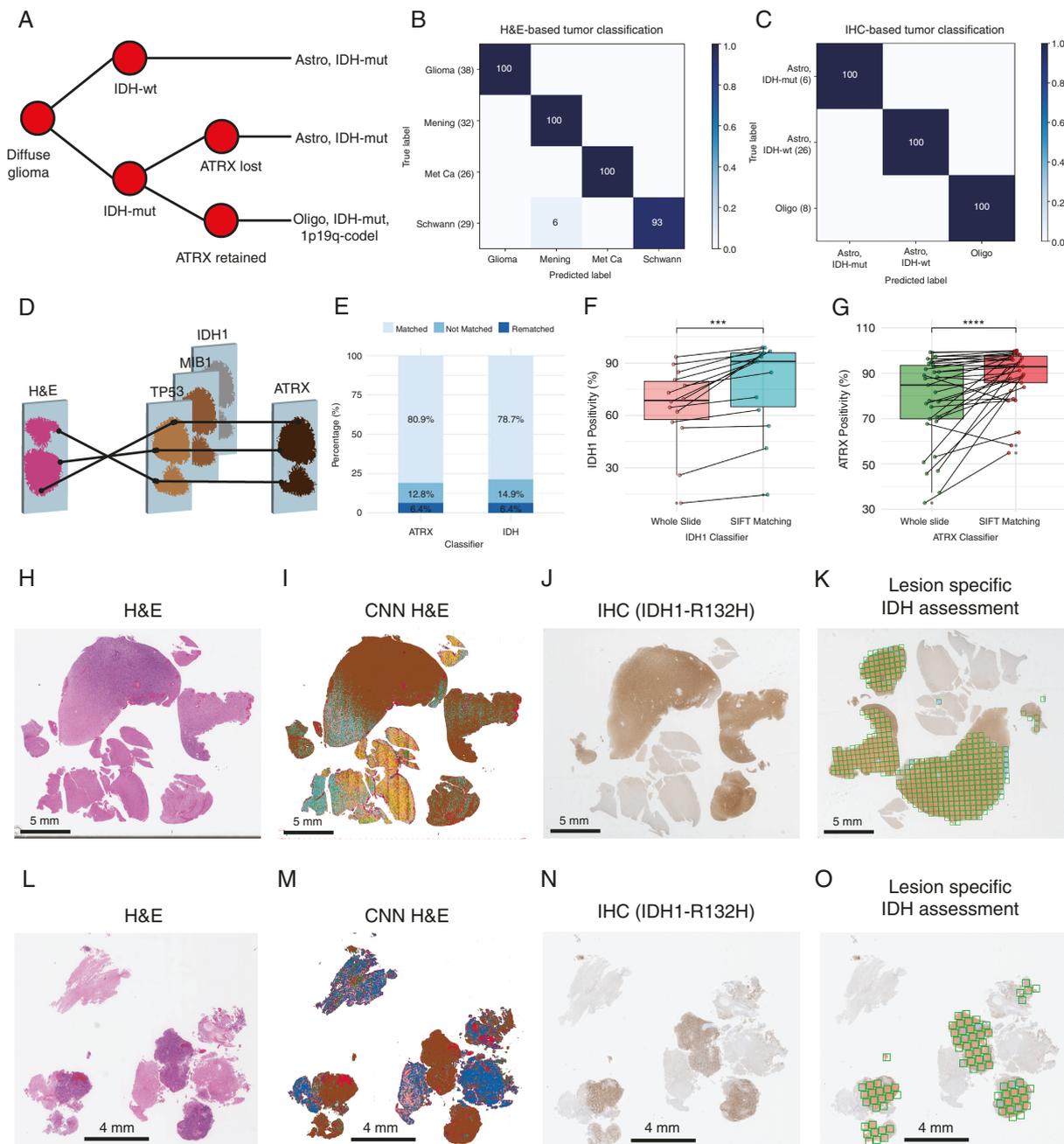
Figure 2. **Molecular subclassification of diffuse gliomas with WSI registration and convolutional neural network**. **a**. Glioma cases can be further subclassified with relevant IHC stains (IDH1-R132H and ATRX) using binary IHC CNNs. **b**. Confusion matrix of 125 classified brain tumor cases using our 16 class H&E model. In addition to providing a morphological diagnosis, image segmentation using this model provides the spatial reference for lesion-specific IHC assessment following alignment. **c**. Confusion matrix of subclassification of gliomas with lesion-specific IHC assessment following alignment. **d**. Cartoon illustrating "bridging" approach between other available IHC WSIs to increase the proportion of aligned cases. **e**. Relative proportion of aligned WSIs in this analysis. **f–g**. Box plots showing the quantitative positive score of IDH1-R132H and ATRX IHC WSIs before and after alignment. *** denote $P < .001$ and **** denote $P < .0001$. **h–o**. Example of spatial assessment of IDH (panels h–k) and ATRX (panels l–o) positivity in only lesional areas. Brown color in the H&E CNN map highlights lesion tissue (ie glioma). Blue and yellow denote areas of necrosis and adjacent brain tissue respectively.

the workflow could in theory be updated in future iterations with more intricate IHC classifiers that differentiate between tumor and nontumor cells (eg blood vessels).

For Ki67/MIB-1 and Olig2 quantification, we developed a mask R-CNN-based workflow which uses a nuclei segmentation approach, classifying each detected nucleus as positive, negative, or miscellaneous (eg blood cells and
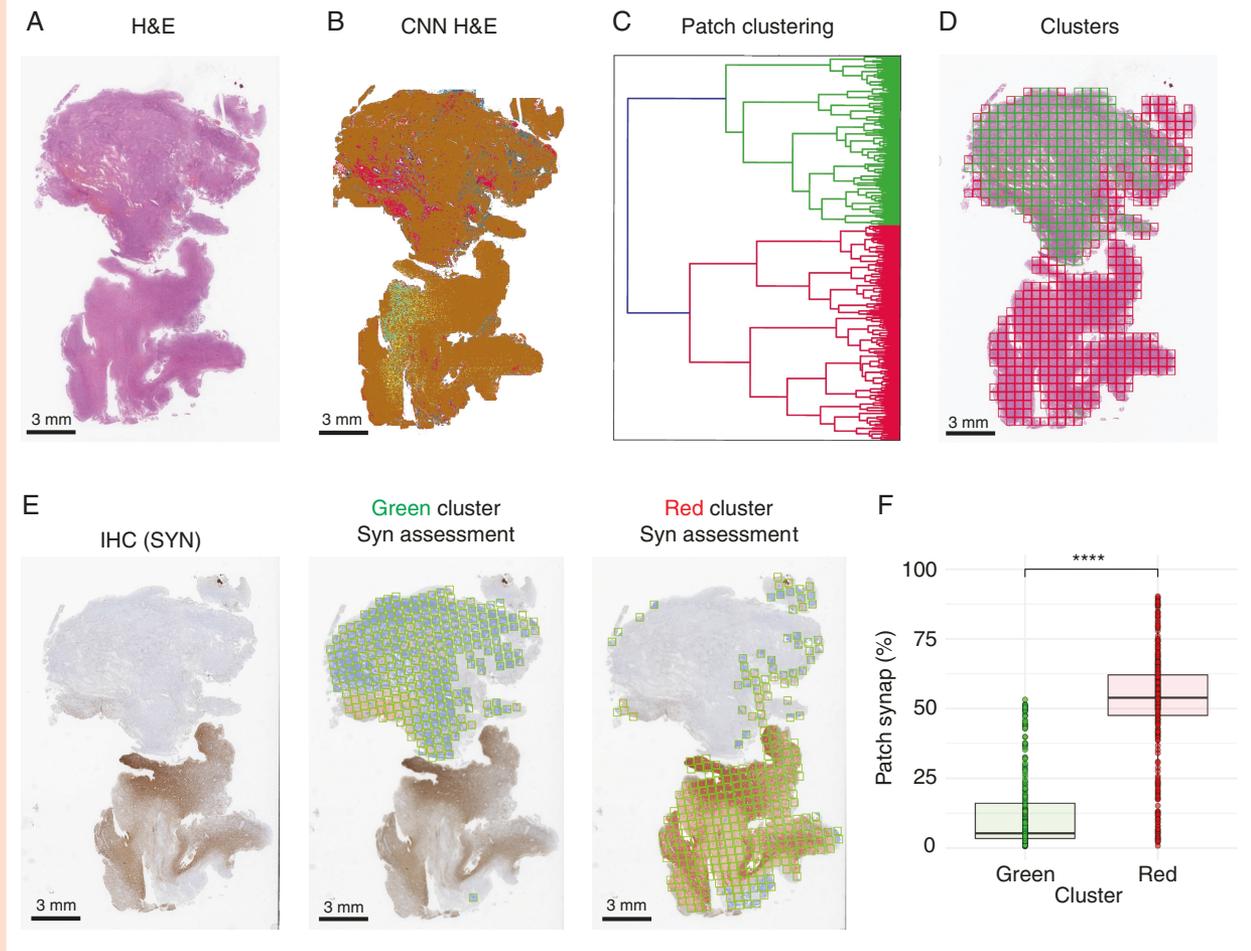
Figure 3. **Resolving intra-tumoral heterogeneity using WSI registration**. **a–b**. Illustrative case in which a WSI of a glioblastoma is comprised largely of lesional tissue (colorized brown). **c–d**. Patch-level feature extraction and image clustering can however define 2 morphologically distinct regions. **e–f**. These regions were found to spatially align with various neuronal IHCs (synaptophysin, neurofilament) defining an infiltrative niche in the red cluster. **** denote $P < .0001$.

tissue folds).[17] Three MIB-1 stained WSIs (ie meningioma, glioblastoma, and small cell neuroendocrine metastatic cancer) were sectioned into 128 × 128 image patches that were manually annotated by a web-based tool,[18] 19,502 annotations across 300 training images. Transfer learning was leveraged to fine tune the ResNeXt-101 R-CNN model which was pre-trained on the COCO 2017 dataset (~37 COCO epochs). The final classifier achieved a box average precision (AP) of 31.83 and mask AP of 30.14 after 10,000 epochs.

Development of Prospective Histopathology Testing Sets of Whole Slide Images

To test the histomorphologic classifier, 125 diagnostic slides from representative and randomly selected clinical cases from our neuropathology service were digitized into whole slide images (WSIs) on the Aperio AT2 whole slide scanner at an apparent magnification of 20× and a compression quality of 0.70. This includes 38 diffuse gliomas, 32 meningiomas, 26 metastatic carcinomas,

and 29 schwannomas. For our glioma cohort, in addition to H&E-stained slides, we also scanned associated IHC slides to show how we could mimic the integrated molecular classification schemes used in routine practice. All histomorphologic diagnoses for these cases were reviewed by at least 2 board-certified pathologists and had corroborating clinical (eg radiology) and molecular (eg immunohistochemistry) information to support the ground truth label. Only the diagnoses relating to individual cases was extracted from the medical records and all WSIs were otherwise deidentified. Our study was reviewed and approved by The University Health Network Research Ethics Board (REB). For the IHC analysis, we took the 38 glioma samples from the initial histomorphologic analysis and added 9 additional samples for a total of 47 samples to ensure we had sufficiently covered all major adult glioma subtypes (32 astrocytomas, IDH-wildtype, 8 oligodendrogliomas, and 7 astrocytomas, IDH-mutant). We note that all the included glioma cases were either IDH wild-type or had the canonical IDH1-R132H mutation detectable by IHC. All cases in which the pathologist could
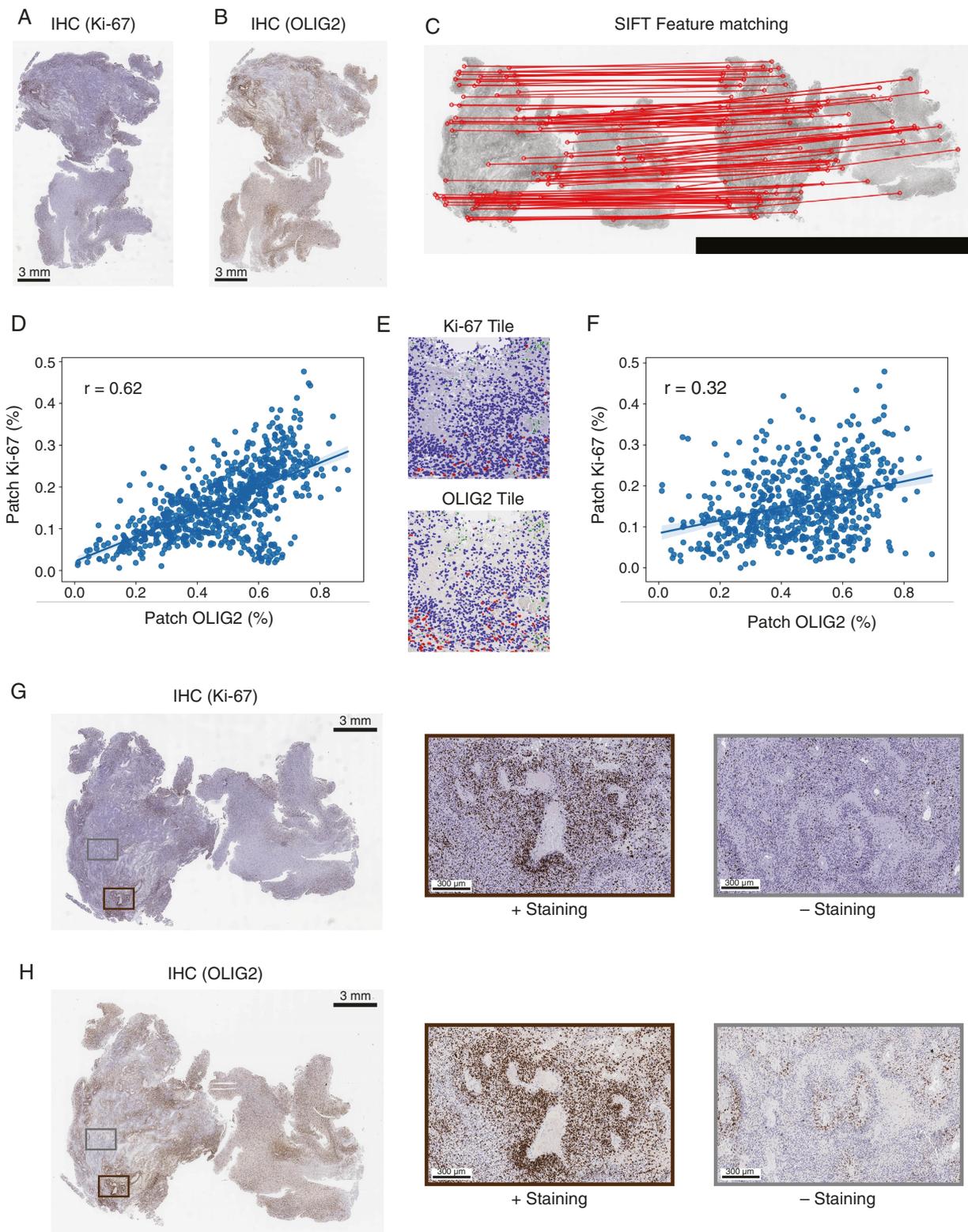
Figure 4. **Spatial correlation of Ki67 and OLIG2 IHC staining using WSI registration and mask R-CNN. a–b.** Illustrative case of Ki67/MIB1 stain and OLIG2 stain of a glioblastoma WSI. **c**. Good matches between the two IHC WSIs evaluated by SIFT. **d**. Quantitative assessment of regional tile-level relationships across studies with SIFT-based alignment. **e**. Example tiles that were aligned showing similar expression pattern. **f**. Quantitative assessment of regional tile-level relationships across studies without prior alignment of original WSIs. **g–h**. Inset of Ki67 stain and OLIG2 stain depicting intra-tumoral regions of high positivity and low negativity.

objectively interpret the IDH1-R132H and ATRX stain in their pathology reports were included in the cohort without any other technical or quality control requirements.

Generating $CNN_{H\&E}$-SIFT-CNN_{IHC}-based Interpretations

To reduce classification errors, our SIFT/CNN-based workflow also employs a number of quality control metrics to flag and remove outlier cases that we previously described.[14] Briefly, to automate classification of H&E slides, image tiles consisting of tissue were extracted and only those containing mostly of lesional tissue were considered (ie P_{tile}(glioma) + P_{tile}(mening) + P_{tile}(met ca) + P_{tile}(schwan) ≥85%). If sufficient lesional tiles are identified (≥15), we output the WSI's probability of inclusion for each of the 16 classes by: (i) The average of P_{tile} across all lesional tiles and (ii) The proportion of $DLFV_{tile}$ belonging to pre-computed true DLFV's of a given class. Both $DLFV_{tile}$ and P_{tile} are established using the CNN's pre-FC and softmax output layers respectively.[14] Similar to our previous work, when these predictions are concordant, the matching diagnosis provides a high confidence overall prediction. Discordant classifications between the two approaches were classified as challenging/anomalous and were omitted as not relevant to the image registration objectives of this study. Only the high confidence predictions were used to estimate the accuracy of the model and workflow ($n = 125$).

From the IHC integration study, we took the 38 glioma samples in our initial H&E cohort and added 9 additional glioma samples to increase our overall cohort size ($n = 47$). For each glioma sample, we obtained the corresponding ATRX, IDH1-R132H, TP53, and MIB-1 IHC stains. These additional slides (TP53, MIB1), while not evaluated during subclassification, proved helpful for aligning additional sets of slides of interest using SIFT. If the slides did not meet the alignment score cutoff, they were not included in the analysis (Supplementary Fig 7). In total, 41/47 ATRX and 40/47 IDH1 slides were either aligned or re-aligned (Figure 2e). Since IDH1 is required to subclassify glioma, we ultimately used a total of 40 gliomas. For this immunohistochemistry study, we simplified the glioma tree (Figure 2a) so that we would only examine the glioma's ATRX and IDH1-R132H status for its subclassification. After SIFT-based alignment (either directly or through an IHC slide bridge) the ATRX and IDH1 slides were quantified by their respective binary classifiers. Based on the optimal cutoff (Youden's index) determined by the ROC curves (62.07% and 10.45% positivity for ATRX and IDH1 classifier respectively) (Supplementary Fig 6), the slides were determined to be either "positive" or "negative" for each stain. Once the glioma's ATRX and IDH1 status were determined, we used a human constructed decision-tree to subclassify the glioma samples.

Statistics

We used R (v4.0.4), Python (v3.8) to analyze data and generate figures. Wilcoxon signed-rank (matched pairs) test was used to compare differences in IHC quantification between whole tissue analysis versus SIFT-based tumor analysis. Student's t-test was used to test the difference between red cluster and green cluster in synaptophysin positivity. Three asterisks (***) denote $P < .001$ and four asterisks (****) denote $P < .0001$.

Results

Evaluation of SIFT-based Alignment of H&E- and IHC-stained WSIs

Towards testing this fully automated workflow, we digitally scanned 107 H&E- and IHC-stained section pairs into WSIs spanning common brain tumor types (diffuse gliomas, meningiomas, metastatic carcinomas, schwannomas) encountered in our neuropathology service (Supplementary Table 1). To overcome size and colorization barriers to image registration, our workflow carries out bicubic resizing of WSIs into small thumbnail-like images (~100 × reduction in size) and normalize pixel intensity distribution of the respective IHC through histogram matching, to mirror that of the H&E image (Figure 1b–e, Supplementary Fig 1). On these resulting images, we use SIFT to find features for image matching and filter "good matches" using Lowe's ratio test. The resulting matches are used to generate a homography, mapping points from one image to another, using random sample consensus.

To evaluate the performance of this approach, we correlate an alignment score, defined by the degree of overall tissue overlap, with: (i) the number of "good matches" from Lowe's ratio test and (ii): a manual binary human assessment of WSI alignment (consensus by three independent blind observers) (Figure 1f). Overall, 97/107 (91%) of cases were defined as appropriately matched by the human evaluators, with poor matches being associated with fewer number of good matches by Lowe's ratio test and tissue overlap scores (Figure 1f, Supplementary Fig 2-3). The latter alignment score provided a particularly sensitive and specific metric for WSI alignment (area under the receiver operating characteristic curve (AUC) = 0.96, Figure 1g). With a pre-defined threshold of good matches as shown in Figure 1f, we are able to align H&E- and IHC-stained WSIs pairs in under 1 s, providing an attractive approach for routine clinical applications and large-scale discovery efforts. Interestingly, review of cases with suboptimal alignments revealed numerous technical limitations such as dramatic differences in the overall tissue shapes or multiple duplicate ("deeper") tissue sections on individual slides (Supplementary Fig 3).

SIFT-based Spatial Quantification of IHC Stains for Subclassification of Glioma

We next wanted to demonstrate how this WSI alignment approach could be coupled with contemporary CNN-based image analysis to automate resolution of various spatially dependent components of inter- and intra-patient tissue heterogeneity. To do this, we developed a 16-class H&E and several IHC-based CNN classifiers relevant to surgical neuropathology. Briefly, we used sets of pathologist-annotated

image patches (516 μm^2) generated from WSIs and transfer learning to train the VGG19 CNN[19] to recognize different histomorphologic brain tumor/tissue patterns (Supplementary Fig 4).[20] This resulting 16-class H&E model included images spanning the four common brain tumor types (meningiomas, diffuse gliomas, metastases, and schwannomas), which in aggregate represent ~80% of routine clinical practice.[1] The remaining 12 classes represented normal/nondiagnostic tissue types to allow for lesion segmentation from nondiagnostic regions (eg gray matter, white matter, cerebellar tissue, acute and chronic hemorrhage, crush artifact, necrosis, radiation necrosis, muscle, dura, blank space, and inflammatory cells). For testing, we used this model to segment and classify tumor regions away from the noncontributory tissue classes and then used our IHC classifiers on appropriately aligned sections by SIFT to carry out lesion-specific immunohistochemical assessments.

To highlight automation potential, we applied our workflow to the immunohistochemical subclassification of diffuse gliomas into (i) IDH-wildtype astrocytomas, (ii) IDH-mutated astrocytomas, and (iii) IDH-mutated 1p19q-codeleted oligodendrogliomas (Figure 2). Most IDH-mutated gliomas (~90%) can be detected by IHC for IDH1-R132H mutation (Figure 2a). Within IDH-mutated gliomas, loss of ATRX staining is a specific marker for the astrocytic lineage. To mimic this molecular- and human-inspired pathological workup of these cases, we therefore developed and applied a serial $CNN_{H\&E}$-SIFT-CNN_{IHC} workflow on a prospectively collected set of brain tumors from our service. We first used the H&E classifier ($CNN_{H\&E}$) to differentiate the four common and distinct brain tumor types (98% accuracy, $n = 125$) (Figure 2b). By aligning IHC WSIs relevant to diffuse glioma cases (eg IDH1, ATRX) using the SIFT-based workflow, we could also effectively integrate molecular information (CNN_{IHC}) needed for further subclassification ($n = 47$) (Figure 2c). Using the alignment cutoffs generated in Figure 1g, we matched 79% and 81% of IDH1 and ATRX WSIs with their respective H&E sections. Interestingly, we noticed we could also overcome some poor alignments due to serial changes in tissue shapes, across the different studies, by using other available WSIs (eg TP53, MIB1 in addition to ATRX, IDH) as "bridges" between the H&E and IDH1-R132H/ATRX images (Figure 2d, Supplementary Fig 5). This led to 3 additional matches for each IHC study and an overall matching frequency of 85% and 87%, respectively (Figure 2e). Importantly, the quantitative interpretation of the IHC studies by CNNs following alignment significantly improved for almost all ATRX retained and IDH1 positive WSIs with higher scores achieved when compared to the overall nonaligned sections (P-value$_{IDH}$ = 1.22 × 10^{-4}; P-value$_{ATRX}$ = 1.02 × 10^{-5}) (Figure 2f and g, Supplementary Fig 6). Dramatic improvements in score came from highly heterogenous samples where alignment allowed for appropriate exclusion of noncontributory tissue elements such as necrosis (Figure 2h–o). Similar to initial testing of SIFT, unmatched cases often had technical limitations such as duplicate tissue sections found only on one slide and dramatic changes in overall tissue shape (Supplementary Fig 7). Overall, this highlights the automation capabilities of framework to integrate histomorphologic and IHC-based molecular information in a spatially preserved and quantitative manner with generalizable implications to automating molecularly driven subclassification and grading tasks.

SIFT-based Evaluation of Intra-tumoral Heterogeneity

We also explored applications of our workflow to various concepts of intra-tumoral heterogeneity (Figure 3). Diffuse gliomas, particularly glioblastomas (GBMs), are known to show "multiform" region-to-region variations in histomorphologies (eg infiltrative, cellular, and hypoxic niches) and molecular markers.[21,22] As our 16-class supervised CNN model was however not designed to resolve intra-tumoral differences (Figure 3a and b), we used a previously diversely trained CNN[15] to serve as a feature extractor and help define areas of morphologic heterogeneity. Unsupervised analysis of image patch-level tissue patterns highlighted two major histomorphologic clusters ($k = 2$ agglomerative hierarchical clustering, silhouette score) (Figure 3c and d, Supplementary Fig 8). Despite these regions initially appearing as cellular tumor areas on H&E, one subregion, following alignment with the synaptophysin IHC, could be better defined to represent an infiltrative tumor component (Synaptophysin positivity: 52% vs. 13% (red vs. green clusters), P-value = 2.2 × 10^{-16}; Figure 3e and f, Supplementary Fig 9). This approach could therefore serve to automate objective selection of highly pure area of tumor of expressed-based molecular analysis or explore the heterogenous expression of specific tumor markers across different tumor niches.

SIFT-based Discovery of Novel Spatial Relationships Between Molecular Markers of Gliomas

We also used this workflow to align and examine the global spatial variation and relationships in the staining patterns of other IHC studies we routinely order for gliomas. Perhaps the most interesting example for this presented case, was a strong correlation (r = 0.62) when aligning mask R-CNN-based estimates of the proliferation marker Ki-67/MIB1 in each image patch (516 μm^2) with the values of the glial-lineage transcription factor Olig2 (Figure 4a–e). This interesting spatial relationship would have likely gone unappreciated without matching (r = 0.32, Figure 4f). Importantly, the specific case used to highlight this relationship was largely all tumor tissue and could not be explained solely based on region-to-region variation in tumor burden (Figure 4g and h). If it is indeed a reoccurring pattern in glioblastoma, this observation could have important implications for emerging clinical trials using Olig2 inhibitors, as it suggests this class of compounds may not effectively reach slowly dividing/quiescent tumor cells that already elude traditional treatment approaches.[23,24] Together, these analyses highlight how this WSI alignment approach can also serve as a discovery tool by objectively and systematically assessing histomorphologic and molecular relationships across multiple studies.

Discussion

Molecular medicine continues to nominate new biomarkers and concepts of inter- and intra-patient heterogeneity for personalized care. These additional parameters however introduce a higher level of complexity for both routine disease subclassification, research investigation, and discovery. Here we develop an automated image registration workflow that can handle the large sizes and color variability of WSIs. We couple this approach with contemporary deep learning tools to highlight applications in molecular subclassification of glioma and resolution of various aspects of intra-tumor heterogeneity. Importantly, this approach can be carried out in less than 1 s and does not require any additional training or manual steps making it particularly generalizable for WSI comparisons and diagnostic workup across other neoplastic and nonneoplastic tissue types. Furthermore, the automated nature of this workflow can serve as an important discovery platform through large-scale, objective, and systematic analysis of complex histopathological and molecular patterns of neoplastic diseases. Excitingly, this slide alignment approach did not require any site-specific validation steps and in theory should be generalize well to other centers. However, future studies to formally evaluate the performance of this approach across laboratories are required given differences in slide generation protocols across different clinical centers (eg slide scanners, staining methods, tissue thickness).[25,26] Furthermore, we note that in some cases these alignments could occur even in the presence of tissue folds and tears in tissue provided that a sufficient number of good matches could be secured in other areas of the slide (Supplementary Fig 2). It is likely that larger tissue sections with more distinctive tissue patterns may be more robust to tissue folds and artifacts when compared to smaller and less heterogeneous tissue fragments (eg uniform circular and rectangular tissue shapes). Such parameters could be further explored in future studies to further refine metrics in which this workflow is most robust and reliable.

Overall, we expect that efficient approaches to WSI registration, such as the workflow presented here, to be an essential step for advanced data integration across multiple histopathologic and molecular studies. We believe our simple and fully automated approach therefore provides an important blueprint toward adoption and translation of computational pathology in the modern molecular era.

Supplementary Material

Supplementary material is available at *Neuro-Oncology Advances* online.

Keywords

computer vision | deep learning | digital pathology | histopathology | image registration | molecular pathology.

Author Contributions

K.F., M.L., and P.D. conceived the idea and approach. P.D. and U.D. supervised the project. K.F. developed the computational workflow. C.F., N.A., M.R., A.G., A.P. and P.D. annotated and generated the image set and interpreted the histological data. K.F., M.L. and A.D. interpreted the alignment data. K.F., M.L., and P.D. wrote the manuscript with input from all other authors.

Conflict of interest statement*:* The authors declare no conflicts of interest.

Data Availability

Histopathology images used to train the VGG19 network can be found at https://bitbucket.org/diamandislabii/faust-feature-vectors-2019. The trained CNN models and code used in this study is publicly available at https://bitbucket.org/diamandislabii/faust-alignment-2021. The original whole slide images used in the study are available from the corresponding author upon request.

References

1. Louis DN, Perry A, Reifenberger G, et al. The 2016 World Health Organization classification of tumors of the central nervous system: a summary. *Acta Neuropathol.* 2016; 131(6):803–820.
2. Kovacs GG. Molecular pathological classification of neurodegenerative diseases: turning towards precision medicine. *Int J Mol Sci.* 2016; 17(2).
3. Campanella G, Hanna MG, Geneslaw L, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med.* 2019; 25(8):1301–1309.
4. Faust K, Roohi A, Leon AJ, et al. Unsupervised resolution of histomorphologic heterogeneity in renal cell carcinoma using a brain tumor–educated neural network. *JCO Clin. Cancer Inform.* 2020; 4(4):811–821.

5. Coudray N, Ocampo PS, Sakellaropoulos T, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat Med.* 2018; 24(10):1559–1567.

6. Harmon SA, Patel PG, Sanford TH, et al. High throughput assessment of biomarkers in tissue microarrays using artificial intelligence: PTEN loss as a proof-of-principle in multi-center prostate cancer cohorts. *Mod Pathol.* 2021; 34(2):478–489.

7. Lowe DG. Object recognition from local scale-invariant features. In: *Proceedings of the IEEE International Conference on Computer Vision.* Vol 2. IEEE; 1999:1150–1157.

8. Gonzalez RC, Woods RE. *Digital Image Processing.* New York City, NY: Pearson; 2018.

9. Muja M, Lowe D. *FLANN - Fast Library for Approximate Nearest Neighbors User Manual.* Published online February 2009. https://usermanual.wiki/Document/flannmanual184.1650808969/view.

10. Lowe DG. Distinctive image features from scale-invariant keypoints. *Int J Comput Vis.* 2004; 60(2):91–110.

11. Borovec J, Kybic J, Arganda-Carreras I, et al. ANHIR: automatic non-rigid histological image registration challenge. *IEEE Trans Med Imaging.* 2020; 39(10):3042–3052.

12. Ward JH. Hierarchical grouping to optimize an objective function. *J Am Stat Assoc.* 1963; 58(301):236–244.

13. Youden WJ. Index for rating diagnostic tests. *Cancer.* 1950; 3(1):32–35.

14. Faust K, Xie Q, Han D, et al. Visualizing histopathologic deep learning classification and anomaly detection using nonlinear feature space dimensionality reduction. *BMC Bioinf.* 2018; 19(1).

15. Faust K, Bala S, van Ommeren R, et al. Intelligent feature engineering and ontological mapping of brain tumour histomorphologies by deep learning. *Nat Mach Intell.* 2019; 1(7):316–321.

16. Diamandis P, Aldape KD. Insights from molecular profiling of adult glioma. *J Clin Oncol.* 2017; 35(21):2386–2393.

17. He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. *IEEE Trans Pattern Anal Mach Intell.* 2017; 42(2):386–397.

18. Skalski P. Make sense. https://github.com/SkalskiP/make-sense. Accessed August 26, 2021.

19. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv:14091556 [cs]. Published online April 10, 2015. http://arxiv.org/abs/1409.1556. Accessed January 21, 2022.

20. Faust K, Xie Q, Han D, et al. Visualizing histopathologic deep learning classification and anomaly detection using nonlinear feature space dimensionality reduction. *BMC Bioinf.* 2018; 19(1):173.

21. Lam KHB, Valkanas K, Djuric U, Diamandis P. Unifying models of glioblastoma's intra-tumoral heterogeneity. *Neuro-Oncology Advances.* Published online August 2020.

22. Puchalski RB, Shah N, Miller J, et al. An anatomic transcriptional atlas of human glioblastoma. *Science.* 2018; 360(6389):660–663.

23. Ligon KL, Huillard E, Mehta S, et al. Olig2-regulated lineage-restricted pathway controls replication competence in neural stem cells and malignant glioma. *Neuron.* 2007; 53(4):503–517.

24. Singh SK, Fiorelli R, Kupp R, et al. Post-translational modifications of OLIG2 regulate glioma invasion through the TGF-β pathway. *Cell Rep.* 2016; 16(4):950–966.

25. Aung TN, Acs B, Warrell J, et al. A new tool for technical standardization of the Ki67 immunohistochemical assay. *Mod Pathol.* 2021; 34(7):1261–1270.

26. Inoue T, Yagi Y. Color standardization and optimization in whole slide imaging. *Clin Diagn Pathol* 2020; 4(1).