Stitching and registering highly multiplexed whole-slide images of tissues and tumors using ASHLAR

Jeremy L. Muhlich 1,2, Yu-An Chen 1,2, Clarence Yapp 1,2, Douglas Russell 1,2, Sandro Santagata 1,2,3 and Peter K. Sorger 1,2,4,*

1 Human Tumor Atlas Network, Harvard Medical School, Boston, MA 02115, USA, 2 Harvard Ludwig Cancer Center and Laboratory of Systems Pharmacology, Harvard Medical School, Boston, MA 02115, USA, 3 Department of Pathology, Brigham and Women’s Hospital, Boston, MA 02115, USA and 4 Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA

*To whom correspondence should be addressed.

Associate Editor: Alfonso Valencia

Received on June 7, 2021; revised on April 4, 2022; editorial decision on July 2, 2022

Abstract

Motivation: Stitching microscope images into a mosaic is an essential step in the analysis and visualization of large biological specimens, particularly human and animal tissues. Recent approaches to highly multiplexed imaging generate high-plex data from sequential rounds of lower-plex imaging. These multiplexed imaging methods promise to yield precise molecular single-cell data and information on cellular neighborhoods and tissue architecture. However, attaining mosaic images with single-cell accuracy requires robust image stitching and image registration capabilities that are not met by existing methods.

Results: We describe the development and testing of ASHLAR, a Python tool for coordinated stitching and registration of 10^3 or more individual multiplexed images to generate accurate whole-slide mosaics. ASHLAR reads image formats from most commercial microscopes and slide scanners, and we show that it performs better than existing open-source and commercial software. ASHLAR outputs standard OME-TIFF images that are ready for analysis by other open-source tools and recently developed image analysis pipelines.

Availability and implementation: ASHLAR is written in Python and is available under the MIT license at https://github.com/labsyspharm/ashlar. The newly published data underlying this article are available in Sage Synapse at https://dx.doi.org/10.7303/syn25826362; the availability of other previously published data re-analyzed in this article is described in Supplementary Table S4. An informational website with user guides and test data is available at https://labsyspharm.github.io/ashlar/.

Contact: peter_sorger@hms.harvard.edu

Supplementary information: Supplementary data are available at Bioinformatics online.

1 Introduction

Multiple approaches have been described for performing 20–60 plex subcellular resolution microscopy on normal and diseased tissues for research and diagnostic purposes (Angelo et al., 2014; Gerdes et al., 2013; Giesen et al., 2014; Goltsev et al., 2018; Lin et al., 2018; Tsujikawa et al., 2017). These methods make it possible to image differentiation markers, signaling proteins, cell cycle regulators, oncogenes and drug targets in a preserved tissue context. The resulting data can be processed to determine the molecular and physical relationships of cells within the tissue to each other, to the local vasculature, and to the non-cellular components within connective tissue or basement membranes. Research has shown that spatial profiling by highly multiplexed microscopy can reveal features of normal and diseased tissues and their responses to therapy that cannot be discerned in other ways (Färkkilä et al., 2020; Goltsev et al., 2018; Launonen et al., 2022; Schürch et al., 2020; Wagner et al., 2019). For this reason, multiplexed spatial profiling of proteins and mRNA is the cornerstone of large scale atlasing projects such as the Human Cell Atlas (Regev et al., 2017), NIH HuBMAP consortium (HuBMAP Consortium, 2019) and NCI Human Tumor Atlas Network (HTAN) (Rozenblatt-Rosen et al., 2020). Such atlases promise to fundamentally advance understanding of tissue development and physiology and improve how diseases are diagnosed and individual patients matched to optimal therapies.

Highly multiplexed imaging of proteins in tissues uses antibodies to detect specific antigens, building on 80 years of experience with immunohistochemistry in research and diagnostic settings (Wick, 2012). Methods such as MxIF, CyCIF, CODEX, 4i and mIHC use conventional fluorescence and brightfield microscopes, whereas

Bioinformatics, 38(19), 2022, 4613–4621
https://doi.org/10.1093/bioinformatics/btac544
Advance Access Publication Date: 16 August 2022
Original Paper
Multiplexed Ion Beam Imaging (MBI) and Imaging Mass Cytometry (IMC) vaporize specimens with ion beams or lasers followed by atomic mass spectrometry (Angelo et al., 2014; Gerdes et al., 2013; Giesen et al., 2014; Goltsev et al., 2018; Gut et al., 2018; Lin et al., 2018; Tsjikawa et al., 2017). Approaches to imaging nucleic acids are based on hybridization (Chen et al., 2018; Lin et al., 2014) and sequencing (Stahl et al., 2016). Some methods require frozen samples, but methods that use Formaldehyde Fixed Paraffin Embedded specimens—the sample type universally acquired for diagnostic purposes—can tap into large archives of human biopsy and resection specimens (Burger et al., 2021).

Existing imaging methods differ in resolution, field of view and number of distinct antigens or genes that can be detected (the assay ‘plex’). Most immunofluorescence methods (e.g. MxIF, CyCIF, mHC, CODEX and Immuno-SABER) are cyclic approaches in which high-plex data are generated by the repeated acquisition of lower-plex images, each of which has two to six channels of information. Each channel represents an image acquired with excitation and emission filters matching one antibody or oligo-coupled fluorophore. As such, a single image represents an image acquired with excitation and emission filters matching one antibody or oligo-coupled fluorophore. Thus, each tile is a multi-wavelength megapixel-scale image that represents a different lateral (x, y) stage position. The number of wavelengths in each tile, the number of tiles and the number of imaging ‘cycles’ (each of which involves the acquisition of a full set of tiles), differs with the microscope and the multiplexing technology. However, it is universally true that tiles from all cycles must be merged accurately into a single high-plex ‘mosaic’ image.

High-plex mosaic images represent the key ‘Level 2 or 3’ data type for all subsequent visualization and quantitative data analysis. The data level concept was introduced by dbGAP for genomics (Tryka et al., 2014) and its implementation to tissue imaging is described in detail in the MITI guidelines (Schapiro et al., 2022a). In this context, ‘data levels’ denote different degrees of data processing, with Level 1 corresponding to single, raw image tiles, Level 2 data to stitched, illumination corrected mosaics and Level 3 to mosaic images that have also been subjected to manual or automated quality control to improve interpretability and accuracy.

It is increasingly clear that the greatest challenges in the acquisition of high-plex image data arise in the data acquisition per se, but in the subsequent image processing steps. For example, even the best microscopes require computational alignment of tiles to form a mosaic, since mechanical tolerances and imperfect calibration introduce uncertainty into recorded tile positions. To enable the assembly of a mosaic, tiles are slightly overlapped during acquisition so that each pair of adjacent tiles contains some identical cells. Image features in these cells are then used as reference points for ‘stitching’ adjacent tiles into a seamless mosaic. In cyclic imaging (Gerdes et al., 2013; Lin et al., 2018), all tiles from the second and subsequent cycles must be aligned to the mosaic through ‘registration’ of image features across corresponding tiles. DNA-stained nuclei serve as an excellent image feature for alignment since they stain well with a variety of fluorescent dyes, are present at suitable density in most tissue types, and have sharp edges with high contrast. Multiple tools exist for registering image stacks and stitching image tiles (Chalfoun et al., 2017; Holtkamp and Goshatsby, 2009; Horl et al., 2019) and some are available in common image analysis software such as ImageJ (Schneider et al., 2012). However, we have found that open-source tools currently available for stitching and registering multi-tile image stacks are not satisfactory when applied to high-plex cyclic images with respect to speed, reliability and accuracy. Some commercial instruments have also integrated stitching routines, but we have found that these methods are only sufficient for visual review and are generally not accurate enough for quantitative single-cell analysis. Existing tools also struggle with very large images and generally require substantial format conversion and file renaming, a non-trivial task when confronted with 100 GB of data contained in 107 megapixel-scale image tiles (a large 10-cycle whole-slide image).

In this article, we report the development of a new open-source Python package, ASHLAR (Alignment by Simultaneous Harmonization of Layer/Adjacency Registration), for coordinated stitching and registration of multiplexed, multi-tile images. The package offers both a command line file-oriented interface and a documented Application Programming Interface (API) for incorporation into other tools. ASHLAR can directly process any image format supported by the widely used Open Microscopy Environment (OME) BioFormats library (Li et al., 2016) and it outputs standard OME-TIFF files. We describe ASHLAR’s design and implementation and compare its performance to existing tools using high-plex CyCIF images. ASHLAR is available as a Docker or Singularity container and has been incorporated into MCMICRO (Schapiro et al., 2022a), the Nextflow-based image processing pipeline developed by HTAN; as part of MCMICRO, ASHLAR has been tested with several hundred CyCIF, CODEX and mXIF images acquired from 12 types of mouse and human tissues at seven different institutions on five different microscopes and slide scanner platforms (Supplementary Table S1). ASHLAR is therefore a robust and practical tool for use with diverse spatial profiling methods.

2 Materials and methods

2.1 Overview of the assembly process for cyclic multi-plex fluorescence images

ASHLAR operates in three broad phases to convert a multi-cycle multi-tile (Level 1) dataset into a cohesive (Level 2) mosaic image (Schapiro et al., 2022b) (Fig. 1): (i) tiles within the first imaging cycle are stitched; (ii) tiles from the second and subsequent cycles are registered to corresponding tiles from the first cycle; and (iii) all tiles from all cycles are merged into a mosaic image. The output of stitching and registration is a list of new, corrected positions for all tiles in each cycle. Only in the final mosaic phase is the actual full-size many-channel mosaic image created. These mosaic images can be very large and contain information spanning length scales from <1 μm (subcellular structures) to cm in dimension (gross tissue morphology). In many cases, the boundary of a tissue specimen is irregular, and a significant fraction of the tiles in a rectangular data collection grid contain few if any cells, posing a challenge for stitching as well as an opportunity to reduce data collection demands by creating irregular-shaped tile sets that closely follow the outlines of the tissue.

To generate an initial estimate of tile positions, ASHLAR uses data from image tiles (grids of pixels), recorded stage positions and physical pixel dimensions. ASHLAR leverages the Open Microscopy Environment Bio-Formats library (Li et al., 2016) to extract the necessary image data and metadata (stage position and pixel size) directly from native image files produced by the great majority of commercial microscopes, obviating the need for image format conversion and manual metadata extraction. For microscopes that do not support BioFormats, ASHLAR accepts a set of TIFF files using a configurable naming convention along with an explicit specification of tile overlap and acquisition order. Subsequent stitching and registration involve aligning one image to another. In stitching, the small overlapping strips of adjacent tiles are aligned (Fig. 1b), and in registration, full tiles that cover the same region of the sample but acquired in different cycles are aligned (Fig. 1c). We performed stitching and registration only on the reference image channel (typically Hoechst 33342-stained nuclei) and applied the resulting positional corrections to all other channels recorded within that cycle. This is sufficient because the chromatic aberration exhibited by research-grade wide-field microscopes is not a major contributor to image inaccuracy at resolutions typically used for tissue imaging (10–40× magnification, 0.3–0.95 NA air objectives).
that almost all of the discrepancy between recorded stage positions and true positions can be modeled by translation alone. Stage position errors encountered between multiple imaging cycles are also purely translational, as long as the slide is always placed at exactly the same angle on the stage; this can routinely be achieved with kinematic mountings that positively register the slide in a consistent position (this is a standard feature of contemporary slide-scanning microscopes).

ASHLAR uses an enhanced method of phase correlation that improves the precision of tile alignment. The accuracy of classical phase correlation is limited to whole pixels. At pixel sizes around 1 μm or larger, this represents a substantial error relative to the size of a single cell. We overcame this limitation by using an improved phase correlation algorithm (Guizar-Sicairos et al., 2008) that offers arbitrary sub-pixel precision with minimal extra computation. An alignment precision of 0.1 pixels produced a discernible improvement in final mosaic quality over whole-pixel alignment, with diminishing returns beyond that. ASHLAR also enhances phase correlation by pre-filtering input images with the discrete Laplacian operator (or Laplacian of Gaussian operator—LoG—for noisy images) to eliminate auto-correlation. It has been understood for at least a century that computing cross-correlation can yield spurious results with signals that exhibit auto-correlation, but this fact is often overlooked in practice (Dean and Dunsmuir, 2016; Yule, 1926)—we are aware of only one open-source image stitching tool, ITKMontage (Zukić et al., 2021), that performs decorrelation. Our work with ASHLAR shows that decorrelation substantially improves confidence in image tile alignments.

### 2.3 Image stitching

The stitching procedure begins with the creation of a node-edge adjacency graph in which nodes represent tiles (Fig. 2, Step A1). Edges are added to the graph to connect overlapping tile pairs, which are initially identified by consulting recorded stage positions and other metadata. By reading recorded stage positions directly from BioFormats metadata, it is straightforward to support samples imaged with non-rectangular grids and irregular layouts. Figure 2a shows the overlap in adjacent tiles associated with one edge in the adjacency graph and reveals that one image is slightly translated relative to the next—this translation represents the stage positioning error we are trying to correct. When the overlap region contains many cell nuclei or other alignment features, phase correlation can accurately and confidently compute the correct translation between the images. Phase correlation will always return some value for translation of any two tiles, even when the overlap region is uninformative and contains only incidental signal or background noise in the registration channel; in these cases, we rely on the recorded stage positions. Uninformative overlaps in tissue sections are most commonly encountered when nuclei are scant, such as in fat, connective tissue or regions of necrosis, and in areas in which no tissue is present, such as along the edge of a specimen, between separate pieces of tissue or between the circular cores that make up a tissue microarray (TMA) (a regular grid of 50–200 0.3–1.2 mm pieces of tissue arrayed on a single microscope slide).

We use normalized cross-correlation (NCC) to score how well the translation returned by phase correlation aligns images, but the threshold dividing an effective alignment from a spurious one varies by dataset. We estimate this threshold by the 99th percentile of NCC values computed from a permutation test that considers 1000 randomly selected pairs of non-adjacent tiles (Fig. 2a, Step A2); this corresponds to the unadjusted one-sided empirical P-value threshold of 0.01. For each tile pair represented by an edge in the adjacency graph, we crop the images to their mutually overlapping region based on recorded stage positions and align them using phase correlation as described above. This yields a corrected X, Y shift and NCC value (Fig. 2, Step A3). For all downstream steps, we use the negative logarithm of the NCC values, hereafter referred to as ENCC (NCC error), which provides a more intuitive ‘lower is better’ error metric and empirically appears to have a more normal distribution. We use this ENCC threshold and a user-provided translation limit parameter to determine whether to discard low-quality tile pair
alignments (Fig. 3b). The value of the translation limit is not particularly critical, as physical translation distances for spurious alignments tend to be fairly extreme (note that the Y-axis in Fig. 3b is on a log scale). When a low-quality alignment is discarded, we delete the corresponding edge from the adjacency graph (Fig. 2, Step A3 and Fig. 3c). At this point, there are almost always more remaining pairwise alignments than tiles, leading to an overconstrained system. We solve this by constructing a minimum spanning tree with the ENCC values as the edge weights and retaining only the alignments corresponding to edges in this tree (Fig. 2, Step A4). This allows us to discard extraneous alignments so the position of every tile is unambiguous. Since the edge deletion process in Step A3 could split the graph into multiple disconnected pieces, we perform the spanning tree procedure independently for each piece.

With this spanning tree, it is straightforward to obtain final corrected positions by walking along the edges from tile to tile (starting at the root) and adding up each pairwise alignment along the way (Fig. 3d). Even though individual pairwise tile alignments correct primarily for local uncorrelated stage position error, taken collectively they also characterize systematic errors such as miscalibrated pixel size or Z-axis camera rotation. To quantify these types of errors, we perform multiple linear regression of the corrected tile positions after stitching (dependent variable) against the original tile positions, we down-sample the data by a factor of 20 and assemble low-resolution ‘thumbnail’ mosaic images for each cycle using the recorded stage positions. We then align the thumbnails

2.4 Image registration

The procedure for registering subsequent cycles against the first cycle uses many of the same tools as stitching, although the goal is aligning whole tiles across data acquisition cycles rather than aligning adjacent tiles edge-to-edge within a single cycle (Fig. 2b). First, we establish a correspondence between each tile in the target cycle (the one to be registered) and the nearest tile in the first cycle by comparing recorded stage positions (Fig. 2b, Step B1). Identifying these tile correspondences is trivial when recorded stage positions are consistent from run to run, and the geometry of the image acquisition grids is identical. However, a significant shift in stage positions can occur between cycles with microscopes that lack a physical ‘homing’ procedure to zero stage position encoders at start-up. Shifts also arise where the planned tile grid is significantly displaced or rearranged between runs. To account for this shift when comparing tile positions, we down-sample the data by a factor of 20 and assemble low-resolution ‘thumbnail’ mosaic images for each cycle using the recorded stage positions. We then align the thumbnails
using phase correlation with sub-pixel precision to obtain a coarse alignment between the first and target cycles. Working with low-resolution images in this step saves compute time and memory while providing sufficient precision to accurately recover inter-cycle tile correspondences. Next, each corresponding tile pair is cropped to mutually overlapping regions and aligned using phase correlation (Step B3). The resulting alignments are then filtered using the user-specified translation limit. Note that we do not currently use a permutation test and E_LNC threshold in the registration phase (Step B2) as the translation limit alone has been sufficient for all images processed to date. For alignments that pass the filter, the target-cycle tiles are positioned by adding the alignment translations to the corrected positions of the corresponding tiles from the first cycle (Step B4). The remaining tiles (generally those with sparse or no tissue, or where the tissue was damaged significantly during inter-cycle sample handling) are positioned using the affine transformation computed in the stitching phase, assuming that the same microscope and calibration conditions were used for both cycles (Step B5). The registration steps described above are then applied to all other cycles, establishing corrected global positions for all remaining tiles. Importantly, our method registers each cycle of tiles against the first cycle (rather than consecutively against each preceding cycle) because each successive alignment step incurs additional error.

### 2.5 Mosaic image generation

The result of the stitching and registration phases is a corrected global position for every image tile (Fig. 4). To generate the final output image mosaic, we create an empty image large enough to encompass all corrected tile positions and copy each tile into it at the appropriate coordinates. Since each pairwise image registration is computed to a precision of 0.1 pixels as described above, the sum of these shifts for a given tile generally yields non-integer values for the final coordinates. ASHLAR defaults to applying sub-pixel translations on the tile images to account for this, but some users may prefer to round the final positions to the nearest pixel instead. Where neighboring image tiles overlap in the mosaic, they are combined with linear blending or one of several other user-selectable blending functions. The final many-channel image is then written out as a standard OME-TIFF file containing a multi-resolution image pyramid to support efficient visualization.

### 2.6 Implementation

ASHLAR is implemented in Python 3 and utilizes many numerical and image processing routines from the numpy, scipy, scikit-image, scikit-learn and networkx packages. The pyjnius Python-to-Java connector provides access to the Java BioFormats library for reading microscopy image files. The user interface is mainly via command-line script, but the underlying Python modules may also be used directly.

### 3 Results

#### 3.1 Evaluation of stitching accuracy

We identified MIST (Chalfoun et al., 2017) as the current state-of-the-art public-domain tool for stitching large, tiled microscopy images. We used the evaluation framework described by Chalfoun et al. to compare the accuracy of stitching by ASHLAR and MIST using that manuscript’s most challenging dataset: a plate of widely-spaced GFP-labeled stem cell colonies that were grown for 2 days and imaged with 10% tile overlap. The Chalfoun et al. dataset contains two image sets acquired via separate mechanisms: (i) images acquired with ‘traditional’ overlapping tiles and (ii) ground-truth images—with each colony centered and wholly contained in a single image field—acquired with a closed-loop microscope stage control algorithm. The Chalfoun et al. evaluation framework assesses the accuracy of a stitching algorithm by applying the algorithm to the overlapping tile set, segmenting the stitched output mosaic into colonies, and finally comparing each colony’s area and position to the ground truth data. Four metrics are reported: False negative count (FN), false positive count (FP), per-colony size error ($S_{\text{err}}$) and per-colony position distance error ($D_{\text{err}}$) (see Chalfoun et al. for full details). MIST and ASHLAR yielded the same false negative and false positive counts ($FN = 47$, $FP = 4$), the size error distributions were nearly the same and had indistinguishable medians (median $S_{\text{err}} = 0.0474\%$), and the distance error distributions were also similar with a difference between medians that was not statistically significant (MIST median $D_{\text{err}} = 10.8$ pixels, ASHLAR median $D_{\text{err}} = 11.5$ pixels, Mood’s median test $P$-value = 0.32) (Fig. 5a). We conclude that MIST and ASHLAR have similar stitching accuracy when applied to a previously described test set involving cells grown in vitro.

#### 3.2 Benchmark dataset for evaluating registration accuracy

As a first test of ASHLAR and MIST on high-plex, whole-slide tissue data, we acquired a $\sim$24 mm $\times$ 14 mm $\times$ 5 mm thick section of a human normal colon sample from the Cooperative Human Tumor Network (https://www.chtn.org/). This sample was subjected to 9-cycle t-CyCIF (Lin et al., 2018) to generate a subcellular-resolution 28-plex image. Cell nuclei were stained with Hoechst 33342 in every cycle, providing reference features for image alignment. Imaging was performed on a RareCyte CyteFinder II HT Instrument with a 20x 0.75 numerical aperture (NA) objective and four excitation and emission filter pairs having peak and full-width at half-maximum.
3.3 Image registration of independently stitched mosaics

Whereas ASHLAR performs stitching and registration in a coordinated process, it was necessary to globally register the MIST output mosaics before evaluating local registration accuracy. We first downsampled the MIST mosaic images by $10 \times$ to obtain a manageable image size and then aligned them with subpixel-precision phase correlation (phase correlation on the full-size images required a computer with more RAM than was readily available to us). We performed the same process on the ASHLAR mosaics to verify that their global registration was already optimal.

3.4 Optical flow computation and evaluation of registration accuracy and robustness

We used dense optical flow fields to quantify and visualize local registration errors in the Hoechst reference channel. Because we could not find any general-purpose dense optical flow implementations capable of processing gigapixel-scale images on a reasonable workstation computer, we implemented our own approximate method suitable for small-magnitude flow fields using a block-based approach which is memory-efficient and highly parallelizable. To perform the same process on the ASHLAR mosaics to verify that their global registration was already optimal.

The local registration error as the magnitude of the flow vector field at each point. The median error was $1.94 \mu m$ for the MIST mosaics ($\sim 3$ pixels) and $0.119 \mu m$ ($\sim 0.2$ pixels) for the corresponding ASHLAR mosaics (Fig. 3b). At a magnification sufficient to see individual cells, the error generated by MIST was readily apparent (Fig. 5e). Visualizing the full vector field on top of the reference channel images (Fig. 3d) showed that the field direction was often consistent across large regions but could change dramatically at tile boundaries. This most likely arises because small local stitching differences propagate across the mosaic in a manner that is uncorrelated between cycles, leading to inter-cycle shifts that cannot be corrected by any rigid adjustment of the entire mosaic. It is important to note that this is not a weakness of MIST per se, but rather a consequence of applying a tool designed for stitching alone to the combined process of stitching and registration, a use case for which MIST was not designed. With the ASHLAR-generated mosaic, vector field visualizations confirmed a much lower level of registration error (Fig. 5e and f). Close inspection of the few regions with high error showed that they corresponded to parts of the tissue in which cells had become physically distorted or detached from the slide as a consequence of the cyclic staining procedure. Thus, the remaining errors are not a result of errors in registration and stitching but rather extrinsic processes that must be identified and accounted for by downstream error-checking procedures.

To demonstrate ASHLAR’s robustness and versatility with different image types, we compared its registration accuracy against MIST on four further datasets encompassing two cyclic imaging techniques, three vendors’ microscopes, and two new tissue types plus a TMA that itself spans a multitude of tissue types and disease states. We also evaluated one additional open-source stitching algorithm, BigStitcher (Hörl et al., 2019), on some of these datasets. The

Fig. 5. Performance comparison of ASHLAR and MIST software. (a) Stitching error metric distributions (kernel density estimate) for MIST and ASHLAR computed according to the stitching evaluation framework of Chalfoun et al. (2017). Dotted lines indicate median values; neither difference in medians was statistically significant. (b-d) Images and data derive from analysis of the multi-tile image of human colon shown in Figures 3 and 4. (b) Local registration error distance distributions for MIST and ASHLAR mosaic images of two CyCIF cycles of a human colon section. Distances at the upper end in this plot as well as in panels d and f were clipped to the 90th percentile of the MIST error values ($\sim 4\mu m$) to highlight the relevant data. (c) Full-resolution view of four regions from the MIST mosaics demonstrating local registration error in different directions. The Hoechst images of nuclei from cycles 1 and 2 are pseudocolored red and cyan, respectively, to visualize the effect of registration error at the single-cell level. The MIST median error of $\sim 2 \mu m$ is around one-quarter of the diameter of the average cell nucleus, a shift that is clearly visible at full resolution. (d) Heatmap of MIST local registration error direction (hue) and magnitude (intensity) at 200-pixel resolution overlaid on the Hoechst image (brighter colors indicate larger errors). Characteristic tile-sized scale of heatmap features suggests inconsistent stitching. Yellow X marks indicate locations highlighted in panel e. (e) The same regions as in panel c, but taken from ASHLAR mosaics. An identical pseudocoloring scheme is used; the red and cyan images, now more accurately registered, combine to appear nearly white. The ASHLAR median registration error of $\sim 0.1 \mu m$ is $\sim 1\%$ of the diameter of a nucleus. (f) Heatmap of ASHLAR local registration error using the same intensity and hue scale as in panel d showing overall lower error and no apparent tile-scale features. Remaining small-scale errors represent damaged tissue that could not be registered (A color version of this figure appears in the online version of this article.)

bandpass wavelengths (in nm) of: 395/25–438/26, 485/25–522/20, 555/20–590/20 and 651/10–692/40, respectively. $2 \times 2$ pixel binning was used during acquisition yielding a four-channel tile with dimensions of $1280 \times 1080$ pixels and a pixel size of $0.65 \mu m$ per pixel. To image the entire specimen, a grid of 609 (29 $\times$ 21) tiles was used and each tile overlapped by $\sim 31$ pixels ($20 \mu m$ or 2–3%) in both directions. Each cycle yielded one OME BioFormats-compatible RCPNL file containing 16-bit imaging data from all 609 tiles, $\sim 7$ GB in size. The entire 9-cycle dataset comprises 5481 image mosaics plus a TMA that itself spans a multitude of tissue types and disease states. We also evaluated one additional open-source stitching algorithm, BigStitcher (Hörl et al., 2019), on some of these datasets. The...
datasets are described in Supplementary Table S4 and the evaluation results are shown in Supplementary Figures S1–S5. Overall, ASHLAR compared favorably to MIST and BigStitcher on all datasets with respect to registration of cyclic datasets.

3.5 Evaluation of commercial stitching algorithms

Slide scanning microscopes include stitching algorithms as part of their data acquisition and analysis software. These algorithms use proprietary file types and in most cases, they erase the original image tiles or strips after generating the final stitched output image. Thus, the detailed evaluation of their performance is not straightforward, but it is possible to evaluate stitching consistency by using our optical flow method to examine two re-scans of the same specimen. We performed such an analysis using a human colorectal adenocarcinoma specimen retrieved from the archives of the Department of Pathology at Brigham and Women’s Hospital with Institutional Review Board (IRB) approval as part of a discarded/excess tissue protocol. The specimen was stained with H&E and scanned in brightfield mode at the Brigham and Women’s Hospital Pathology Core Facility using three different slide scanning microscopes: a Leica Aperio GT450, Leica Aperio VERSA and Hamamatsu NanoZoomer 2.0-HT. We also imaged the adenocarcinoma specimen in brightfield mode using a GE INCell 6000 microscope to produce tiles suitable for processing with ASHLAR. The specimen was imaged twice on each instrument to emulate a cyclic imaging workflow. Pre-stitched mosaic image pairs generated by the three scanners and the ASHLAR-stitched mosaic pair assembled from the INCell 6000 tiles were subjected to the registration accuracy evaluation described above. The results, shown in Figure 6, demonstrate that all of the tested systems exhibit worse stitching consistency than ASHLAR. Inspection of the underlying images reveals obvious stitching errors that would confound single-cell level analysis. We also observed that the error field images also contained ‘signatures’ of each instrument’s internal design, such as line sensor vs. area sensor and sensor size and orientation. Thus, commercial algorithms included with existing slide scanners do not appear to fully correct for intrinsic limitations of the instrumentation. Finally, we evaluated the registration accuracy of the stitching feature in Zeiss’s Zen software which was recently used to generate a publicly available 50-plex rat brain dataset (Maria et al., 2021) based on cyclic fluorescence imaging on a Zeiss Axio Imager Z2 microscope. We compared the DNA channel images from two imaging cycles from this dataset with our evaluation framework and found conspicuous errors here as well (Supplementary Fig. S6). Thus, while commercial algorithms may stitch well enough for visual review and gross structural analysis, they have weaknesses that are very likely to impact single-cell quantification—especially with cyclic imaging methods.

3.6 Runtime and memory usage

The runtime and RAM usage for the stitching and registration phases of ASHLAR are each comparable to that of MIST stitching running on a single CPU core. On the first two cycles of the human colon dataset, ASHLAR required 306 s (149 s to stitch cycle 1 and 157 s to register cycle 2) and 2.1 GB RAM. MIST-FFTW required 228 s (114 s per cycle) and 2.5 GB RAM to compute corrected tile positions. If we include the 30 s per cycle required to convert the datasets from the microscope vendor’s native file format into MIST’s required single-TIFF format (ASHLAR requires no such conversion step) the total for MIST is 288 s. MIST does execute more quickly when allowed to use multiple CPU cores or a GPU; we have not enabled parallel processing in ASHLAR but expect a similar increase in speed. ASHLAR’s runtime per imaging cycle varies linearly with the total number of pixels in all tiles of the reference channel (Supplementary Fig. S7). All measurements were taken on a 3.5 GHz Intel Xeon E5-1620 v3 CPU with 32 GB of RAM and an SK Hynix SH920 512 GB SSD running Ubuntu Linux 20.04. Software versions were as follows: ASHLAR 1.14.0, MIST 2.0.7, Python 3.8.10 and OpenJDK 1.8.0.

4 Discussion

To date, ASHLAR has been used to stitch several hundred whole-slide images collected using diverse acquisition technologies and instruments (Supplementary Table S1). ASHLAR, therefore, provides a robust and efficient way to generate large, multi-channel, mosaic images of tissues and other biological specimens by assembling individual megapixel image tiles collected at multiple wavelengths over multiple imaging cycles. The key innovation for image quality is joint optimization of stitching and registration as opposed to stitching individual cycles independently and then attempting to register mosaic images against each other. Joint optimization becomes increasingly important as the size of the specimen increases. Coupling ASHLAR with tile-based image acquisition and cyclic data collection makes it possible to optimally balance the resolution, size and plex of a tissue image for reliable analysis of spatial features on a wide variety of scales. Although many recent highly multiplexed studies have relied on small fields of view and TMAs, whole-slide imaging is emerging not only as a diagnostic necessity (Center for Devices and Radiological Health, 2015) but also as a key requirement for basic research into the spatial organization of tissue and tumor microenvironments (Lin et al., 2021). ASHLAR is optimized for these data acquisition requirements and is more rapid and
ASHLAR is effective not only with conventional rectangular image acquisition grids but also with images involving multiple non-overlapping regions of interest and tiles that do not form a rectangu-
lar grid. The ability to manage irregular and disconnected specimens has emerged as a key requirement in the broader application of tis-
sue imaging. By instructing a microscope to avoid emptying injury space lying outside of the margins of the tissue, imaging time and file size can be reduced, often by a factor of two or more (a signifi-
cant advantage as datasets approach terabyte scale). We have suc-
cessfully used ASHLAR to assemble images of TMAs, in which 0.3–
1 mm diameter ‘cores’ from multiple tissue specimens are positioned in a regular array on a glass slide, making it possible to analyze over 100 specimens in parallel. This represents a potentially challenging stitching problem since much of the slide is devoid of sample, and in-
dividual cores are often divided among multiple fields. Core biopsies and fine needle aspirations are other samples in which the collection of non-rectangular images is highly advantageous. Such biopsies are typically long and thin (0.3 × 10–20 mm) and rarely aligned along the axis of the slide, making it necessary to collect tiles on a diag-
onal. The ability of ASHLAR to reject spurious alignments using a permutation test followed by pruning of adjacency graphs make the algorithm robust to regions of the images such as these that contain little if any data in the registration channel.

Much of the recent discussion about multiplexed tissue imaging has focused on the importance of the number of channels (assay-
plex) (Baharlou et al., 2019), since more channels allow more pro-
teins or genes to be assayed and yield more detailed molecular insights. However, two other considerations are at least as important: image resolution and field of view (speed also matters for high-
volume applications). In the case of fluorescent imaging of a single tile, resolution and field of view are functions of the optics, primar-
ily the numerical aperture of the objective lens, the properties of the transfer optics and the number of camera pixels—which together specify pixel size (Ghiran, 2011). For large whole-slide images assembled from many tiles, the accuracy of image stitching and registration also becomes critical. ASHLAR directly addresses this requirement. In most applications, it is also combined with other software to optimize the quality of image mosaics. Prior to stitching and registering tiles using ASHLAR we perform illumination correc-
tion using BaSc (Peng et al., 2017), which exploits low-rank and sparse decomposition to correct for uneven shading and background variation in microscope images. This is essential because the illum-
ination of each tile is typically brightest at the center of the field (along the optical axis) and dimmer at the edges.

### 4.1 Limitations

We have found that the spanning tree approach used to combine in-
dividual pairwise tile alignments is broadly effective. However, one recurrent weakness observed with large specimens is that tiles at the margin of the tissue that are adjacent in physical space are often found to lie far apart in the adjacency graph. When corrected posi-
tions are determined, uncorrelated error accumulates as pairwise shifts are added up along the edges in the spanning tree. The result-
ing stitching error is rarely noticeable by eye in the resultant mosaic image, but we have identified this as an area for future improvement of the algorithm.

To achieve reasonable processing speed, ASHLAR makes some compromises with respect to the factors it accounts for during stitch-
ing and registration. For example, ASHLAR currently performs only rectilinear correction of tile position and assumes tiles have the same magnification. When images from different microscopes must be combined it is usually necessary to account for changes in camera angle due to the rotation of individual cameras and their microscope stages relative to each other. Scaling is also frequently different across instruments, even when the same objective is used, due to dif-
fferences in transfer optics and sensor configurations. We have never encountered the need to assemble an image from multiple micro-
scopes (partly because many other batch effects are introduced) but corrections for image rotation and scaling can be performed through minor additions to the alignment procedure (Gonzalez, 2011); we will add these features to ASHLAR as the need arises, most likely arising from a requirement to combine multiple different imaging modalities.

### Acknowledgements

We thank Jerry Lin, Zoltan Maliga, Connor Jacobson, Artem Sokolov and other members of the Laboratory of Systems Pharmacology for their assistance. The normal colon specimen used for this work was acquired from the Cooperative Human Tumor Network (https://www.chn.org/). We thank the members of the HTAN Consortium for their support; a full list can be found at humantumoratlas.org.

### Funding

This work was supported by the National Institutes of Health [U2C-
CA233262 and U2C-CA233280]; the Ludwig Cancer Center; and the Bill and Melinda Gates Foundation [IN-027106].

**Conflict of Interest:** P.K.S. is a member of the scientific advisory board or board of directors of Glencoe Software (which supports a commercial version of the OME software), Applied Biomath and RareCyte Inc. (which built some of the microscopes used in this study) and has equity in these companies; he is on the Scientific Advisory Board of NanoString Inc. and a consultant to Montai Health and Merck. In the last 5 years, the P.K.S. lab has received re-
search funding from Novartis and Merck. P.K.S. declares that none of these relationships altered the conduct or reporting of this research. The other authors report no outside activities.
